HAL
open science

Collaboration massivement distribuée et gestion du savoir en ligne: le cas de la communauté WordPress (2003-2008). "Code is poetry", but a "documentation resource should be managed"

Emmanuel Ruzé

To cite this version:

Emmanuel Ruzé. Collaboration massivement distribuée et gestion du savoir en ligne: le cas de la communauté WordPress (2003-2008). "Code is poetry", but a "documentation resource should be managed". Humanities and Social Sciences. Ecole Polytechnique X, 2009. English. NNT : . pastel-00006078

HAL Id: pastel-00006078

https://pastel.hal.science/pastel-00006078

Submitted on 26 May 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Collaboration massivement distribuée et gestion du savoir en ligne: le cas du wiki de la communauté WordPress (2003-2008).

« Code is Poetry », but « a documentation resource should be managed ».

Thèse présentée pour obtenir le grade de
DOCTEUR DE L'ECOLE POLYTECHNIQUE

Domaine
ECONOMIE ET SCIENCES SOCIALES

Spécialité
GESTION

Soutenue publiquement le 4 décembre 2009 par

Emmanuel RUZÉ

Devant un jury composé de :

Pierre-Jean BENGHOZI	Directeur de recherche au CNRS, *directeur de thèse.*
Paul BOURGINE	Directeur de recherche au CNRS, *co-directeur de thèse.*
Jean-Michel DALLE	Professeur à l'Université Pierre et Marie-Curie, *rapporteur.*
François-Xavier DE VAUJANY	Professeur à l'Université Paris-Dauphine, *rapporteur.*
Anca METIU	Professeur à l'ESSEC.
Nicolas AURAY	Maître de conférences à Telecom ParisTech.

L’Ecole polytechnique n’entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

REMERCIEMENTS

Cette thèse est d'abord dédiée aux membres de ma famille, à ma grande-tante Odile Huber décédée le 27 juillet 2009 et qui est certainement à l'origine de ma vocation d'enseignant, et naturellement à mes très chers parents qui ont beaucoup donné pour me soutenir pendant une dépression de près de deux ans.

Une fois n'est pas coutume, nous commençons par remercier *une communauté*, celle du CRG. Une thèse sur une communauté se prête à ce genre d'écarts. Nous avons l'intention de continuer à y travailler comme chercheur associé.

Nous remercions Pierre-Jean Benghozi notre directeur de thèse, pour le temps qu'il a passé à nous encadrer et ses relectures, et Paul Bourgine, notre co-directeur de thèse, qui nous a aussi aidé à mettre en forme nos idées et a été une source d'inspiration essentielle sur le plan scientifique.

Nous remercions sincèrement tous les membres du jury : Jean-Michel Dalle, entre autres pour ses nombreuses remarques et discussions fructueuses en particulier lors de la pré-soutenance, François-Xavier de Vaujany, dont l'exemple aura été décisif, et enfin Anca Metiu et Nicolas Auray qui ont été aussi source d'inspiration par leur connaissance des communautés open-source. Je les remercie d'avoir consacré du temps à lire ce travail.

Nous remercions parmi les chercheurs du CRG pour leurs conseils et/ou leurs relectures fructueux : Pascal Sindonino, Matthias Den Besten, Hervé Dumez, Régine Teulier, et Alain Jeunemaître.

Parmi les doctorants, nous nous permettrons une pensée particulière pour, sans ordre particulier : « Nicki » Mirc, Colette Depeyre, Lise Gastaldi, Anne-Hélène Vergote, Emmanuelle Rigault, Sylvain Bureau, Jean-Baptiste Suquet, Jean Béhue.

Nous avons aussi une pensée très particulière pour Michèle Breton qui nous a beaucoup soutenu. Mes sincères remerciements vont aussi à Caroline Matthieu qui a relu l'intégralité du manuscrit et aidé à la mise en forme.

Au CREA, nos remerciements vont en particulier à Jean-Philippe Cointet et Carla Taramasco, entre autres pour leur aide en matière d'analyse de réseau.

A l'Ecole polytechnique, nous remercions vivement Michel Rosso, l'école doctorale et les membres successifs de X-Doc.

Nous désirons remercier toute la communauté scientifique qui gravite autour du Master 2 « économie des institutions », plus particulièrement Olivier Favereau, mais aussi plus particulièrement, à Nanterre, Bruno Chaves, avec qui nous avons eu de longues discussions sur le logiciel libre et qui nous a orienté vers la communauté WordPress.

Nous remercions aussi certains de nos élèves, avec lesquels nous avons pu nouer des liens d'amitié, de leurs encouragements.

Enfin, nous terminons par un hommage aux contributeurs bénévoles de la communauté WordPress.

Sommaire de la thèse

Sommaire de la thèse 5
Introduction 13
Chapitre 1 : présentation du sujet, problématisation, et mise en perspective de notre travail 19
1.1. Introduction : 19
1.2. Propositions de mise en perspective générale de notre thèse : d'un aspect historique essentiel du déploiement du numérique à une problématisation en sciences de gestion 20
1.2.1. Mise en perspective historique de nos orientations de recherche 20
1.2.2. Vers l'étude des wikis, élément du « Web 2.0» émergent. 22
1.2.3. Les frontières d'une approche en gestion 26
1.3. Que sait-on des formes de gouvernance dans les communautés open-source et les communautés de contenu ouvert (« open-media ») ? 30
1.3.1. La gouvernance des communautés open-source mériterait des approfondissements sur de nombreux points 31
1.3.2. La gouvernance des wikis : une dimension indispensable à leur viabilité, mais insuffisamment étudiée 33
1.4. La « collaboration massivement distribuée » et les communautés open-source : quelles articulations ? 34
1.4.1. Le problème de l'insuffisance de l'analyse et de la conceptualisation de la « collaboration massivement distribuée » 34
1.4.2. « Collaboration massivement distribuée » et « open-source », quelques clarifications 36
1.5. Progression et conséquences de la constitution de notre problématique 39
1.5.1. Retour sur le concept de gouvernance 39
1.5.2. Les conséquences associées au choix de notre problématique et de notre terrain pour comprendre la gouvernance des communautés reposant sur des wikis. 41
1.5.3. Présentation détaillée du choix des sous-questions de recherche et de la structuration de notre travail. 42
1.6. Nos orientations théoriques 45
1.6.1. Une approche économique est compatible avec un travail en management des systèmes d'information 45
1.6.2. Des théories économiques employées parfois en système d'information à une orientation en économie cognitive 46
1.6.3. Spécification des éléments théoriques choisis 48
1.7. Ouverture vers la suite de la thèse 52
Chapitre 2 : les limites de l'analyse par la littérature de la gouvernance des wikis dans les communautés open-source 55
2.1. Introduction 55
2.2. Orientations de la revue de littératures 56
2.3. Pourquoi une telle limite dont nous cernons ici les contours ? 58
2.3.1. Une littérature pour l'essentiel centrée sur Wikipedia 58
2.3.2. Le terrain primordial des communautés open-source a été négligé 59
2.3.3. La question des tâches humbles dans les communautés open-source a été négligée, ce qui permet de donner une perspective plus générale à ces limites de la littérature 63
2.3.4. Le cas général de la question du document numérique mériterait aussi de plus amples développements 64

2.4. Les limites méthodologiques et techniques de la littérature 65
2.4.1. De nombreux travaux déroutants 65
2.4.2. Des résultats lacunaires en sciences de gestion, ou en matière de questionnements gestionnaires dans la littérature qui n'est pas étiquetée en tant que telle 66
2.4.3. Les quelques thèses émergentes 67
2.4.4. Autres limites méthodologiques en matière d'étude de cas, d'analyse qualitatives, d'études longitudinales, et d'articulation entre approche empirique et outils théoriques 68
2.5. La littérature sur la gouvernance des wikis 69
2.5.1. Les routines : une approche exploratoire centrée encore sur Wikipedia 70
2.5.2. Les questions démographiques et de motivations des contributeurs 71
2.5.3. L'analyse quantitative des formes de collaboration au sens large et de la structuration du fonctionnement des wikis 74
2.5.4. Appropriation, usage et efficacité d'un wiki 77
2.6. Conclusions 79
Chapitre 3 : une approche historique des archives d'une communauté 81
3.1. Introduction 81
3.2. Les fondements du choix de notre étude de cas 82
3.2.1. Justification du choix de l'étude de cas 82
3.2.2. Justification du choix du terrain, la communauté WordPress 83
3.2.3. Justification et conséquence du choix d'un cas unique avec plusieurs items, sélection des unités d'analyse et du découpage temporel. 86
3.3. « Rendre à César ce qui est à César », l'intérêt d'une approche essentiellement historique 87
3.3.1. Les limites des autres approches, pourtant envisagées, et leur utilité à la marge .. 88
3.3.1.1. Apports à la marge et limites de « l'ethnographie virtuelle » 88
3.3.1.2. Apports à la marge et limites de la « grounded theory » 90
3.3.1.3. Quelques précisions sur l'absence d'interviews ou de questionnaires électroniques dans le cadre de cette thèse 91
3.3.2. Au-delà de l'analyse longitudinale, la nature des données appelle une analyse de type historique 91
3.3.3. Les approches historiques en sciences de gestion et en management des systèmes d'information sont appréciées, mais des travaux supplémentaires sont bienvenus. 93
3.3.4. L'analyse limitée par les historiens de l'avènement du numérique néglige la question des archives 94
3.3.5. Une tentative de micro-histoire dans un sujet d'histoire économique « immédiate » où sont traitées des problématiques de gestion 96
3.4. Présentation, critique des sources et formes de triangulations 97
3.4.1. Nos choix des sources parmi celles qui étaient disponibles 98
3.4.2. Présentation des sources sélectionnées et contextualisation 99
3.4.2.1. Le site de référence « Web archives » est indispensable mais insuffisant pour une analyse fine, que ce soit en micro-histoire ou en gestion. 99
3.4.2.2. L'entrepôt principal des listes de discussions 101
3.4.2.3. Comment analyser de façon satisfaisante un wiki ? 102
3.4.2.4. Les listes de discussions des programmateurs 104
3.4.2.5. La liste de discussions de la documentation du wiki 107
3.4.2.6. Les archives de chat IRC 107
3.4.3. Problèmes généraux de critique des sources et biais détectés 109
3.4.3.1. Biais matériels 109
3.4.3.2. Biais sémantiques 111
3.4.4. Formes et limites des triangulations 113

3.4.4.1. Le croisement des sources ... 113
3.4.4.2. La triangulation des méthodes d'analyse ... 114
3.5. Des déceptions récurrentes en matière d'outils informatiques de traitement de données ... 115
3.6. Synthèse des orientations méthodologiques ... 116
3.6.1. Retour sur les présupposés épistémologiques de cette thèse ... 116
3.6.2. Tableau de synthèse des usages des sources ... 117
3.7. Conclusions ... 118
Chapitre 4 : contextualisation du cas par quelques éléments d'histoire de la communauté ... 121
4.1. Introduction ... 121
4.2. Présentation de la communauté, du produit (le blog), et de la start-up « Automattic ». ... 122
4.2.1. De la communauté WordPress au succès de son produit ... 122
4.2.2. Le système d'information de la communauté WordPress ... 125
4.2.3. La start-up ... 126
4.3. L'histoire de la documentation est celle de l'acceptation progressive du wiki ... 129
4.3.1. Remarques méthodologiques préalables ... 129
4.3.2. Une première tentative de documentation au moyen d'un wiki ... 130
4.3.3 Le lancement du deuxième wiki ouvre une période de croissance forte et d'organisation adhocratique. ... 135
4.3.4. Une phase importante de documentation managée par Lorelle Van Fossen, future administratrice, contribue largement à aboutir à une documentation plus exhaustive... 139
4.3.5. On observe à partir de janvier 2006 remises en question et mise en quasi-sommeil de la régulation de la documentation pendant presque une année ... 146
4.3.6. On observe une des reprises périodiques de l'activité de coordination à partir de décembre 2006, caractérisée des évolutions notables mais aussi des formes de permanences ... 148
4.3.7. Propositions de synthèse par des repères chronologiques ... 152
4.4. Evolution des interventions des administrateurs : ... 153
4.5. Rythmes des activités et mise en évidence de dynamiques à partir des séries chronologiques ... 156
4.5.1. L'activité de programmation ... 156
4.5.1.1. Le rythme de publication des nouvelles versions du logiciel ... 156
4.5.1.2. Les activités de programmation ... 157
4.5.2. La fréquentation de la communauté et le développement en conséquence des activités de support ... 160
4.5.2.1. La fréquentation de la communauté : ... 161
4.5.2.2. Le forum de discussions et l'activité marchande ... 162
4.5.3. L'activité sur le wiki ... 164
4.5.3.1. La liste de discussions de coordination de l'activité de documentation ... 165
4.5.3.2. L'évolution du nombre d'articles ... 166
4.5.3.3. La reconstitution du nombre d'éditions sur le wiki ... 168
4.5.4. Comparaison des rythmes des différentes activités de la communauté et mise en évidence de la singularité de la documentation. ... 170
4.6. Synthèse, conclusions, limites de notre approche et ouverture vers les autres chapitres d'approfondissements ... 172
Chapitre 5 : « just my 2 cents » ? Analyse des contributions des participants au Codex et de leurs motivations. ... 177
5.1. Introduction ... 177

5.2. Eléments de méthodologie 178
5.3. La documentation est une activité laborieuse et guère reconnue 179
5.4. Le problème d'incitation du premier ordre a été résolu par la communauté, mais d'autres problèmes de création du bien collectif se posent 181
5.5. Eléments quantitatifs sur la structuration des contributions au Codex 184
5.5.1. La structuration des articles 185
5.5.2. La structuration des contributions 186
5.5.3. La structuration temporelle des contributions 188
5.5.4. La structuration des discussions 190
5.5.5. Comparaison entre les éditions de contribution et les éditions de discussions 192
5.5.6. Eléments sur la structuration du bien collectif à partir d'un échantillon d'articles 193
5.6. Qui sont les contributeurs ? 196
5.7. Les motivations des contributeurs d'après les données qualitatives 197
5.7.1. Du désir d'aider à une pratique habituelle de don contre don 198
5.7.2. De l'amusement à l'addiction 200
5.7.3. Purisme et professionnalisme: le désir du travail bien fait pour lui-même 201
5.7.4. Quelques contributions "au passage" 203
5.7.5. Contribuer à un projet qui fascine ? 203
5.7.6. Quelques éléments de motivation intéressée 204
5.8. Conclusions, limites, et recherches ultérieures 205
Chapitre 6 : l'émergence de routines hybrides 209
6.1. Introduction 209
6.2. Comment étudier les routines sur notre terrain si particulier ? 210
6.2.1. Justification de l'usage du concept de routine 210
6.2.1.1. Une thématique a priori paradoxale, d'où quelques exemples concrets 210
6.2.1.2. Définition du concept de routine parmi les définitions possibles et articulation avec d'autres concepts 211
6.2.2. Spécification des concepts théoriques mobilisés pour comprendre les routines dans ce contexte 212
6.3. Conception de la sous-question de recherche 214
6.3.1. Sélection des sources et précision des biais : 214
6.3.2. Spécification du codage 215
6.3.3. Ordre des opérations pour l'analyse et l'établissement des routines 216
6.4. Deux aspects institutionnels préalables 217
6.4.1. Le choix de la licence confirme la légitimité significative de la logique du code dans la communauté de la documentation. 217
6.4.2. Davantage qu'un problème d'incitation, on remarque l'émergence de la nécessité de la gouvernance suite aux problèmes d'organisation du premier wiki 219
6.5. Les routines de structuration du contenu 222
6.5.1. Les routines de la « gouvernante » (« housekeeping », i.e. le contrôle continu de la qualité de la documentation par des tâches ingrates) 222
6.5.2. La standardisation des pratiques de rédaction : de quelle façon formate-t-on le contenu sur le wiki ? 224
6.5.3. Le contrôle de l'architecture du Codex 227
6.5.4. Les brouillons sur page utilisateurs limitent l'application de la routine « edit mercilessly » 230
6.5.5. Les limites de la parabole de la « soupe aux pierres » est qu'il faut choisir les ingrédients : que met-on de façon routinière dans le Codex ? 237
6.6. Les routines de gouvernance de la population de contributeurs au Codex 241

6.6.1. « Massivement » ? Quel est le degré d'ouverture du wiki? 241
6.6.2. La routinisation des outils de coordination de l'activité : quelle coexistence ? ... 243
6.6.3. « No one bites here » : les routines de résolution des conflits sont assez marginales 249
6.6.4. Les routines de supervision, entre autorité et responsabilisation : l'autonomisation de l'équipe de documentation 252
6.6.5. Une routine de remerciement emphatique suffit à traiter la question de la motivation dans l'équipe de documentation..... 259
6.7. Synthèse : les liens entre routines 261
6.7.1. Synthèse concernant les cités mentionnées 263
6.7.2. Schéma de synthèse des routines 265
6.8. Conclusions 266
Chapitre 7 : analyse des formes de collaboration permettant la gouvernance du Codex..... 269
7.1. Introduction 269
7.2. La constitution d'une base de données à partir du dépouillement des archives..... 271
7.3. Caractéristiques démographiques de la population des discutants à la liste de discussions « wp-Docs » 271
7.3.1. Statistiques descriptives 272
7.3.2. Formes d'encadrement 273
7.4. Les caractéristiques en coupe (« cross-section ») de la population qui intervient sur la liste de discussion :..... 274
7.4.1. Statistiques descriptives 275
7.4.2. Des formes de dissymétrie à l'intérieur des historiques d'interventions des membres de la liste de discussion. 276
7.4.3. Quelques corrélations significatives..... 277
7.5. La structure dynamique des interventions sur la liste de discussions 278
7.5.1. Statistiques descriptives 279
7.5.2. Les caractéristiques du noyau central des intervenants..... 284
7.5.3. Les caractéristiques mensuelles de dispersion et de dissymétrie 287
7.5.4. Dispersions et dissymétrie dans les interventions sur la liste..... 288
7.5.5. Relation entre démographie et dispersion des interventions..... 289
7.6. Comment comprendre les interactions internes aux fils de discussions ?..... 291
7.6.1. Statistiques descriptives de l'initiation de fils de discussions et premières approches..... 291
7.6.2. Mise en évidence d'un effet structurel d'homophilie..... 298
7.6.3. Interprétation et implications 300
7.7. Une approche de la structuration de la liste de discussions à partir de l'analyse de réseau (Social Network analysis) 300
7.7.1. Le processus de traitement des données..... 301
7.7.2. Visualisations et remarques exploratoires..... 302
7.7.3. Analyse numérique du réseau 304
7.8. Conclusions 307
7.9. Annexe du chapitre 7..... 310
7.9.1. Structuration de la base de données 310
7.9.2. Autres explications et tentatives de réfutation du phénomène d'homophilie mis en évidence dans la partie 7.6. 311
Chapitre 8 : appropriation, usages et efficacité pour les parties programmation et aide aux utilisateurs dans la communauté..... 315
8.1. Introduction 315

8.2. Conception de la sous-question de recherche: 316
8.2.1. L'origine de la sous-question de recherche, une approche historique « opportuniste » 316
8.2.2. Choix des parties de la communauté étudiée et sélection des sources associées. 317
8.2.3. Extraction des données et saturation des références possibles au wiki 318
8.2.4. Un codage dans une visée d'analyse historique sans faire usage de CAQDAS . 321
8.3. Résultats concernant le forum d'aide aux utilisateurs 322
8.3.1. Eléments sur les résultats quantitatifs du codage 322
8.3.1.1. Statistiques de qualité du codage 322
8.3.1.2. Apports des données quantitatives des fragments extraits et leurs limites. .. 323
8.3.2. Vers une complémentarité entre Codex et forum 325
8.3.2.1. Un système d'information qui ne fait significativement pas consensus 325
8.3.2.2. D'une politique ambitieuse, la communauté, face aux difficultés, est passée à des objectifs plus réalistes de complémentarité entre SI. 327
8.3.2.3. La complémentarité entre les deux SI suppose une forme de régulation par une population commune de volontaires. 331
8.3.3. Dans quelle mesure peut-on qualifier la qualité de la documentation pour les utilisateurs et contributeurs du forum ? 334
8.3.3.1. Un outil comportant des insuffisances significatives mais non rédhibitoires, et perçu comme orienté « expert ». 334
8.3.3.2. Efficience du Codex, une centralisation nuancée 337
8.3.4. Mécanismes favorisant ou limitant l'usage du Codex 337
8.3.4.1. L'usage du Codex suppose l'émergence de pratiques d'intermédiation et d'orientation des utilisateurs 337
8.3.4.2. La valorisation de l'expertise et des conseils donnés aurait un effet limitant les citations, et donc l'usage de la documentation. 340
8.4. Résultats pour les informaticiens : efficience de la documentation, qualité perçue, les mécanismes observés 342
8.4.1. Aspects quantitatifs 342
8.4.1.1. Statistiques de codage des listes de discussion des informaticiens. 342
8.4.1.2. Une permanence de l'usage du Codex d'après les statistiques des fragments 343
8.4.2. Aspects qualitatifs : dans quelle mesure observe-t-on que le Codex est trouvé satisfaisant ? 344
8.4.2.1. Usages et efficacité sur la liste de discussion « Trac » 344
8.4.2.2. Usages et efficacité sur la liste des testeurs 346
8.4.2.3. Usages et efficacité sur la liste des hackers 348
8.5. Comparaison, discussion des résultats, conclusions, limites et autres orientations de travail 354
8.6. Annexes du chapitre 8 : 357
8.6.1. Annexe 8-1 : visualisation du processus de codage 357
8.6.2. Annexe 8-2 : courriels longs importants 358
Conclusion générale 361
Apports descriptifs, compréhensifs et explicatifs 361
Apport managériaux pour les contributeurs aux communautés, ceux qui sont amenés à les gérer, voire au-delà 363
Apports méthodologiques et théoriques 365
Limites méthodologiques, théoriques et recherches futures envisageables 366
Annexes générales 373
Annexe 1 : liens vers les principales archives en ligne sélectionnées 373

Annexe 2 : autres liens menant vers la communauté WordPress 373
Annexe 3 : glossaire de vocabulaire informatique 374
Annexe 4 : script du programme d'extraction (« parsing ») en langage python écrit par Jean-Philippe Cointet 379
Annexe 5 : autres indicateurs d'analyse de réseau 381
Annexe 6 : description précise des interactions entre intervenants dans le réseau de la liste de discussions de la documentation 384
Annexe 7: "codebook" correspondant à la liste de discussion des hackers 389
Annexe 8: "codebook" du codage de la liste de discussion du forum 390
Annexe 9 : tableau de quelques régressions du chapitre 7 391
Régression de la partie 7.6.2. 391
Régression de la partie 7.3.2. 393
Régression de la partie 7.5.1. (figure 62) 395
Régression de la partie 7.6.2. (figure 77) 397
Régression de la partie 7.6.1. (figure 82) 399
Régression de la partie 7.6.2. (figure 73) 401
Bibliographie 403
Ouvrages et articles généraux, en particulier en management des systèmes d'information et théories des organisations 403
Ouvrages et articles de méthodologie, et théories mobilisées 404
Ouvrages et articles traitant d'histoire, d'enquête, méthodes qualitatives et quantitatives 404
Articles et ouvrages concernant les théories mobilisées 406
Ouvrages et articles sur les wikis 409
Articles tirés de revues à comité de lecture, ou chapitres d'ouvrages. 409
« Working papers » ayant subi le test d'une conférence ou d'un colloque. 412
Autres travaux pertinents ou autres articles sans comité de lecture (Working paper non soumis à conférences, mémoires, etc.) 418
Ouvrages et articles sur l'open-source, les communautés virtuelles, l'économie du logiciel 419
Articles de presse 428
Liste des figures 429
Liste des tableaux 432

Introduction

Le travail décentralisé individuel et collectif à l'extérieur des murs des organisations voire au-delà de leurs frontières est un élément essentiel des systèmes économiques contemporains. Il sera probablement davantage crucial à l'avenir étant donné que les coûts de déplacements des individus augmenteront suite à la croissance prévisible des coûts de l'énergie. La montée en puissance de l'interconnexion entre les organisations et les individus, des formes de réseaux entre ordinateurs, de l'échange de données informatisées, du télétravail, et enfin l'émergence de la numérisation des contenus ne sont que quelques éléments essentiels qui ont contribué à l'émergence du phénomène désormais à une grande échelle. Elle ne concerne pas que les entreprises à but lucratif, mais aussi les communautés de chercheurs, les associations, voire les institutions publiques.

Certaines jeunes disciplines en sciences sociales ont même vu le jour afin d'étudier entre autres ces formes de communication et de collaboration synchrones ou asynchrones impliquant des ordinateurs. Leurs contours sont encore mal définis, elles comportent dans leur intitulé par exemple les mots « computer-mediated-communication » ou « systèmes d'information ».

L'une des concrétisations récentes de ce phénomène général est qualifiée avec le recul de « déploiement du numérique » par les historiens du numéro 43 d'*Entreprises et histoire* (2006). Des systèmes d'information spécifiques reposant sur des outils adaptés comme les forums, les listes de discussions, etc. ont émergé lors de cette histoire. Au-delà des technologies employées, des processus de création individuelle ou collective de contenus par les internautes, appelés « user-generated content » et souvent « Web 2.0. » sont un objet d'étude de premier plan pour comprendre nos sociétés et économies contemporaines marquées par de récentes évolutions technologiques rapides qualifiées de « nouvelle économie » (OCDE, 2007). Sur le plan organisationnel, cette création de contenu suit, dans une certaine mesure, l'exemple communautaire emblématique des communautés d'informaticiens qui conçoivent des logiciels libres et dont les principes de collaboration seraient adaptés ailleurs.

Cet aspect communautaire du « déploiement du numérique » appelle plusieurs questions en sciences de gestion. En effet, au-delà des seuls aspects technologiques qui sont insuffisants en management des systèmes d'information, on peut se demander comment les

participants à de telles communautés produisent, interagissent, voire s'organisent. Plus précisément, quels sont les processus de production, les formes de collaboration, les motivations, les modes d'appropriation du contenu et d'insertion des contributeurs au contenu collectif dans une communauté dans son ensemble? Dans quelle mesure observe-t-on des formes de gouvernance de ces communautés?

Dans ces communautés, un outil entre autres a émergé depuis 1994 pour faciliter les contributions décentralisées dans des proportions potentiellement considérables: le wiki (CUNNINGHAM, 2001). Un nombre conséquent de communautés en ligne très diverses et de taille potentiellement gigantesque ont adopté cet outil : celles des bénévoles de catastrophes naturelles, des encyclopédies collaboratives, des communautés de scientifiques (biologistes…), des passionnés de jeux vidéos ou de produits de la « culture de masse », mais aussi des communautés open-source, etc.. L'importance du phénomène n'a pas encore engendré une recherche exhaustive, quantitative et historique concernant les communautés et organisations les utilisant[1], en particulier les communautés open-source.

Notre présente thèse propose, afin d'aborder les thématiques précédentes, une étude en profondeur et historique d'une communauté en ligne qui a implémenté un wiki pour capitaliser ses connaissances (« repository of knowledge ») par l'apport collectif de contenu. Notre travail porte donc sur les usages des wikis dans les communautés open-source, un terrain pertinent comportant des spécificités organisationnelles. Ces communautés sont essentielles dans l'histoire de la Toile, et les wikis sont largement répandus dans ces communautés, mais leur étude a été négligée sur ce terrain précis par la recherche en sciences sociales, alors qu'il s'agit d'un usage historiquement premier.

Notre problématique de thèse qui précise les questionnements *organisationnels* précédents peut se formuler de la façon suivante :

> Nous nous demandons de quelles façons les communautés open-source en particulier gèrent leur savoir dans un système d'information comportant un wiki.

[1] Pour une tentative de dénombrement elle-même collaborative et à manier *avec les plus grandes précautions*, voir : http://meta.wikimedia.org/w/index.php?title=List_of_largest_wikis (consulté le 19/09/09). Par ailleurs, beaucoup de wikis sont utilisés à l'intérieur d'intranets, ils n'entrent pas dans notre propos bien qu'il s'agisse d'un objet de recherche légitime.

Plus précisément, dans quelle mesure la « collaboration massivement distribuée » sur un support wiki caractérise-t-elle ces processus dans ce contexte?

Notre thèse permet de comprendre, à partir du cas WordPress, pourquoi dans les communautés open-source, un terrain où les wikis ont une importance cruciale et historiquement première, l'auto-organisation de telles communautés conduit à l'émergence de formes d'adaptations qui relèvent de la gouvernance d'un système d'information[2].

Nous soutiendrons cette thèse dans plusieurs sous-questions de recherche en étudiant l'émergence de *routines*, de formes de *collaboration* plus centralisées, *la gestion de l'articulation* entre le système d'information wiki et le reste de la communauté (rythmes, appropriation du savoir en ligne), l'émergence des contributions au système d'information et la pluralité des motivations observées.

Au-delà de l'outil informatique, la « collaboration massivement distribuée » est un concept qui sert à désigner l'organisation informelle de Wikipedia, qui fait l'objet de recherches conséquentes. Le principe de l'encyclopédie avait été inspiré par le mouvement open-source. Cependant, cette thèse souligne que les logiques de fonctionnement de l'open-source et de l'open-media seraient significativement différentes. Le point d'articulation entre les deux phénomènes organisationnels éminemment *interdépendants* dans notre contexte est pertinent à étudier. Ils sont souvent confondus alors qu'ils conduisent à la production de biens différents (code et contenu diversifié en ligne).

La mise en œuvre de cette thèse comporte une spécificité, en particulier en management des systèmes d'information. La conception de cette étude de cas relève de l'analyse longitudinale et plus précisément d'une approche de type *historique* : sa spécificité réside dans l'analyse et la prise en compte des données imparfaites auxquelles nous avons été confrontés, comme des traces sur une période finie, ou, au contraire de données en quantité démesurée. Ce travail mobilise aussi bien les approches qualitatives - dont l'analyse de traces -, quantitatives - dont l'analyse de réseaux -, compatibles avec une approche de type

[2] Nous le précisons au besoin, il n'y a pas opposition entre « auto-organisation » et « gouvernance », mais spécification de la première par la compréhension du phénomène émergent caractérisant la seconde.

historique. Nous ferons appel largement à la triangulation des données ou mobilisation discutée de sources multiples.

Nous avons choisi de traiter cette question par une étude de cas approfondie, celle du wiki nommé « Codex » de la communauté WordPress. Cet outil supporte le système d'information central servant à rendre accessible l'ensemble des savoirs concernant le logiciel produit par la communauté.

Cette communauté open-source a en effet conçu et diffusé avec succès de 2003 à 2009 un moteur de blog haut-de-gamme devenu un outil de référence pour les connaisseurs de la publication en ligne. Les blogs sont l'un des vecteurs majeurs contemporain de la diffusion d'information décentralisée et souvent indépendante. L'étude des conditions de viabilité de cette « nouvelle imprimerie » numérique permettant la diffusion d'informations à grande échelle est susceptible de piquer la curiosité. Nous rappelons aussi que plus de la moitié des sites internet sont des blogs. Une start-up proposant des services payants à partir du logiciel est née en 2005, mais elle n'aurait pu émerger sans une communauté préalable, ce qui montre l'importance de celles-ci pour comprendre les « modèles d'affaires » de « l'économie numérique ».

L'angle d'analyse théorique choisi pour ce travail en gestion est économique mais adapté à l'examen de problématiques de gestion et cohérent avec l'approche historique : nous mobilisons des outils tirés de l'économie cognitive au sens large (évolutionniste qualitative et quantitative, analyse des réseaux, économie des conventions, économie de la connaissance…) pour *comprendre les processus de gouvernance* observés dans quatre sous-questions de recherche présentées ci-dessous. L'articulation entre économie hétérodoxe préoccupée par les questions de théories des organisations et sciences de gestion est une préoccupation significative des chercheurs de ce courant pluraliste. Sur le plan institutionnel, notre travail est, rappelons-le, une collaboration entre deux disciplines, les sciences de gestion et les sciences cognitives, et concrètement entre le Centre de Recherche en Gestion et le Centre de Recherche en Epistémologie Appliquée de l'Ecole polytechnique.

Les questions de recherche développant les thématiques de thèse seront précédées par une histoire partielle et une contextualisation de notre communauté qui servira à les introduire. Nous résumerons nos contributions scientifiques en fin d'introduction, en présentant le plan de notre travail.

Sous-question de recherche 1 : *comment comprendre les contributions des participants au Codex et leurs motivations ?*

Sous-question de recherche 2 : *comment comprendre les routines qui ont émergé durant l'histoire de la conception de la documentation ?*

Sous-question de recherche 3 : *de quelle façon les contributeurs interagissent-ils dans la sous-partie de la communauté qui s'occupe du wiki ?*

Sous-question de recherche 4 : *comment comprendre l'appropriation du produit de la collaboration massivement distribuée ?*

Nous détaillons à présent l'articulation du plan de notre thèse. Nous avons conçu notre thèse en deux grandes parties. La première expose des savoirs préalables au développement des résultats dans trois chapitres traditionnels. La seconde, de cinq chapitres, présentera les résultats.

Le premier chapitre prolongera la présentation de notre sujet et notre problématisation par une mise en perspective historique, gestionnaire, et théorique.

Le deuxième chapitre est une revue de littérature sur les wikis, un bilan partiel qui est aussi un apport en soi. Il montrera dans quelle mesure les principaux points de notre sujet ont été négligés par la littérature antérieure.

Le troisième chapitre présentera la méthodologie de notre thèse. Il montrera qu'à partir des matériaux disponibles, traiter notre problématique suppose d'adopter une approche historique. Ce retour d'expérience à partir de notre matériau est un apport en soi, car il peut être utile à d'autres chercheurs travaillant sur le « déploiement du numérique » (RUZE, 2009).

Le quatrième chapitre, une transition, aidera à la contextualisation de notre communauté, permettra de dégager ses principaux rythmes et de cerner son histoire. Les tendances globales observées permettent déjà de dégager des enseignements mais aussi d'introduire les thèmes des sous-questions de recherche qui font l'objet des quatre derniers chapitres.

Le cinquième chapitre traitera la première sous-question de recherche. Il examinera la formation du bien collectif informationnel qu'est le wiki de la communauté, en particulier la structuration des contributions par un noyau de volontaires permanent, les motivations diverses des contributeurs et le caractère satisfaisant de résolution du paradoxe économique de bien public (« public good »).

Le sixième chapitre traitera la deuxième sous-question de recherche. Il permettra de mieux comprendre l'émergence et l'adaptation des routines, formes institutionnelles *informelles* de gouvernance des processus d'édition en ligne associées à la collaboration massivement distribuée. Il permet de comprendre l'émergence d'une forme d'hybridation entre les pratiques de l'open-source et de l'open-media.

Le septième chapitre traitera la troisième sous-question de recherche. Il s'agit d'une étude quantitative de la dynamique de mise en relation entre contributeurs et de la structure des éventuelles collaborations. Cet aspect fondamental de la gouvernance du wiki est mis en oeuvre sur une liste de discussions.

Le huitième chapitre traitera la quatrième sous-question de recherche. Nous la concevons comme une étude de l'appropriation du wiki par le reste de la communauté. L'effet du wiki sur ses membres, la qualité du contenu, sont des résultats, mais ils supposent des formes d'appropriation du produit. Nous montrerons en quoi on peut observer des spécificités pour les utilisateurs novices et les informaticiens, deux catégories de membres de la communauté WordPress.

Nous proposerons enfin dans une conclusion générale des recommandations managériales à destination des volontaires qui gèrent ces communautés, une mise en perspective de nos résultats, et enfin plusieurs pistes de recherches, déjà plus ou moins élaborées et qui découlent des limites de notre travail.

Chapitre 1 : présentation du sujet, problématisation, et mise en perspective de notre travail

« On parle de la disparition du livre : je crois que cela est impossible »
Jorge-Luis BORGES

« Il faut bien reconnaître que le livre n'est pas un objet particulièrement bien inventé : il attire la poussière, il se déglingue facilement, il est fragile et pas pratique, et ça en tient de la place dans une bibliothèque (...). Plus de livres ? Pourquoi pas ? Il y a bien eu des œuvres avant l'imprimerie, pourquoi n'y en aurait-il pas après ? »
Raymond QUENEAU[3]

1.1. Introduction

Nous développons ici la présentation de notre sujet de thèse à partir d'une tentative de mise en perspective historique générale, celle du texte numérique et du Web 2.0. Elle converge vers une problématisation en sciences de gestion, un choix qui demandait une explicitation. Nous prolongeons cette problématisation en montrant qu'elle implique un problème significatif en théorie des organisations, à l'articulation entre modes d'organisation « open-source » et « open-media » : si le second est une adaptation ultérieure des principes du premier, ces deux modes d'organisation diffèrent, ce qui implique des problèmes de gestion lorsqu'ils s'articulent dans une communauté.

Nous présentons enfin les orientations théoriques de notre thèse, qui consiste en l'usage des outils de l'économie cognitive sur plusieurs thématiques, les sous-questions de recherche. Nous justifierons également, bien que cela ait déjà été dégagé partiellement, l'articulation en sous-questions de recherche au service d'un axe de questionnement principal.

Ce chapitre sera donc allusif sur certains points que nous évoquerons longuement plus loin (présentation de la communauté, choix du terrain, méthodologie, perspective historique sur notre cas).

[3] Nous avons repris des citations d'un rapport du Sénat : elles nous semblaient appropriées (http://www.senat.fr/rap/r06-468/r06-4680.html, consulté le 10/10/09)

1.2. Propositions de mise en perspective générale de notre thèse : d'un aspect historique essentiel du déploiement du numérique à une problématisation en sciences de gestion

Nous proposons de partir d'une mise en perspective générale de notre thèse pour arriver, en concentrant et ciblant nos propos, à une problématisation en gestion.

1.2.1. Mise en perspective historique de nos orientations de recherche

La question des modes de formation des textes et de leur utilisation est une question historique de première importance et une matière à interrogations incontournable en systèmes d'information[4]. En la matière, nous ne saurions épuiser la question, nous n'en proposons que quelques aspects. WAHLSTROM et SRUTON (1997) proposent une mise en perspective des changements historiques semblables à ceux de l'Internet, par exemple le passage du Codex[5] au livre imprimé. CHARTIER (2006) dans le numéro spécial d'*Entreprises et Histoire* sur le « déploiement du numérique » met en perspective les changements dus à cette dynamique, et met en évidence les *trois grandes mutations* de la nature des textes qui caractérisent l'ordre numérique : le brouillage des genres, la lecture discontinue, et surtout l'intervention du lecteur et de l'effacement de l'auteur. Il évoque en conséquence la concurrence entre livre sur support papier et livre en ligne. Cet « horizon livresque » remis en question par d'autres dispositifs plus efficaces est consubstantiel à notre problématique de gestion. Notre présent sujet portant sur les wikis est emblématique de cette thématique de transition et de ces transformations.

Le passage surprenant suivant qui anticipe dans une certaine mesure ce que sera Internet date du *Traité de documentation* (1934), l'ouvrage de base de la documentation moderne écrit par Paul Otlet, visionnaire méconnu, inventeur de la classification décimale universelle et du « Mundaneum », une utopie qui consistait à classer en un seul lieu tout le savoir humain :

> « Ici, la table de travail ne serait plus chargée d'aucun livre. A leur place, se dresse un écran et un téléphone. Là-bas, au loin, dans un édifice immense, sont tous les livres et tous les renseignements… De là, on fait apparaître sur l'écran la page à lire pour connaître la réponse aux questions posées par

[4] Il s'agit d'une discipline qui, il est bon de le rappeler, n'est aucunement liée nécessairement à des supports comme l'ordinateur et aux technologies même si c'est le plus souvent le cas. On pourrait tout à fait par exemple effectuer une thèse en systèmes d'information sur la production et la circulation des manuscrits entre abbayes cisterciennes au Moyen-Age.

[5] Il s'agit de l'ancêtre du livre, support de l'écrit de l'Empire romain à l'invention de l'imprimerie. Le Codex de la communauté est, lui, un wiki.

> téléphone avec ou sans fil. (…). Utopie aujourd'hui, parce qu'elle n'existe encore nulle part, mais elle pourrait bien devenir la réalité pourvu que se perfectionnent encore nos méthodes et notre instrumentation. »[6]

Un lecteur gestionnaire dirait qu'il aurait pu aussi préciser : « nos méthodes de gestion ». Ce passage décrit précisément ce que propose à des utilisateurs désireux d'obtenir des réponses à leurs questions au plus vite et sans efforts, WordPress, avec son Codex, comme tant d'autres communautés open-source, mais sur un serveur plus décentralisé que ce que proposait Otlet, qui n'évoque pas non plus les formes d'organisations sous-jacentes émergeant de façon décentralisée. De plus, un tel système reste en concurrence avec des livres papiers actuellement, ce qui n'était pas forcément attendu par Otlet, qui voyait les livres disparaître[7].

Cependant, la thématique que nous étudions ne se réduit pas à la nature textuelle, c'est d'ailleurs sur cette limite de la vision du réseau d'Otlet que nous poursuivons. La question générale de la nature de l'Internet est plus ambiguë. Qu'est ce qu'Internet ? Répondre à une telle question dépend du contexte et de la discipline choisie. Nous répondons en citant un expert du travail de terrain :

> « In the previous section, Internet was described as a site for interaction which, although it might not entail face to face communication, was still in some sense ethnographically available. This assumption is based on the argument that what goes on on the Internet *is social interactions*. Another way of looking however, would see the cyberspace *as composed of text, rather than being interactive*. There is no definitive line between those two concepts."[8]

C'est de cette idée concrète d'entre deux (qui caractérise le wiki), encore imprécise mais générale, que nous nous autorisons de partir : Internet est un lieu d'interactions sociales, mais aussi un lieu de production de contenu[9]. Or, ces deux dimensions générales de l'Internet soulèvent des difficultés spécifiques qui intéressent directement la recherche en gestion: quelle est la nature des relations sociales nécessaires à la production du savoir en ligne ? Peut-on les qualifier de collaboration ? En deçà, quels sont les modes de coordination efficaces ? Quelle est la qualité du contenu en ligne ? Comment le formater et le produire?

Internet est un Système d'Information dont les parties ayant des finalités productives supposent des processus d'organisation et des pratiques de management. En effet, Internet est

---

[6] Nous renvoyons à http://www.mundaneum.be/index.asp?ID=247 (consulté le 20/07/09) pour une biographie entre autres. Les éléments mentionnés ici viennent du ,n°186 de *Sciences humaines* (2007).

[7] Nous précisons au lecteur que la concurrence et la transition entre les deux supports, livresques et numériques est générale et comporte donc d'autres dimensions en système d'information qui complexifient la problématique, en particulier si la discipline prend en compte l'idée de développement durable : un livre papier peut être plus écologique qu'un livre électronique ou de multiples requêtes sur un moteur de recherche, pour ne parler que de cela. La concurrence entre les deux SI est complexe.

[8] HINE (2000), c'est nous qui soulignons.

[9] Ce contenu ne consiste pas seulement en « textes », contrairement à ce que suggère Hine, mais cela ne change rien à l'idée générale que nous proposons de suivre.

aussi un système productif, ce qui est en particulier le cas dans le secteur du logiciel, qui nous intéresse plus particulièrement : on sait que les éléments essentiels d'un produit aussi complexe qu'un logiciel sont conçus en ligne, et il ne s'agit pas seulement des processus de coordination de la production, mais aussi de la relation avec les utilisateurs et de leur intégration dans une communauté, en particulier le processus de documentation et l'usage des documents numériques. Les processus de conception sont à la fois « lien » et « texte ».

Le cas des wikis est pertinent pour étudier une telle frontière entre lien et texte, et les problèmes qui s'y posent : la nature du « texte » *et* la nature des « relation » sont tous les deux problématiques, étant donné que la spécificité des wikis est justement de laisser aux internautes la possibilité de modifier textes et liens, et que la mise en relation des individus n'est en fait pas seulement une médiation de type tableau noir (« blackboard »), pour reprendre un concept de GENSOLLEN (2005) qui décrit les interactions des participants via un artefact électronique ressemblant au tableau des salles de classes permettant l'ajout de texte de façon conséquente.

Le contexte général de ce qui nous intéresse est ainsi celui de *la nature évolutive de l'Internet à un moment donné de son histoire, et des problèmes de gestion (du contenu, des contributeurs qui forment des liens) qui s'y posent*.

1.2.2. Vers l'étude des wikis, élément du « Web 2.0. » émergent

L'évolution de l'Internet qu'est le Web 2.0 fait débat (O'REILLY, 2007) et mériterait d'autres travaux. Le Web 2.0 permettrait une implication nettement plus significative des utilisateurs sur les sites Internet. Une telle évolution ou, au contraire, un tel *retour* aux possibilités premières offertes par le numérique comporte, au sens large, des enjeux économiques (OCDE, 2007) au point que désormais des questions significatives de régulations se posent (RUZE, 2009) et que des approches gestionnaires soient justifiées. Nous relevons plusieurs secteurs mentionnés dans la littérature grise, la banque (HENG et al., 2007, ASHTA et ASSADI, 2008), les biens culturels (MABILLOT, 2007), et tout ce qui concerne l'innovation dans le domaine de l'Internet, ascendante ou pas, en particulier l'économie du logiciel (O'REILLY, 2007). Nous mentionnons aussi l'application de ses principes à l'entreprise elle-même en interne comme à ses frontières, i.e. au management en lui-même (KOUSHIK, 2009), ou des formes d'adaptation au management et secteur public (MERGEL et al., 2009). Cependant, tout le potentiel du Web 2.0 n'est pas encore pleinement exploité :

nous renvoyons par exemple à DUBOSSON et BROILLET (2008) pour le cas de la vente des montres de luxe ou à GINSBERG et al. (2009) pour le cas des librairies juridiques. Le lien encore problématique entre activité économique productrice et Web 2.0. est donc une question méritant réflexion. La recherche sur le Web 2.0. , évoquée ci-dessus, en est à un stade significativement exploratoire.

Par ailleurs, les nouvelles formes d'entreprises sur le Web et certaines de leurs composantes appelées « communautés » sont encore mal connues pour de nombreuses raisons, entre autres la collaboration difficile entre chercheurs en informatique et en sciences sociales[10]. Nous nous intéressons ici plus particulièrement aux communautés à composantes technologiques, et plus précisément celles qui relèvent de l'open-source, qui constituent un phénomène spécifique d'ampleur conséquente en histoire contemporaine[11]. Une « bonne » communauté est, dans un projet efficace, un élément stratégique fondamental qui permet de comprendre le succès d'un logiciel : en effet, cette communauté propose interaction et aide bénévole, ce qui rend confiant un futur « newbie » (l'utilisateur novice) plus prompt alors à installer l'outil en question. Nous ne parlons pas seulement des développeurs, mais de ceux qui assument d'autres tâches plus ingrates comme la constitution de la documentation sur laquelle repose l'aide en ligne, comme le montrera le chapitre 4. Ainsi, ces communautés sont d'une importance capitale pour comprendre l'entrepreneuriat et l'innovation dans le domaine des nouvelles technologies et ont un rôle économique remarquable.

Dans cette thèse, nous choisissons d'aborder plus particulièrement le sujet à partir d'un cas prometteur, une communauté de blogueurs du « Web 2 » dont la réputation chez les blogueurs en général est excellente (nous renvoyons sur ce point au chapitre 4) : celle qui supporte le moteur de blog open-source WordPress. La partie stratégique de la communauté qui nous intéresse est l'équipe chargée de la documentation, écrite sur un wiki nommé « Codex ». Le contexte organisationnel sous-jacent de la formation de cette documentation est celui de la « collaboration massivement distribuée » dont l'archétype serait Wikipédia. Dans quelle mesure cependant un tel ensemble de contributeurs hétérogènes, d'origine dispersée, etc. débouche-t-il sur l'émergence d'une communauté viable et efficace ?

Il est tout à fait cohérent d'étudier un wiki dans une communauté open-source pour étudier le phénomène. On peut penser a priori justement que notre terrain serait décalé par

[10] Nous pensons sur ce point à notre expérience personnelle dans le projet PERSI de 2004 à 2005.
[11] Nous renvoyons pour cela à la base de données du MIT qui centralise une part significative de la recherche sur la question : http://opensource.mit.edu/ (consulté le 10/07/09).

rapport à ce mouvement Web 2.0. . Le mouvement open-source est en effet indépendant de ce mouvement récent. Ici, un rappel historique est nécessaire : le premier wiki créé en 1994 était un outil d'informaticiens, le wiki C2 de Ward Cunningham[12], et ce n'est qu'ensuite que le principe du wiki a été adopté par Wikipedia pour créer une encyclopédie en ligne. De plus, l'un des produits essentiels du Web consiste en logiciels, un secteur qui emploie, on ne le sait pas toujours, autant que l'industrie automobile (HORN, 2004), pour comparer avec un secteur de taille. Etudier les usages des wikis dans un tel secteur est donc justifié.

Une proposition d'étude approfondie des wikis dans leur contexte original manquait au début de notre travail de thèse, malgré l'émergence récente de travaux sur cette thématique. Nous développerons ce point dans la revue de littérature.

Notre étude de cas sur la communauté WordPress est la première du genre, personne n'avait encore étudié ce terrain, alors que d'autres travaux sur l'open-source peuvent porter plusieurs fois sur un même terrain (par exemple Linux[13]). Pourtant, le segment des logiciel de gestion de contenu (« Content Management system », dit CMS), un type de logiciel répandu[14], n'a guère fait l'objet de travaux de recherche : le seul cas que nous connaissons est le cas du logiciel SPIP, étudié par DEMAZIERES et al. (2007). Ceci est en effet cohérent avec l'essentiel des travaux déjà nombreux sur le logiciel libre qui s'intéressent à des projets complexes de taille significative (Debian, Apache, etc.) ; nous présentons au contraire un projet de taille moyenne.

Pour cette thèse, nous ferons usage d'une définition minimale de « wiki » : un wiki est un site web dynamique où l'utilisateur peut modifier contenu du site et architecture des pages. C'est ce qui se dégage par exemple de deux définitions dans des travaux faisant autorité :

> "While a precise definition of Wiki does not exist (Lamb, 2004), the general consensus is that a Wiki is *a collective website where any participant is allowed to modify any page or create a newpage using her Web browser*." (DESILETS et PAQUET, 2005).
>
> "Wikis are a set of web-based software applications that allows visitors to edit and contribute to the content already on the online page. Users access the content from most web browsers, and no special software is required. Common wiki uses include brainstorming, developing frequently asked questions (FAQs) and responses, collaborative drafting of positions or statements, planning, scheduling, family histories, course materials and projects. Wikis are similar to but distinct from blogs, which may, though

[12] Le premier wiki de l'histoire du déploiement du numérique, le C2, a été créé en 1994 (source : http://c2.com/cgi/wiki?WelcomeVisitors).

[13] Voir par exemple sur ce point la base de données du MIT sur l'open-source : http://opensource.mit.edu/online_papers.php?lim=1000 (consulté le 22/07/09).

[14] Les logiciels de gestion de contenu, ou « content management system » (CMS) sont l'une des grandes catégories de logiciels proposées par Framasoft (source : http://www.framasoft.net/, consulté le 10/06/09).

do not always allow visitors to add comments and create an ongoing conversation. Wikis can be limited to doing this alone, but they also have other functions. For example, they can be configured, to allow users to change their previous postings: add, edit, delete, etc. Used in this way, wikis tend to be more collaborative than conversational, with content developed and owned communally by all participants." (CHALLBORN et REIMANN, 2005).

La question des frontières de l'objet ne va pas toujours de soi, car les fonctionnalités des outils sont multiples en fonction du type d'outil wiki considéré et de sa période de vie. Les logiciels wiki sont divers : cela va de l'outil le plus minimaliste à l'interface la plus complexe. Les rapports avec d'autres outils relationnels comme les blogs, les CVS, les courriels, ne sont pas immédiats non plus, car les fonctions des outils peuvent se recouper. Pour synthétiser, nous présentons une matrice datant de 1995 qui permet par une remise en perspective de montrer que les wikis comportent de fortes spécificités, tôt détectées :

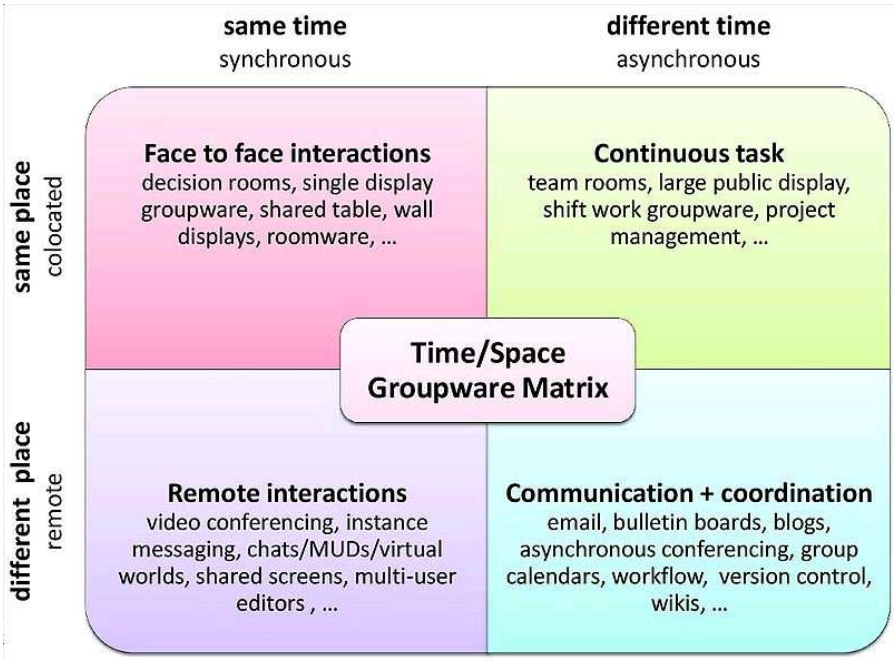


Figure 1 : Matrice de synthèse des outils de communication sur Internet (BAECKER et al. 1995)

Les fonctionnalités du wiki de la communauté et certaines de ses caractéristiques qu'il faut comprendre pour suivre notre travail de recherche seront présentées ultérieurement dans le chapitre 3.

1.2.3. Les frontières d'une approche en gestion

Ici, nous présentons en quoi notre sujet est une recherche en gestion[15] : nous inscrivons notre recherche en premier lieu dans certaines formes de tradition du CRG, ce que nous explicitons par un certain nombre d'éléments concrets, et enfin nous montrerons quelles sont les thématiques de gestion significatives qu'il recoupe : *l'étude des communautés* a trouvé une légitimité en management des systèmes d'information, ainsi que la thématique de la *gestion du savoir*, et, enfin, depuis peu, la *gouvernance* de ces systèmes. C'est en particulier à cette question que nous nous attaquerons, même si la notion de « gouvernance » est particulière dans notre contexte. Nous verrons dans le chapitre prochain, lors de la revue de littérature, en quoi notre travail est novateur de façon plus approfondie et systématique.

Jacques Girin avait tenté de construire une théorie de l'action située et cela l'avait amené à définir une situation de gestion. Etant donné que notre sujet est quelque peu décalé et que sa dimension gestionnaire ne va pas de soi, il est utile d'y revenir :

> « Une situation de gestion se présente lorsque les participants sont réunis et doivent accomplir en un temps déterminé une action collective conduisant à un jugement externe »[16].

Précisément, la formation d'un bien collectif de nature informationnelle comme une documentation logicielle, au moyen d'un TIC donné, sur une durée significative et pendant l'histoire complexe d'une communauté, au moyen de formes « d'actions organisantes » (restant à comprendre) sous contrainte du jugement d'utilisateurs divers et exigeants est une problématique de gestion. De plus, la formation d'un tel bien collectif est difficile, mal reconnue et se fait par bénévolat : c'est ce que nous verrons ultérieurement dans cette thèse. Rappeler cette définition est utile, car les présupposés de la « situation de gestion » sont ici imparfaits même si la définition reste pertinente: les participants sont *ponctuellement* réunis (les contributions sont potentiellement instables) et le « temps déterminé » n'est pas toujours défini clairement : le Codex suit ou tente de suivre l'activité de programmation ou l'accumulation du savoir imparfaitement codifié du forum, mais il n'y a pas de date limite de réalisation du travail (« deadline »), et le rythme du Codex peut être lent. C'est probablement cette « gestion à la limite » dans une « situation limite » qui peut même faire l'intérêt de notre recherche : la notion de « management » dans la communauté émerge progressivement.

Tentons à présent un développement de notre problématique à partir de la page d'accueil de la communauté, celle où les novices se rendent d'abord. Deux formes

[15] Cette partie est utile pour les lecteurs qui ne connaissent pas significativement ce champ de recherche. Pour les autres, elle aide à cerner les orientations de notre travail.
[16] GIRIN (1990).

d'accumulation du savoir sont dans une certaine mesure *substituables*, ce qui est un autre aspect essentiel de la question de l'efficacité :

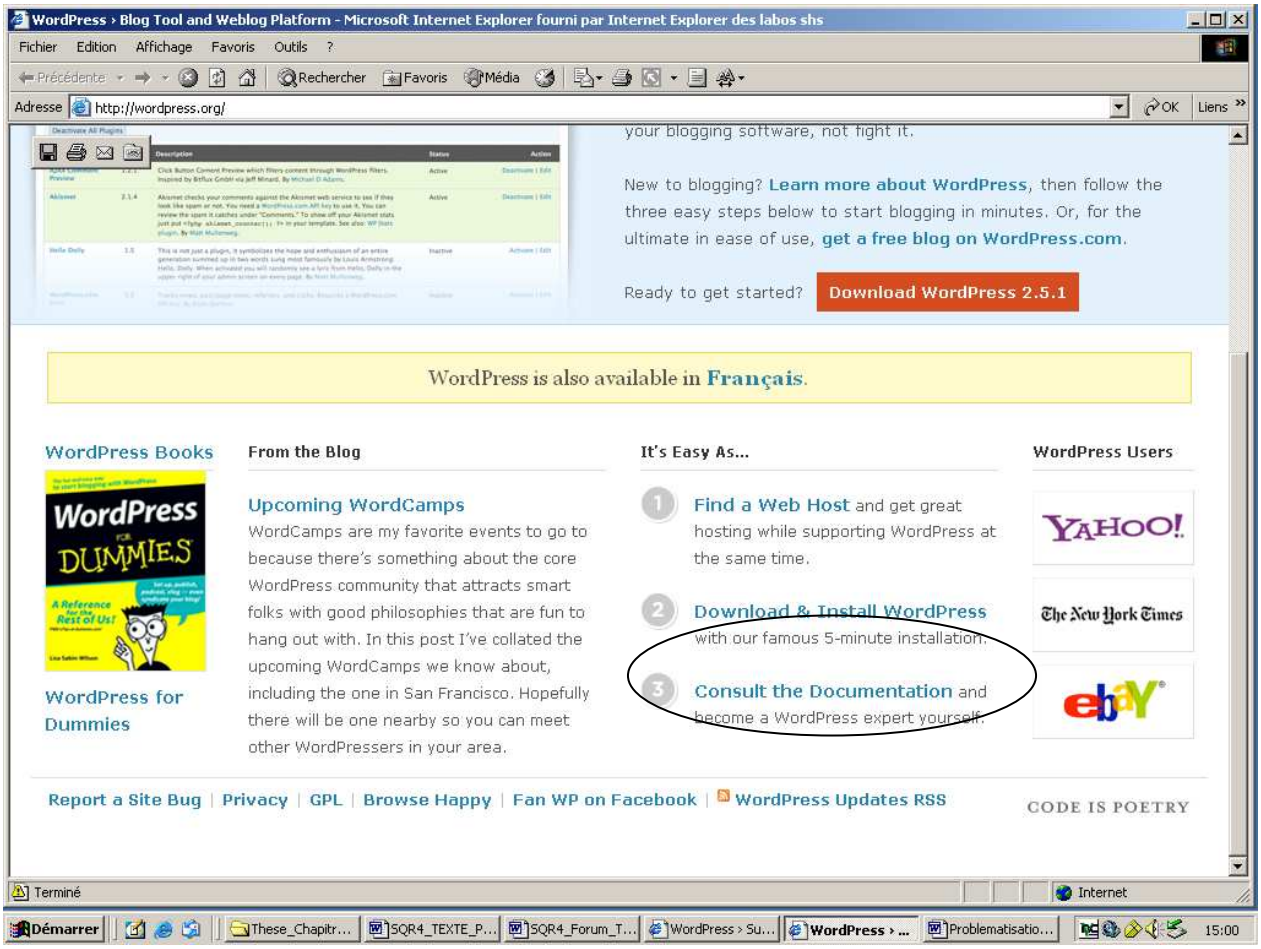


Figure 2: Capture d'écran du site principal de la communauté (effectuée le 5 juin 2008)

A gauche ci-dessus, nous avons un exemple de manuel de référence « classique » avec un seul auteur, Lisa Sabin Wilson, paru récemment[17] et validé par une préface de Matt Mullenweg, l'un des leaders de la communauté. A droite, nous avons la documentation dynamique originelle sur le wiki nommé Codex (nous entourons le *lien* qui y mène sur la capture d'écran), qui a obtenu le statut de manuel officiel en décembre 2005 alors qu'il s'agit du produit de la rédaction collaborative de plusieurs milliers de contributeurs. Notre connaissance de l'histoire de la communauté montre que ce dernier système d'information collaboratif a été au minimum précieux voire indispensable[18]. Ordinairement, la

[17] Il y en a plusieurs ; il suffit de chercher sur Amazon à « WordPress ».

[18] Pendant longtemps, il n'y a pas eu de manuel, ce sur quoi insistaient les contributeurs au forum et à la documentation (source : http://wordpress.org/support/topic/37984); les autres formes de savoir plus ou moins codifiées étaient dispersées sur Internet (tutoriaux) ou plus tacites (expertise accumulée dans le forum).

documentation d'un logiciel est associée à un ouvrage papier : il s'agit des gros manuels des outils bureautiques, rarement lus par les utilisateurs et qu'on trouve dans les librairies spécialisées comme Eyrolles. Le passage au support wiki par opposition au support papier est une problématique d'internalisation à l'intérieur d'un projet d'une fonction habituellement externe. Ce cas n'est pas isolé : un nombre considérable de communautés open-source comportent des wikis servant à la documentation sous des formes diverses et variées.

Notre sujet paraît donc d'une aridité certaine, et ne semble pas vraiment présenter d'intérêt : quoi de plus banal que la rédaction d'un manuel technique (même au sens large) et son usage ? C'est nettement moins exaltant que l'innovation en matière technique par exemple en matière de conception du code: les logiciels sont en effet des œuvres d'art et le produit le plus complexe de l'ingéniosité humaine (HORN, 2004). En fait, la documentation est considérée comme l'une des limites du logiciel libre par comparaison avec le logiciel propriétaire[19]. Sur ce point, dans les communautés open-source, les wikis ont été introduits entre autres pour remédier au caractère vite obsolète de la documentation (MEISZNER A., 2007). C'est donc aussi une solution de gestion, qui demande compréhension et évaluation.

Dans le cas de WordPress et certainement dans d'autres cas de logiciels libres, se pose donc la question de savoir quel est le meilleur mode de documentation du logiciel. La documentation fait participer l'utilisateur à la communauté des contributeurs, ce qui suppose des formes de gouvernance. On observe d'ailleurs des formes de concurrence non plus entre logiciels libres et logiciels propriétaires, mais entre outils « libres », par exemple WordPress et Dotclear. La question de l'interaction avec les utilisateurs par la documentation est absolument cruciale pour comprendre le succès d'un logiciel par comparaison avec un autre.

Ensuite, les communautés forment un aspect essentiel de la chaîne de valeur des nouvelles formes industrielles de la Toile (BITOUZET, 2001, BENGHOZI, 2006), ce qui en fait un objet pertinent pour la gestion. La question de la chaîne de valeur implique ici de ne regarder qu'un aspect, au niveau de la frontière entre firme et communauté. Des questionnements en gestion peuvent y être soulevés de façon pertinente bien qu'il ne s'agisse pas de la firme. Certes, les communautés de l'Internet ont fait l'objet d'une ample littérature, que nous ne rassemblerons et commenterons pas pour le moment. En la matière, force est de constater que l'étude de la question reste exploratoire. DAHLANDER et MAGNUSSON (2005) montrent dans le cas des firmes nordiques que les relations entre firmes et

[19] Source : http://www.volle.com/travaux/offrelogiciel.htm .

communautés peuvent être complexes, plus ou moins intéressées ; différents modes de management sont la conséquence de cette diversité de relations. LIN (2005) dégage les facteurs clefs caractérisant les relations entre communautés et entreprises associées. Enfin, et c'est fort étonnant, aucun travail à part PETERSEN (2008) ne mentionne les risques d'exploitation des productions collectives par d'autres, en particulier les entreprises dont les modèles d'affaire reposent sur ces communautés. Pourtant, de telles problématiques de gouvernance sont absolument essentielles pour comprendre la viabilité de la motivation des contributeurs, qui sont loin d'être des naïfs en matière économique.

Dans le cas qui nous intéresse, WordPress, notre propos de thèse n'est en effet pas l'aspect visible du succès de la « jeune pousse » WordPress, *mais les débuts humbles*[20] *et cachés de la communauté nécessitant un retour vers ses archives*. L'objet d'étude choisi est son système d'information d'importance stratégique, organisé autour du wiki baptisé par la communauté « Codex », élément essentiel de sa « chaîne de valeur ».

Notre thèse fait ensuite partie d'un questionnement général sur la gouvernance des systèmes d'information, qui est une thématique de plus en plus étudiée dans la communauté des chercheurs en SI. La littérature sur la gouvernance des systèmes d'information est conséquente : 8 références sur SSRN, mais bien davantage dans sciencedirect. C'est aussi un terme fréquemment mentionné dans la littérature en gestion dite « grand public »[21] . Cependant, les approches et les définitions de « gouvernance » s'appliquent bien aux firmes, mais plus difficilement à d'autres objets comme les communautés, comme celle-ci par exemple, pourtant académique : "IT governance is the process by which firms align IT actions with their performance goals and assign accountability for those actions and their outcomes." (WEILL et ROSS, 2004). Cependant, il apparaît que la question de la gouvernance des communautés et l'approche usuelle en SI forment des *littératures séparées*, ce qui nous incite à nous en tenir là pour le moment, la littérature nous orientant dans une direction qui n'est pas des plus pertinentes, mais nous reviendrons sur ce point dans ce chapitre. Il est prudent de penser que la question est complètement différente dans une communauté open-source, et une telle approche requiert une définition spécifique.

Enfin, notre thèse fait partie d'un questionnement de la discipline sur la gestion du savoir (« knowledge management »). Certes, la question de la gestion du savoir en ligne peut

[20] Les fonctions de support dans la communauté sont moins prestigieuses que la programmation.
[21] C'est ce qu'on trouve chez Eyrolles par exemple. « La « Gouvernance des Systèmes d'Information » ou « Gouvernance informatique » (en anglais « IT gouvernance ») renvoie aux moyens de gestion et de régulation des Systèmes d'Information (SI) mis en place dans une entreprise pour atteindre ses objectifs. A ce titre, la gouvernance IT fait partie intégrante de la gouvernance d'entreprise. ». (http://www.journaldunet.com/solutions/0506/050608_tribune.shtml)

se poser avec d'autres outils antérieurs aux wikis[22]. Cependant, en matière de knowledge management, les wikis sont une pratique d'une nouveauté significative (LEE et LAN, 2007). Dans les communautés open-source[23], la question du partage du savoir en ligne est absolument centrale, mais, et c'est surprenant, elle n'a pas pour autant été significativement étudiée. Des requêtes sur SSRN, sciencedirect, où il n'y a rien, et la base de données du MIT sur le mouvement open-source montrent que la thématique du management du savoir dans ces communautés n'a pas été abordée de façon conséquente. ZHANG et STORCK (2001) montrent que les membres dits « périphériques » des communautés contribuent de façon significative en matière de « knowledge management », dont ils profitent d'ailleurs en retour. LANZARA et MORNER (2003) est un travail exploratoire sur l'écologie d'agents et d'artefacts dans deux communautés, suivi de HEMETSBERGER et REINHARDT (2004) sur la communauté KDE et qui l'est encore davantage. SOWE et al. (2007) étudient le partage du savoir entre développeurs et non-développeurs dans la communauté Debian. MADANMOHAN et NAVELKAR (s.d.) étudient *seulement* l'impact des rôles des différentes parties prenantes à une communauté sur le knowledge management. La littérature reste exploratoire, alors qu'un tel questionnement sur l'accumulation du savoir a pourtant des conséquences macroéconomiques de taille (FREEMAN, 2007).

1.3. Que sait-on des formes de gouvernance dans les communautés open-source et les communautés de contenu ouvert (« open-media ») ?

Dire comme nous venons de le faire que nous allons étudier la question des wikis dans les communautés open-source comme situation de gestion suffirait à justifier notre thèse dans cette discipline. C'est en effet un processus délicat et avec de forts enjeux : il ne s'agit pas moins que de la *pérennité du mouvement open-source*, qui ne dépend pas que de la qualité du code. Cependant, les questions sous-jacentes comportent des enjeux plus larges, entre autres conceptuels, en théories des organisations, que nous souhaitons aborder ici.

[22] Voir par exemple pour le cas des « groupwares » GUNNLAUGSDOTTIR (2003) ou, pour la problématique du Knowledge Management dans le management de projet en ligne, HAMERI et PUITTINEN (2003). Pour une remise en perspective sur la longue durée, voir GIJBERS (2004).

[23]…mais pas seulement, comme le rappelle la communauté « Tela Botanica » ou d'autres communautés épistémiques et de pratique.

1.3.1. La gouvernance des communautés[24] open-source mériterait des approfondissements sur de nombreux points

La littérature sur l'open-source, un phénomène désormais d'une certaine profondeur historique, *a fait l'objet d'une masse croissante de travaux, tout à fait justifiée en raison de l'importance économique et historique du phénomène*. Cela ne veut pas cependant dire que l'on ait déjà une vue et compréhension *vraiment exhaustive* de tout ce qui est associé au phénomène. Nous mentionnons quelques limites des travaux précédents en matière de gouvernance. L'analyse explicite de cette thématique est observée dans le contexte des communautés open-source (DEMIL et LECOCQ, 2005). O'MAHONY et FERRARO (2007) développe justement une articulation entre autorité et démocratisation dans un échantillon de communautés. Une telle approche par l'angle gouvernance ne pose pas de difficulté en sciences de gestion. Le phénomène open-source est parfois considéré comme *lui-même porteur d'une forme originale de gouvernance et de régulation* (JENKINS, 2003, DEMIL et LECOCQ, s.d.), susceptible d'être appliquée ailleurs (MARRELLA et YOO, 2007, LJUNGBERG 2000, LESSIG, 1999). Cette originalité des modes de gouvernance pourrait aller au-delà des conceptions traditionnelles en théories des organisations (LATTEMANN et STIEGLITZ, 2005, OSTERLOH et al., 2004).

On observe une focalisation de l'analyse de la gouvernance sur les questions d'incitations et motivations. MAGGIONI (2002), par exemple, rappelle en effet que les incitations font partie de la gouvernance. Certains débats ont fait l'objet de travaux récurrents depuis les articles de LERNER et TIROLE (2000 et 2004). Toujours dans une optique économique, OSTERLOH et ROTA (2007) montrent de quelle façon la gouvernance des communautés résout un problème de bien collectif de second ordre, celui de la régulation comme « bien collectif » en tant que tel, réguler ayant un coût. La thématique des incitations et motivations des développeurs (par exemple SCHIFF, 2002), ou la thématique de la participation, ont fait l'objet d'une focalisation marquée. Ce qui se passe réellement est cependant plus nuancé que ce que suggèrent les premiers articles d'économistes, qui ont suscité des réactions diverses que nous devons *prendre en compte ici*. ROBERTS et al. (2007) pour citer une belle étude longitudinale, ou ALEXY et LEITNER (2008) montrent la complexité du lien entre motivations et efficience d'un projet, ainsi que celle du lien entre

[24] Le concept de communauté est l'un des plus employés en sociologie. Nous avons ici l'intention d'en étudier un cas concret, ce qui est un préalable aux définitions générales. Pour une réflexion plus globale sur la question, voir par exemple GUERIN (2004).

motivations intrinsèques et extrinsèques. En opposition avec la littérature économique, BITZER et al. (2004), BITZER et al. (2007), SHAH et KESAN (2003), LAKHANI et WOLF (2003) et OREG et NOV (2008) montrent au contraire l'importance première des motivations intrinsèques.

La gouvernance des communautés[25] open-source étudiée dès le début de la littérature sur la question recouvre pourtant des aspects variés : une forme de « gestion des ressources humaines » (« membership management »), des règles et des institutions, des formes de « monitoring » et sanctions, une forme de focalisation sur la réputation comme source de motivation et de contrôle, une culture partagée (MARKUS, 2000) et à coup sûr des problématiques de *management des systèmes d'information*.

On remarquera en particulier qu'on en sait peu sur l'organisation des tâches autres que la programmation, en particulier les tâches plus humbles, bien que cette thématique d'importance pour le succès d'un projet ait été soulignée depuis plusieurs années. Les tâches humbles dans les communautés open-source ont été négligées, alors que LATTEMANN et STIEGLITZ (2006) mettent en avant la nécessité d'obtenir un travail de qualité à tous les niveaux du projet, ce qui oblige à mettre en place des processus de gouvernance. LERNER et TIROLE (2004), probablement l'article le plus cité sur l'open-source qui parle de gouvernance n'évoque pas celle des parties les plus humbles. Sur ce point, on ne peut citer que LAKHANI et VON HIPPEL (2003) et KRISHNAMURTHY et TRIPATHI (2006). On sait aussi peu d'éléments sur la gouvernance du savoir, à quelques exceptions (LEE et COLE, 2003). Pourtant, ces questions sont fondamentales pour comprendre l'efficience des projets[26].

Les travaux sur le logiciel libre comportent aussi des limites méthodologiques cruciales si on s'intéresse à la gouvernance, sur lesquelles nous travaillerons ici. Nous mentionnons en particulier le manque d'études longitudinales, une approche pourtant indispensable si on a pour projet d'étudier la gouvernance sous un angle dynamique ; en effet, il n'y a aucune raison de penser que des évolutions en cette matière n'existent pas dans les communautés au fil de leur histoire. Deux exceptions ont été relevées pour l'organisation des communautés, WYNN (2003) et SADOWSKI et al. (2008) qui étudient spécifiquement l'émergence d'une forme de gouvernance dans la communauté Debian.

---

[25] Nous signalons que la notion de communauté ne s'applique en fait qu'à une petite minorité de projets : l'essentiel des projets sont organisés autour d'un développeur isolé (KRISHNAMURTHY, 2002). Ce n'est cependant pas le cas de notre communauté.

[26] Pour certains projets de taille, d'autres fonctions que la programmation apparaissent et elles ne sont guère étudiées non plus, comme celle du marketing (KRISHNAMURTHY, 2005), qui recoupe celle de la documentation.

1.3.2. La gouvernance des wikis : une dimension indispensable à leur viabilité, mais insuffisamment étudiée

Nous n'abordons la littérature sur les wikis que de brève façon ici, ce qui était pourtant indispensable : elle fera l'objet d'un chapitre entier après celui-ci qui en développera certains aspects.

Nous nous contentons ici de préciser que la question de la gouvernance est d'une importance fondamentale pour comprendre la croissance et la viabilité des communautés reposant sur des wikis (ce que montrent ROTH et al., 2008 dans l'un des rares points de vue « macrosociaux »). Cependant, ce qu'on en sait est insuffisant. MADINIER (2006) rappelle (et semble confirmer) que « selon Jimmy Wales, la gouvernance de Wikipedia est « un mélange de consensus, de démocratie, d'aristocratie et de monarchie ». Sur ce terrain particulier et non représentatif, déjà, la question comporte plusieurs dimensions. Le concept de gouvernance commence à apparaître dans certains travaux (WONG W. et WONG, L., 2005, Zacklad, 2007) qu'il serait clairement utile d'approfondir. GIORDANO (2007) montre par une étude de cas l'importance "en creux" d'une bonne gouvernance, caractérisé notamment par le manque d'intégration et de norme à l'intérieur des groupes. FROOMKIN (2005) montre que l' « auto-organisation », à l'opposé *a priori* du concept usuel de « gouvernance », ne va pas de soi, une idée à retenir. GEORGE (2007) montre que la gouvernance de Wikipedia qui a pour but de maximiser le nombre de contributions utiles est délicate, car elle doit d'un côté limiter le nombre d'éditions nocives et maintenir le zèle des principaux contributeurs. En France, les travaux sur la question restent rares. CARDON et LEVREL (2009) est l'un des principaux travaux exploratoires en français à avoir tenté de comprendre l'originalité de la gouvernance de communautés de grande taille comme Wikipedia.

1.4. La « collaboration massivement distribuée » et les communautés open-source : quelles articulations ?

1.4.1. Le problème de l'insuffisance de l'analyse et de la conceptualisation de la « collaboration massivement distribuée »

L'idée sous-jacente dans les développements qui suivent est que les modes d'organisations sous-jacents aux communautés faisant usage de wikis sont de façon *décisive* originaux, et pas seulement à cause de l'artefact technologique en question. Le créateur du concept de « collaboration massivement distribuée » était clair là dessus :

> « The sudden and unexpected importance of the Wikipedia, a free online encyclopedia created by tens of thousands of volunteers and coordinated in a deeply decentralized fashion, represents a radical new modality of content creation by *massively distributed collaboration*. [We] will discuss the unique principles and values which have enabled the Wikipedia community to succeed and will examine the intriguing prospects for application of these methods to a broad spectrum of intellectual endeavors"[27]

Les concepts servant à décrire cette activité nouvelle sur des wikis sont multiples. Il en est ainsi du concept de Yoshai Benckler, « peer based production » (BENKLER, 2001), mentionné pour décrire ce qui se passe dans Wikipedia (et repris par LOUBSER et DEN BESTEN, 2008). Cependant, ce terme doit être pris avec précaution, étant donné que le cadre théorique de Benckler est spécifique (théorie des coûts de transaction). On parle aussi ailleurs d' « open-content communities » (SCHROER et HERTEL, 2007), « open-media communities », « massively virtual collaboration », « mass collaborative authoring » (HANSEN, 2007) « user-created content », voire de « web 2.0. » comme synonyme, etc. . Mais nommer le fait qu'il existe des formes de travail particulières[28] dans les communautés faisant usage de wikis ne suffit pas. Curieusement, nous n'avons quasiment pas trouvé d'article problématisant ce concept de « collaboration massivement distribuée », ce qui montre les problèmes de conceptualisation associés.

Dans quelle mesure ces concepts décrivent-ils effectivement une activité spécifique aux communautés reposant sur des wikis ? Cela peut se discuter. Sur ce point, nous citons ensuite une autre définition opposée, dont « l'assiette » est en effet plus large. Mais cette autre façon de décrire ce type de phénomène, qui souligne la spécificité de Wikipedia mais qui

[27] L'auteur est un expert en matière d'open-source: "Mitchell Kapor, 55, is the President and Chair of the Open Source Applications Foundation, a non-profit organization he founded in 2001 to promote the development and acceptance of high-quality application software developed and distributed using open source methods and licenses." (source: http://www.ischool.berkeley.edu/newsandevents/events/dls11092005). Ainsi, lorsqu'il parle d'originalité significative, notamment par rapport à l'open-source, sa parole a du poids.

[28] « Collaboration » vient de « cum » et « laborare ».

l'étend bien au-delà, réduit sans doute son originalité. Le concept de « massive virtual collaboration » (CROWSTON et FAGNOT, 2008) est plus étendu, alors qu'une question prioritaire serait de définir ce que cet idéal-type permet effectivement de décrire de façon pertinente :

> "The term massive virtual collaboration highlights the following signal features of these phenomena: large numbers of contributors, commensurate with the popularity of the activity but ranging from dozens to tens of thousands; mostly unpaid contribution by contributors, for reasons subject to much speculation but little data; and jointly focused activity, in which contributors collectively create new content, structure, presentation, and/or computer software with possible value to some larger audience. Wikipedia is the most dramatic (but hardly unique) example of MVC. This online encyclopedia has expanded rapidly (over seven million articles in more than 200 languages) due to the huge number of text contributions from voluntary contributors (more than 4 million account holders) who help to develop and edit content for the site. But MVC includes also smaller-scale collaborations, such as contributors to blogs and discussion groups on a wide variety of topics and the teams of programmers and users who develop free/libre open source software (FLOSS), contributing software, documentation and bug reports."

Les auteurs vont jusqu'à inclure sous un même concept Wikipedia, des blogs, et les communautés open-source : on peut donc au contraire discuter la pertinence d'un concept aussi large, ne serait-ce que par la démographie hétérogène des populations intervenant dans les systèmes d'information en question.

Comment alors caractériser la spécificité de la « collaboration massivement distribuée » ? On peut dire que la différence entre « organisation distribuée » ailleurs sur Internet et « collaboration massivement distribuée » vient du concept de « masse », probablement de même que dans l'expression « culture de masse ». Il s'agirait d'une modalité *extrême* d'ouverture aux contributions, qui ne caractérise aucunement toutes les communautés. Le wiki sert justement à faciliter des contributions, à *abaisser* la barrière à l'entrée autant que possible, non pas seulement en terme de nombre de contributeurs, mais aussi en matière d'expertise (moindre) associée à ces contributions, voire d'une diversité encore plus poussée. Cependant, l'existence d'un nombre considérable de contributeurs effectifs ou potentiels pose des problèmes de *gouvernance* spécifiques et dynamiques (routinisation des pratiques, structuration de la collaboration, motivations…), que les définitions mentionnées ici ne permettent pas de cerner et que nous aborderons donc dans cette thèse. En matière démographique seule déjà, NEUS (2001) rappelle une loi connue qui énonce qu'une telle ouverture pose des problèmes de coordination, et n'est donc pas souhaitable:

> “These information quality problems, found in many virtual communities today, appear to confirm the old saying “Too many cooks spoil the broth”. A modern and perhaps more “scientific” version of this notion has become known as Brooks’ Law, named after Frederick P. Brooks, author of the classic book “The Mythical Man-Month”. Brooks talks about the inherent complexities of coordination and states that as the number of involved programmers N rises, the work performed also scales as N, but the complexity and vulnerability to mistakes rises as N2, in accordance with the number of communication paths required to coordinate the contributors. To achieve quality, Brooks therefore recommends a minimum of contributors: “Conceptual integrity in turn dictates that the design must proceed from one mind, or a very small number of agreeing resonant minds.” Conversely, Brooks’ Law predicts that “a project with thousands of contributors ought to be a flaky, unstable mess”, as Eric Raymond put it.”(NEUS, 2001)

Pourtant, des projets arrivent à passer outre cette « loi » (?) en réussissant à atteindre dans une mesure significative leurs objectifs. C’est le cas de Wikipedia, dans une certaine mesure. *C’est donc ce qui fait entre autres l’originalité du phénomène*. Il reste à comprendre comment.

Ainsi, avant de pouvoir répondre définitivement à la question de l’originalité des formes organisationnelles, il serait nécessaire de disposer d’une compréhension satisfaisante définitive du phénomène wiki. Notre revue de littérature montrera que ce n’est pas encore le cas. Une autre voie de recherche possible est de constater des écarts entre quelques pratiques associées aux wikis et des pratiques d’autres formes organisationnelles, comme celles de l’open-source, ce que nous avions évoqué plus haut et que nous développons spécifiquement à présent.

1.4.2. « Collaboration massivement distribuée » et « open-source », quelques clarifications

Le concept dit « open », qu’on trouve dans « open-source » ou « open-media » demande examen. DENG (2008) est d’ailleurs consacré à l’analyse de la notion d’ouverture dans ces *institutions*. Il la problématise en montrant qu’elle peut prendre diverses formes : “open institutions, in whatever forms, are essentially about decentralized production of a collective good (or "commons") that relies on voluntary collaboration of highly variable human-related input”. Curieusement, il ne traite aucunement des wikis et de Wikipedia alors que ce sujet est prioritaire. Cette conceptualisation de la diversité renforce notre argumentation précédente qui insistait sur la spécificité des wikis.

Fréquemment, les configurations « organisationnelles » qu’on trouve dans les communautés open-source et les communautés de contenu ouvert (« open content ») sont

confondues en un même concept ou mouvement. La question de leur articulation peut faire l'objet de confusions, même dans les travaux scientifiques alors que cela devrait être au contraire *problématisé*. Un exemple révélateur dans une revue à comité de lecture serait OKOLI et OHB (2007) : ils tentent de passer, pour l'étude de la reconnaissance comme motivation de l'open-source à l'open-media en parlant pour Wikipedia de « largest open source content community ». Cette expression lourde est une façon pour nous de montrer que l'amalgame est problématique. La base de donnée connue du MIT sur le phénomène open-source elle-même présente quelques travaux portant sur Wikipedia, alors qu'il ne s'agit pas de logiciel, ce sur quoi porte l'archivage en question.

« Collaboration massivement distribuée » est un concept qui a été forgé à partir de l'exemple-type de Wikipedia, qui elle-même a été inspirée par les principes du mouvement open-source (GEORGES, 2007) : on parle de « culture open-source »[29]. Or, il plus que probable qu'il existe des différences entre les deux « modèles » : plus de contributeurs, moins d'expertise dans le mouvement dit « rise of the amateur », implication moindre, pas d'incitations explicites, barrière à l'entrée moindre, etc.). Il y a donc adaptation, évolution, différenciation, peut-être vers une « popularisation » de certains principes de l'open-source.

Nous proposons quelques approfondissements qu'il a fallu chercher dans la littérature. CIFFOLILLI (2004) montre que les deux approches soulèvent des problèmes différents, en particulier une sensibilité plus forte de l'open-media au piratage. Ensuite, c'est sur le cas extrême de Wikipedia que se sont focalisées les critiques du modèle, comme celle de l'anti-élitisme (SANGER, 2004) qui est à l'inverse consubstantiel aux grandes communautés open-source : les élites sont les experts en informatique. Les critiques de Wikipedia et du phénomène open-source existent, mais elles ne se recouvrent pas (MONTESSUIT, s.d.). La frontière est en particulier problématisée en matière de qualité de l'information chez DUGUID (2006), qui montre que la « loi de la qualité » tirée des communautés open-source ne serait pas valable pour les communautés « open-content » comme Wikipedia[30]. L'origine de l'absence de qualité commence ailleurs à être problématisée, sur Wikipedia encore une fois :

[29] La « culture open-source » serait une extension de l'esprit de l'open source, qui concernait seulement le logiciel au départ. Elle peut englober l'art, le savoir, la politique, etc. Ce n'est pas notre sujet. Cependant, il faut se rappeler que cette extension de l'open-source au-delà du logiciel ne concerne pas que Wikipedia, et qu'elle peut poser dans d'autres contextes des problèmes d'adaptation. Le concept de collaboration massivement distribuée a été lui-même ensuite étendu à d'autres objets (blogs, partage de photos, etc.), mais une telle transposition pose problème, demanderait une étude sérieuse en soi, et nous ne tiendrons pas cela pour acquis étant donné que les wikis comportent de fortes spécificités structurantes.

[30] Cependant, marqué par une expérience négative, cet auteur peut être ici discuté.

> *"This thesis has been developed in fascinating detail by James Surowiecki in his recent book,* The Wisdom of Crowds*. It is also the validating belief of the open source software movement, summed up in the motto – Given enough eyeballs, all bugs are shallow.*
>
> Maybe. But in the case of specialized subjects where quality criteria are more judgemental (unlike software bugs), or where relevant expertise is spread very thinly, the "crowd" is unlikely to be sufficiently wise. So there will always be a secure niche for expertise in the traditional sense. Indeed, that conviction led Wikipedia's co-founder, Larry Sanger, to leave what he had created out of despair over the hostility toward expert authority that dominates Wikipedian corporate culture. Sanger is now creating a new on-line authority, Digital Universe, that seeks to provide both expertly-created as well as collaboratively developed content." (NICHOLSON, 2006)

Le wiki aurait permis cependant de changer l'échelle de la construction de sites Internet, en permettant la création de communautés à *une plus large échelle* (LEE et LAN, 2007), au moyen des contributions des plus humbles comme des experts (KING, 2007) et avec une barrière à l'entrée limitée (« login » et « gravatar » seuls par exemple), et donc en changeant les comportements des internautes. Ensuite, les TIC supportant les deux institutions ne sont pas les mêmes. Un wiki n'est pas un CVS par exemple. Cela peut avoir un rôle décisif pour comprendre les formes organisationnelles observées (ce que suggéraient déjà WATSON et al., 2004). L'outil wiki est fait pour rendre les contributions très faciles, ce qui n'est pas le cas de l'essentiel des textes de code, même si des outils d'aides à la programmation existent. La différence que nous explicitons est donc bien une problématique de SI, avec à la fois une dimension technologique et une dimension organisationnelle.

Enfin, la notion de « production de contenu » ne correspond pas au type de production logicielle, celle du code. Un processus de catégorisation comme « folksonomy » concerne l'open-media, mais n'apparaît pas dans la production de logiciel. L'opposition entre code et contenu verbal est d'ailleurs l'un des arguments de ceux qui s'opposent aux Wikis (MONTESSUIT, s.d.). Mais il faut aussi nuancer : pour un manuel utilisateur, qui énonce des procédures, il y a rapprochement…

Ainsi, la séparation entre les deux formes institutionnelles « ouvertes » est significative.

Pourtant, ce n'est que dans des travaux marginaux que les spécificités des deux modes « d'organisation » sont envisagées. On trouve cette idée de séparation entre « peer production » et « open-source » chez WAGNER et MAJCHRZAK (2007); même si l'étude est exploratoire, elle montre qu'il y a des différences significatives, moins d'exigence d'expertise par exemple, et que la prise en compte des seuls facteurs provenant de l'open-

source ne suffisait pas à décrire les originalités observées dans les cas[31]. WATSON et al. (2004), un travail en système d'information sur la gouvernance des communautés comportant un cadre théorique fondé sur les coûts de transaction, comporte un intérêt taxonomique. Il ne place pas dans ses typologies Wikipedia et les communautés open-source dans les mêmes cases : la première serait une communauté « adaptative », l'autre une communauté « inventive ». On pourra toujours discuter le cadre analytique, mais l'article montre que la question est sérieuse. Enfin, SUNSTEIN (2006) montre que les deux types de communautés soulèvent des interrogations spécifiques.

L'articulation entre open-source et open-media, deux phénomènes significatifs, est une question pertinente historiquement pour étudier le « déploiement du numérique » : quelles sont les différences et formes d'interactions alors que le premier est à l'origine de l'autre ?

1.5. Progression et conséquences de la constitution de notre problématique

1.5.1. Retour sur le concept de gouvernance

Le mot gouvernance est à la mode, jusque dans la presse spécialisée grand public[32], et pas seulement en système d'information (« gouvernance d'entreprise »). En économie, Eric Brousseau parle plus particulièrement de « self-governance » pour les communautés[33], ce qui est presque une contradiction dans les termes par comparaison avec l'usage du concept en management des systèmes d'information, à moins de donner une extension conséquente au concept. Il devient alors synonyme « d'auto-organisation », un concept pertinent pour les communautés. Précisément, dans la communauté WordPress entre autres, les comportements effectifs et admis se font de façon informelle et décentralisée. Cependant, négliger ce concept n'est pas une bonne idée, comme le montre cette affirmation dans une thèse abordant les wikis, qui ne décrit pas ce qui est observé :

[31] Que le lecteur nous croit lorsque nous disons avoir eu aussi cette intuition de notre côté, comme le montre notre contribution dans La Tribune de mars 2007, écrite dès décembre 2006 : nous faisions alors précisément référence à la distinction entre open-source et open-media qui relevaient de deux logiques de légitimation différentes.

[32] Ainsi, le site InternetActu comporte une section spécialisée « gouvernance » (http://www.internetactu.net/category/categories/politiques-publiques-gouvernance/, consulté le 05/07/09). L'Url montre qu'il y a confusion ici comme ailleurs entre ce concept et politique publique. Or, nombre de questions concernant « la Cité » sont aussi des problématiques de gouvernance, et ne font pas pour autant partie des politiques publiques : la question de l'open-source comporte des enjeux de gouvernance, et les politiques publiques ont pris du temps pour les prendre en compte.

[33] http://www.brousseau.info/en/home/index.php (consulté le 05/07/09).

> « What happens when everyone in the Open Community is an editor, in the absence of centralised control”? (PAQUET, 2003, page 70).

Ou encore ailleurs:

> « Wiki represents the power of many and this power is distributed collectively to improve content quality. Each author is able to change the contributions of other authors, refining the quality of the knowledge assets” (HASAN et PFAFF, 2006).

Cette constatation peut se discuter : il peut exister justement des formes de contrôle certes informelles, mais centralisées dans une mesure significative tout de même. Une communauté n’est pas forcément l’anarchie, dans notre cas comme ailleurs[34]. Même dans un contexte où il y a justement peu de débats et de conflits d’éditions entre contributeurs en désaccord (« edit wars »), le processus peut être contrôlé. On peut discuter le fait qu’une absence totale de gouvernance soit viable. La question est donc pertinente, et elle le reste d’ailleurs même lorsqu’émerge une forme de contrôle informel, comme pour notre communauté.

Il restait à définir le terme de façon adéquate dans cette perspective. Mais il est difficile de poser une définition initiale, elle devra au contraire être un aboutissement. Pour cette recherche, elle doit être significativement large, car la notion d’organisation appliquée aux « communautés » soulève des difficultés, et relève davantage de l’action organisante (« organizing ») que de l’organisation formelle. Nous adoptons donc la *perspective* de DUMEZ (2004), qui, dans *Gouverner les organisations*, propose une approche qui comporte cette propriété, correspond à la tradition du CRG et a le mérite de revenir au sens originel du concept, au-delà de définitions ultérieures potentiellement restrictives, ce qui permet de reprendre la question sans idée préconçue :

> « Gouverner, au sens propre, c’est mener un navire d’un point à un autre, en tenant compte des caractéristiques du navire et de son équipage, des courants, des vents, de l’état de la mer. C’est un art, combinant calcul, chance, flair, expérience, saisie des opportunités et des sautes de vent. Croire qu’il pourrait se réduire à la mise en place d’un système d’incitations ou toute autre forme de pilotage automatique (même si cette dimension est essentielle), paraît très éloigné des réalités »

Cette perspective est pertinente, étant donné que le concept s’applique d’après DUMEZ (2004) aux communautés sur Internet : Géraldine DeSanctis propose dans l’ouvrage en question justement une analyse de la structuration des forums, un sujet proche du nôtre, dont l’analyse pose des difficultés. La question de la gouvernance des « common pool resources » n’a pas été suffisamment abordée dans cette contribution sur ce TIC[35]. Les wikis sont un

[34] Pour le lien entre éthique « hacker » et anarchisme, voir par exemple IMHORST (2005).

[35] Nous renvoyons sur ce points aux propos de la relectrice Anca Metiu dans la partie « débats » à la fin du chapitre en question.

terrain pertinent pour supporter également un tel questionnement sur la gouvernance[36], pour ne parler que de cet aspect.

1.5.2. Les conséquences associées au choix de notre problématique et de notre terrain pour comprendre la gouvernance des communautés reposant sur des wikis

Notre mise en perspective montre que notre processus de problématisation comporte une ambiguïté, qui est à la fois une difficulté et la marque de la richesse potentielle de notre sujet.

Comprendre la spécificité des formes « organisationnelles » associées aux wikis demande examen. Un des présupposés de ce travail est de montrer qu'un concept insuffisamment conceptualisé peut être abordé avec profit lorsqu'il se trouve dans son terrain *originel et usuel*, celui des communautés open-source : non seulement ses *spécificités* apparaissent par contraste, mais encore il s'agit du terrain *originel* où le TIC qui supporte cette forme organisationnelle est le plus pertinent, comme nous l'avons vu. Ces deux objectifs, recherche des spécificités et pertinence du terrain peuvent être antagonistes.

En première approximation, on pourrait dire que le modèle du Codex est une organisation hybride entre le modèle Wikipedia et le modèle de l'Open-Source, qui ont tous les deux de fortes spécificités (DUGUID, 2006). En particulier, on peut dire que cette communauté est à la frontière entre le mode « documentaire » (un manuel est un document, comme l'encyclopédie Wikipedia) et l'écriture du « code » : on trouve du code et des pages de régulation de sa production sur le Codex[37]. On pourrait affirmer que nous sommes à la frontière entre deux modes de régulations, l'open-media et l'open-source[38].

Une autre façon peut-être encore plus pertinente d'envisager cette recherche serait de se demander si les caractéristiques observées plus loin dans notre thèse *ne seraient pas, justement, le cas général*, ce d'autant plus que le terrain des communautés open-source est de première importance. En effet, conceptualiser la collaboration massivement distribuée par rapport à Wikipedia pose problème, étant donné qu'il n'y a pas de raison de partir de ce cas

[36] Lorsqu'on parle de gouvernance, Michel Foucault est souvent utilisé, y compris en SI (voir par exemple BOYLE, 1998). Foucault utilise (un peu comme nous le terme de gouvernance) le terme « gouvernement » comme un terme générique qui englobe toutes les méthodes qui consistent à diriger la conduite d'individus ou de groupes. Cette perspective large est conforme à la nôtre, même si les axes théoriques choisis par nous diffèrent.

[37] La question de la localisation de la documentation dans les projets open-source ne va pas toujours de soi : on trouve des formes de documentation à l'intérieur du code d'ailleurs.

[38] La communauté a pris conscience de cette différence d'ailleurs (source : http://lists.automattic.com/pipermail/wp-docs/2005-September/001024.html).

particulier. En réalité, le phénomène wiki est avant tout un outil de « geeks », et il serait préférable d'examiner la question sur son terrain originel. Ainsi, dans cette optique, si la pertinence de l'ensemble étudié augmente, les frontières entre les deux phénomènes sont en revanche plus difficilement observables, mais permettraient plus sûrement d'examiner leurs spécificités.

D'un autre côté, l'étude de la collaboration massivement distribuée (dont le modèle ou l'idéal-type serait Wikipedia) dans son contexte originel si exigeant serait *une forme de mise à l'épreuve* d'un mode d'organisation dérivé de celui de l'open-source, de son utilité, efficacité, ou robustesse. Wikipedia peut servir de modèle à une communauté dans une mesure significative : l'idéal-type peut avoir une fonction performative et servir de modèle, et c'est le cas justement pour WordPress. L'articulation qui nous intéresse comporte ainsi une dimension dynamique : dans quelle mesure observe-t-on alors des adaptations et quels sont les mécanismes ou processus significatifs sous-jacents ?

Il serait téméraire de donner des réponses définitives à ces interrogations avant d'avoir une vue pertinente du phénomène dans son contexte originel. Or, notre travail est encore pionnier en la matière.

1.5.3. Présentation détaillée du choix des sous-questions de recherche et de la structuration de notre travail

L'étude du passage entre auto-organisation et formes émergentes de gouvernance dans un travail en gestion imposait d'étudier celles-ci là où des problèmes de gestion significatifs pouvaient apparaître, tout en s'adaptant aux contraintes du traitement des données. Une telle démarche est inductive, du moins dans sa conception générale, sous forme d'aller et retours, au fur et à mesure de l'avancement par tâtonnement de notre réflexion et des opportunités de notre cas. Ces choix correspondent à des thématiques qui sont apparues comme cruciales et prioritaires au fur et à mesure de la prise de connaissance de l'histoire de la communauté. Les sous-questions de recherche permettent d'aborder des aspects essentiels de l'organisation de la production collaborative de contenu informationnel dans un wiki: son histoire et son contexte, son organisation et management, l'émergence de contributions en nombre suffisant, les formes de structurations de la collaboration, et une compréhension de son appropriation.

Une telle orientation de travail est légitime dans une étude de cas avec plusieurs unités d'analyse (YIN, 2003) correspondant à autant d'aspects essentiels de ce qui est examiné. Cela

permettait aussi de minimiser les risques en cas d'échec de telle ou telle sous-question, et de maximiser la probabilité de résultats. Cela peut ressembler à une logique de « gestion de portefeuille ». Enfin, notre thèse est conçue pour permettre une autonomie des chapitres permettant une lecture partielle et/ou dans le désordre, ainsi qu'un découpage en quelques articles publiables ultérieurement: le chapitre méthodologique, qui est central dans une thèse en gestion, a déjà permis la publication d'un article indépendant (RUZE, 2009) alors qu'il n'était pas prévu pour cela.

D'autres thématiques ne sont pas à exclure dans des études ultérieures sur WordPress. Nous pensons par exemple à une approche plus « culturaliste » de la communauté. Mais il n'est pas sûr que cela soit décisif pour comprendre sa viabilité et son succès, et une telle approche plus sociologique ne se range pas aisément dans un questionnement de ses aspects économiques. En matière économique du moins, nous n'avons pas détecté d'autres thématiques d'études pertinentes. Nous restons ouverts à la discussion en matière d'orientation de travail, étant donné que nous avons l'intention de continuer à travailler ce cas.

Voici à présent quelques justifications plus approfondies, qui permettent également de donner un aperçu *partiel* des résultats que nous *développerons* plus loin.

Le chapitre 4 donne une vision minimale de l'histoire de la communauté et plus particulièrement de la partie de la communauté qui traite de la documentation. Il permet de mieux comprendre la dynamique spécifique de la documentation et son articulation avec le reste de la communauté. Au sortir de ce chapitre, nous disposerons d'une vision dynamique et englobante plus claire qui permettra déjà de dégager des régularités pertinentes et significatives en matière de rythmes comparés des différentes parties de la communauté, voire d'en tirer des enseignements. Il permet aussi d'introduire les autres sous-questions de recherche.

Le chapitre 5 porte sur l'étude du processus de formation du Codex, en particulier sur l'apport de contributions par la communauté. Ce type de questionnement est pertinent car la littérature sur l'analyse du mouvement open-source évoque de façon significative la question de la formation d'un bien collectif et des motivations associées (chapitre 2), une thématique qui appartient clairement à la littérature économique, expérimentale ou pas. Notre but sera de dégager les caractéristiques quantitatives du processus, les motivations des contributeurs, et de montrer le caractère satisfaisant du niveau de contribution qui en a découlé.

Le chapitre 6 porte sur l'étude des formes de routinisation dans la communauté, qu'il s'agisse de la mise en forme du contenu informationnel ou de la gestion de la communauté. Nous verrons que nous pouvons observer une forme d'hybridation entre les pratiques usuelles de collaboration massivement distribuée caractéristique des wikis et de Wikipedia en particulier et les pratiques de l'open-source. Ces formes informelles de rigidification des pratiques sont cruciales pour comprendre la viabilité du projet de documentation, et sont essentielles à la compréhension de l'organisation et du management de cette partie de la communauté.

Le chapitre 7 porte sur l'analyse des formes de collaboration qui supportent les processus précédents. *Cette gouvernance est mise en oeuvre en premier lieu sur une liste de discussion centralisée*. Nous étudions de façon fine les formes de structuration du processus en question, au moyen d'une approche quantitative, y compris une analyse de réseau. Au-delà des caractéristiques dissymétriques du processus et de l'approfondissement de ses spécificités dynamiques et historiques, nous mettrons *entre autres* en évidence qu'il est caractérisé par un phénomène d'avalanche, d'attachement préférentiel entre catégories de participants, et par des formes faibles de connectivité cependant renforcées par des formes significatives de réciprocité.

Le chapitre 8 porte sur les formes d'appropriation du contenu informationnel du wiki. Par une étude des archives des autres listes de discussions des autres parties de la communauté, en particulier des courriels évoquant le wiki, nous tentons de comprendre et comparer les usages des informaticiens chevronnés et des utilisateurs novices. Nous verrons que pour ces derniers, l'appropriation ne va pas de soi et suppose des formes de régulations et de complémentarité entre systèmes d'information (forum et wiki).

1.6. Nos orientations théoriques

1.6.1. Une approche économique est compatible avec un travail en management des systèmes d'information[39]

Le choix d'une approche économique se justifie par l'existence de deux questions générales de raretés qu'il est nécessaire de comprendre. Mais l'objet de cette tentative de compréhension dans cette présente thèse consiste en processus de gestion (de cette rareté).

La première relève de la problématique de *constitution d'un bien collectif*. Elle est classique en économie, et a été l'objet de nombreuses d'interrogations. Pour donner une version imagée, notre travail examine un « *paradoxe de la soupe aux pierres* », pour reprendre une « parabole » de nature économique *fréquente* dans le milieu de ceux qui s'intéressent aux nouvelles formes d'organisation du web[40]. Seulement, contrairement à ce que suppose l'histoire simpliste en question, qui raconte la formation bénévole d'un *bien collectif*, un tel processus suppose non seulement des *régulations*, mais aussi des *interactions* entre les contributeurs ; c'est ce qui nous intéresse, on l'aura compris, en priorité, sur ce terrain où le bien collectif en question est plus complexe que la préparation d'une soupe.

La seconde est le corrélat de la première, ou une autre façon de la décrire. Dans le contexte complexe de l'Internet, tel que défini précédemment, une problématique générale qui est donc sous-jacente et qui en est le corrélat est celle du *formatage de l'information*. L'une des raretés principales dans l'économie du numérique, c'est l'information « bien » formatée (et la définition de ce « bien » est problématique). « On ne manque pas d'information, on manque de formatage de l'information" disait Herbert Simon. Une autre façon de parler de ce problème est de le décrire comme un bien public de second ordre (VON KROGH et al., 2003), formater et organiser l'information ayant un coût significatif. Le concept de formatage, différent de celui de standardisation, suggère donc à la fois un bien rare (pourquoi un individu d'une « communauté » accepterait-il de donner pour faire émerger ce bien public ?) mais

[39] Le choix d'une approche économique est acceptable d'une façon générale en sciences de gestion. Par exemple, certains programmes de recherche du CRG autorisent largement la mobilisation de théories économiques (DUMEZ, 2005).

[40] Pour résumer, il s'agit d'une fable qui raconte l'histoire de deux mendiants qui arrivent dans un village d'une province ravagée par la famine. Ils demandent à être nourris. Les villageois refusent. Ils installent alors sur un feu, à l'étonnement général, une marmite remplie d'eau avec une pierre dedans au milieu de la place et annoncent qu'ils font de la soupe aux pierres pour tout le monde. L'un d'eux dit que la préparation est convenable, mais qu'elle serait meilleure avec du sel. Un villageois daigne ajouter une pincée de sel. L'autre mendiant dit que la chose s'améliore, mais que ce serait meilleur avec du persil. Un autre villageois jette dans la marmite une branche de persil. Et ainsi de suite. Les deux mendiants parviennent ainsi à rassembler suffisamment d'ingrédients pour servir à la population une soupe convenable alors qu'il y avait famine.

aussi un processus, impliquant des individus travaillant ou non ensemble, ce qui dépasse la perspective économique usuelle en terme d'équilibre.

1.6.2. Des théories économiques employées parfois en système d'information à une orientation en économie cognitive

Les approches économiques (ou d'économistes) existent mais ne sont pas les plus courantes en MSI[41]. D'autres théories sont nettement plus dominantes, comme la théorie de la structuration prônée par Orlikowski. Sur le site « IS World » qui fait référence, les seules théories citées sont la théorie de l'agence, la théorie des asymétries d'informations sur les marchés (« lemon market theory »), et la théorie des coûts de transaction[42]. Dans la même veine, dans un manuel de référence en MSI comme LAUDON et LAUDON (2006), la perspective se limite à cette dernière théorie. Les approches mobilisées sont donc clairement orientées vers ce qu'Olivier Favereau appellerait la « théorie standard étendue », au détriment d'approches « non standard ». On observe par ailleurs des approches économiques dans les travaux sur l'open-source, mais le constat est le même que précédemment : la littérature porte principalement sur les questions d'incitations (RUZE, 2009), ce que confirme encore une fois une requête sur la base de donnée du MIT répertoriant les travaux sur la question. Mais les incitations ne sont qu'une partie de « l'économie des organisations »[43], un champ de recherche dépassant les orientations précédentes réclamé par DALLE et al. (2004). Par ailleurs, ces travaux ne se situent pas en général clairement en MSI[44].

Si on devait encore mettre en perspective plus générale notre sujet, ici sur le plan théorique et économique, nous pourrions poser la question ainsi : comment comprendre les dimensions économiques essentielles des nouvelles formes industrielles du Web ? Or, nous pensons que les concepts analytiques de l'économie standard souvent cités à propos des communautés open-source sont d'une portée limitée : celle-ci décrit une « économie simple »

[41] Michel Volle, Thierry Pénard à ICIS 2008 par exemple.
[42] Source : http://www.fsc.yorku.ca/york/istheory/wiki/index.php/Main_Page (consulté le 21/07/09)
[43] Certes, quelques autres articles évoquent certes des questionnements économiques, mais ne relèvent pas de la théorie des organisations (adoption de l'open-source et modèle économique associé) : nous ne citons pas ces travaux qui sont hors-sujet.
[44] Les problèmes de frontières lorsqu'il est question des travaux sur les communautés et l'open-source ne vont pas de soi.

(LERNER et TIROLE, 2002, et 2004, où ils persistent et signent[45]), qui reposerait sur des incitations claires, comme la construction d'une réputation d'expert en programmation (pour se faire recruter), le plaisir éprouvé à programmer, à résoudre un bogue, etc. Le matériau empirique *implicite* à une telle conceptualisation est la population des programmateurs d'un projet.

Mais une telle approche n'est à première vue pas pertinente pour comprendre des aspects essentiels de la « chaîne de valeur » d'un projet logiciel, en particulier la « logistique » du processus de documentation dynamique qui est la « colonne vertébrale » de l'aide aux utilisateurs: il s'agit de tâches bénévoles, ingrates, mal reconnues, au contact d'utilisateurs exigeants. Que signifie pour nombre de contributeurs au wiki acquérir une réputation dans notre communauté lorsqu'on ne connaît même pas le langage PhP, qu'on ne s'en cache pas, et que les hackers sont eux-mêmes réticents à documenter leurs propres « plugins » ? Pourquoi écrire par bribes sur un wiki, participer à une rédaction collective (ce qui n'est pas toujours amusant, contrairement à la programmation, comme on le verra au chapitre 5), alors qu'il est peu probable de se faire recruter dans une entreprise pour si peu ?

Une étude approfondie des aspects économiques clefs du projet supposerait plutôt la mobilisation des outils de l'économie hétérodoxe et un positionnement de ce type de recherche dans le champ de la théorie des organisations qui est plus large, et justifie également une thèse en gestion. Intuitivement, notre objet d'étude, le Codex, est un « repository of knowledge », un objet de gestion dont nous étudions différents aspects. Mobiliser l'économie cognitive pour étudier les problèmes d'efficacité, de coordination, dans un sujet centré sur la « connaissance » est tout à fait défendable[46]. Enfin, les wikis sont étudiés au Centre de Recherche en Epistémologie Appliquée de l'Ecole Polytechnique (CREA)[47].

Notre projet est de considérer dans quelle mesure des approches en économie cognitive au sens large sont complémentaires plutôt que concurrentes dans notre étude de cas comportant plusieurs objets (les « items » de YIN, 2003). C'est là aussi une façon de spécifier les théories, comme le demandait DUMEZ (2005). Certains cadres théoriques sont en effet

---

[45] Nous pensons que l'article de 2004 ne répond pas sérieusement aux critiques de l'article de 2002. Une partie de la littérature a en effet réagi contre de telles simplifications, qui posent problème en gestion (et en histoire, discipline qui se méfie de ce qui serait « simple »).

[46] De fait, nous avons aussi fait le choix du terrain des wikis afin de trouver un terrain d'application pour l'économie cognitive, ce qui est pertinent d'après nos discussions avec M Gensollen, spécialiste des communautés, rencontré au 2e trimestre de l'année 2005.

[47] Voir par exemple ROTH (2007 et 2008), et plus généralement une communauté interne au CREA (source : http://chavalarias.com/tiki-index.php?page=chadisc&bl=y , consulté le 14/07/09). Nous avons peut-être eu une influence sur ce point. Sur le Web 2.0, voir RAYNA et STRIUKOVA (2008), qui est une approche économique par un doctorant du CREA.

pertinents pour certaines sous-questions de recherche en matière de gouvernance et pas pour d'autres. C'est la mise en concurrence qui aboutit à une spécialisation de chacune des spécifications de l'économie cognitive au sens large. Par ailleurs, rien n'empêche que notre cas *alimente* à la marge ou de façon décisive des approches théoriques existantes ou embryonnaires, ce qui est une autre façon d'en cerner les limites. Les apports peuvent fonctionner dans les deux sens, des cas empiriques pouvant par exemple compléter une thématique théorique « en friche », ou des outils et concepts éclairer des matériaux complexes.

D'une autre façon, c'est montrer en quoi les courants théoriques dits « hétérodoxes » en économie[48] peuvent être féconds. Par ailleurs, notre projet se caractérise par un croisement des traditions de recherches de deux laboratoires voisins qui ne collaborent pourtant pas de façon significative[49]. Notre travail est donc bienvenu sur ce point.

Notre position en matière théorique n'est qu'une partie de la spécificité de notre travail, un autre aspect de notre thèse étant l'usage d'une approche historique, comme on le verra dans le chapitre méthodologique (voir aussi RUZE, 2009). Il existe sur ce point une homogénéité entre les outils théoriques choisis et l'approche historique[50], comme perspective ou comme outil de traitement de données, ce qui ne veut pas pour autant dire qu'un tel rapprochement ait été fréquent.

1.6.3. Spécification des éléments théoriques choisis

En ce qui concerne *le chapitre 4*, essentiellement historique, la théorie ne joue pas un grand rôle, étant donné qu'il a un rôle introductif de contextualisation et est volontairement descriptif. Néanmoins, le processus de centralisation de la documentation en un seul SI observé est assimilable au processus de « lock-in » caractéristique de la diffusion des standards technologiques (ARTHUR, 1994). Cette fois, il s'agit d'un standard interne à une communauté, i.e. un système de knowledge management. Ce chapitre exploratoire n'était pas sensé apporter des résultats en tant que sous-question de recherche explicite, c'est pourquoi l'articulation entre théories et données n'était pas prévue.

[48] C'est une catégorie largement popularisée du fait de ses enjeux politiques de taille et reconnue par la classification du *Journal of Economic Literature*.

[49] Des complémentarités existent d'après ce qui apparaît des discussions avec des doctorants des deux laboratoires.

[50] LESOURNE et al. (2003) prône en conclusion le dialogue avec l'histoire, et l'historien Chandler a rendu un hommage appuyé aux approches évolutionnistes en particulier (CORIAT et WEINSTEIN, 1995).

En ce qui concerne *le chapitre 5*, voici à présent le thème théorique d'économie cognitive sous-jacent. La théorie économique sur la formation des biens collectifs (informationnels ou non) explique qu'elle pose problème : tout le monde peut dans ce type de situation bénéficier de l'information, bien rare, apportée par les autres, ce qui incite à maximiser ses gains en ne contribuant pas tout en en profitant autant que possible. Si tout le monde agit ainsi, le bien collectif, informationnel ou non, ne se forme jamais, car chacun espère que les autres contribueront à sa place. Ce type de questionnement (dit « public good formation »), nous tenons à le rappeler, est d'une portée générale et d'une grande importance sociétale. C'est la raison pour laquelle l'économie cognitive a largement étudié ce type de question en particulier sous l'angle expérimental (« public good game »[51]). Cela n'interdit cependant aucunement de travailler cette question, dans la perspective qui est la nôtre, de façon plus inductive, contextualisée et historique. Or, ici, c'est le contraire qui s'est produit dans une mesure significative. Nous tenterons donc d'examiner de façon inductive nous insistons, les processus qui permettent de mieux comprendre le dépassement de ce type de paradoxe économique. Si le questionnement de départ est un thème économique, sa résolution, étudiée par une analyse inductive qualitative et quantitative, peut être qualifiée de gestionnaire.

En ce qui concerne *le chapitre 6*, les concepts théoriques mobilisés *demandent de trop longs développements pour être placés ici*. Nous renvoyons donc au début du chapitre en question. Nous précisons seulement que nous allons tenter d'observer les routines qui ont émergé dans la partie de la communauté qui s'occupe de la documentation sur le Codex. L'observation d'une organisation naissante est une opportunité unique d'observer « the nature of organizational imprinting » et la cristallisation du système de routines locales (WARGLIEN et al., 1997). La littérature orientée vers l'ethnographie virtuelle met aussi en avant l'importance du concept de routine pour analyser les problématiques de coordination (METIU, 2008). Nous allons pour comprendre les routines tenter d'articuler l'économie évolutionniste et l'économie conventionnaliste, deux écoles hétérodoxes qui ont fait de la

[51] Les archives de « EconPapers » conservent nombre de travaux abordant cette thématique : http://econpapers.repec.org/scripts/nep.plex?list=nep-exp (consulté le 26/0/09). Pour une explicitation de la relation entre économie expérimentale et cognitive, voir NOVARESE et RIZZELLO (2004).

cognition un objet d'étude essentiel en particulier en ce qui concerne de telles approches microanalytiques[52].

En ce qui concerne *le chapitre 7*, sur les formes de structuration de la collaboration, il n'y a guère de difficultés en matière de spécification sur le plan théorique. Certains thèmes abordés dans cette sous-question de recherche, comme « émergence », « avalanche », « homophilie », « réseau », sont familiers aux chercheurs de l'économie cognitive, tradition économique intéressée par les processus dynamiques, au-delà de l'analyse de réseaux qui s'est prolongée récemment au CREA vers une étude des systèmes complexes au sens large. Il y existe un programme de recherche sur la cognition sociale. Signalons que l'analyse des réseaux est devenue un champ d'étude en soi en économie[53]. Seulement, des courants orthodoxes et hétérodoxes utilisent tous les deux ce type d'approche, et nous nous situons clairement en économie cognitive, sous-discipline qui s'est intéressée aux questions de complexité et aux processus évolutionnaires (LESOURNE et al., 2003, BOURGINE, 2004), ce qui est le cas des phénomènes sous-jacents à la série chronologique étudiée dans ce chapitre 7. L'analyse de la complexité et l'économie cognitive sont étroitement liés au CREA.

En ce qui concerne *le chapitre 8*, le cadre théorique sous-jacent est encore limité et il demande au contraire à être enrichi par des travaux empiriques préalables, ce qui est l'un des *objectifs possibles* d'une étude de cas. Voici un rappel d'un aspect économique important de l'appropriation de l'information par un Maître :

> "What information consumes is rather obvious: It consumes the attention of its recipient. Hence a wealth of information creates a poverty of attention." (SIMON, 1971).

Cette question économique de la rareté de l'attention soulève la nécessité de l'étude gestionnaire d'un processus, à savoir la question de *l'appropriation* par les utilisateurs de l'information avec le moindre coût « attentionnel » possible : est-ce toujours le cas, dans quelle mesure si ce n'est pas le cas, et comment, à quelle(s) condition(s) ? "One questionable assumption of the hard approach[54] is that making information available will automatically

[52] Le concept de routine pose en économie des problèmes théoriques et méthodologiques que nous ne pouvons traiter de façon exhaustive ici : nous préciserons le minimum nécessaire au chapitre 6. Par ailleurs, abordée souvent en intelligence artificielle ou en économie expérimentale, leur analyse « sur le terrain » n'est pas évidente et insuffisamment développée (NOVARESE et RIZZELLO, 2004).

[53] Ce point est trivial et nous ne le documentons pas. Pour une vue d'ensemble de cette tradition de recherche, nous renvoyons seulement aux archives de la communauté scientifique en question qui diffuse les travaux en la matière : http://ideas.repec.org/n/nep-net/ (consulté le 12/07/09).

[54] Il s'agit des outils usuels de « knowledge management » qui rendent le savoir explicite et codifié, par opposition aux approches informelles et au savoir tacite.

make people more knowledgeable" (SHIN, 2004). Penser qu'il s'agirait d'un processus évident et spontané est un a priori, qui peut être parfois vérifié, ou au contraire aboutir au constat de l'existence de pratiques de gouvernance informelles.

C'est donc en ce sens que nous parlerons « d'économie du knowledge management ». Intégrer ce type de questionnement dans l'économie cognitive est défendable étant donné que le lieu de la rareté, l'attention, est d'ordre cognitif. Comprendre les fondements économiques des usages et de l'appropriation montre l'intérêt d'une approche d'économie cognitive inductive sur le terrain, « in the field ». Le Codex est en effet une « knowledge base » dont il subsiste des traces d'usages, des commentaires divers, etc. qui dépendent du contexte d'utilisation (informaticiens, newbies). Ils peuvent décrire des problématiques de rareté, d'arbitrages, etc., donc, des aspects économiques.

La littérature concernant cette approche est encore exploratoire. BROUSSEAU (2007), probablement l'un des auteurs à l'origine du concept de « economics of kowledge management », propose seulement un cadre d'analyse et ne développe pas spécifiquement la question de l'appropriation du savoir accumulé dans les communautés qu'il évoque pourtant. FOSS et MAHNKE (2003) est dans la même veine (cadre théorique englobant et orthodoxe). La perspective de SHIN (2004) cherche à intégrer les théories précédentes et se situe donc aussi à un niveau général. Les références en la matière sont donc rares, et toutes sont orientées vers ce qu'Olivier Favereau appelle la « théorie standard étendue ». Une approche plus hétérodoxe, plus inductive, et assumant une perspective microanalytique est donc bienvenue, surtout pour traiter des questions économiques impliquant la cognition, à propos desquelles les précédentes théories sont a priori moins pertinentes. Sur ce point, en conséquence, c'est plutôt notre cas qui *complète* l'orientation théorique que nous avons choisie, ce qui permet d'en souligner les limites tout en désirant les compléter dans l'optique que nous avons choisie, une approche historique. Notre approche est ici clairement inductive et se justifie par la rareté actuelle des travaux : au delà de la théorie, il s'agira donc aussi d'un changement d'approche.

Ainsi, nous avons choisi un périmètre large pour définir l'économie cognitive, au point d'inclure l'économie des conventions et « l'économie du knowledge management ». Une telle ouverture est justifiée par les recoupements significatifs[55] qu'on observe dans les communautés scientifiques qui traitent les questions de cognition en économie et par des

[55] Par recoupements, nous entendons que les « working papers » sont souvent *les mêmes* dans les listes de diffusions des travaux scientifiques dans ces communautés.

terminologies qui ne sont pas encore stabilisées[56]. Nous n'allons pas effectuer ici une présentation détaillée des courants de recherche. Il faut retenir en revanche est qu'à l'articulation entre économie cognitive et histoire, ce qui est le positionnement de notre travail, beaucoup reste à accomplir, et nous lancerons quelques pistes : analyse des routines, des motivations essentiellement intrinsèques, analyse de la complexité des interactions et de l'émergence de réseaux, analyse des mécanismes d'appropriation du savoir.

1.7. Ouverture vers la suite de la thèse

Nous nous proposons en fin de chapitre introductif un bref retour sur notre parcours. Notre thèse peut mieux se comprendre par une vue rétrospective sur notre cheminement, fait de projets, de contraintes et de hasards.

Ayant terminé 2e du DEA « Economie des institutions » et ayant obtenu une allocation et un poste de moniteur, nous avions le droit de venir dans un laboratoire de l'Ecole polytechnique. Présenter notre thèse en gestion était déjà envisagé. Nous avons cependant passé deux ans dans un laboratoire d'économie, le FORUM à Nanterre, où nous n'avions pas du tout notre place sur le plan méthodologique et théorique. Nous ne nous étendrons pas davantage sur cette expérience pénible. Ce furent pour résumer deux années perdues, qui, avec le recul, nous ont seulement permis de connaître l'importance de la thématique des communautés open-source et des wikis pendant notre insertion dans le projet PERSI sur les réseaux sociaux lancé par les informaticiens du laboratoire LIAFA. Notre intérêt pour l'économie cognitive qui a émergé lors d'un séminaire de DEA avec Paul Bourgine en 2003 s'est en revanche maintenu, avec la diffusion d'un travail sur les ERP (RUZE, 2005). Cependant nos centres d'intérêts sont passés de l'économie expérimentale à une approche plus empirique et historique de l'économie cognitive. Nous sommes passés fin 2005 de l'économie à la gestion, tout en conservant un intérêt pour les questions économiques. Notre thèse en gestion a donc commencé en octobre 2005 avec un sujet renouvelé pour l'essentiel. Elle nous a amené à choisir un terrain et le cadre de l'étude de cas. Nous nous sommes tournés du fait des contraintes imposées par notre matériau et du caractère ouvert des sciences de gestion vers une méthodologie de type historique.

[56] Nous relevons essentiellement: « cognitive and behavioral economics », « evolutionnary economics », « experimental economics », « institutional economics ».

Nous reprenons encore une fois notre problématisation :

A partir du cas WordPress, nous nous demandons de quelles façons les communautés open-source gèrent leur savoir de façon efficace sur un support wiki.

Plus précisément, dans quelle mesure la « collaboration massivement distribuée » sur un support wiki caractérise-t-elle ces processus dans ce contexte?

Une façon d'y répondre est donné par le sous-titre que nous avons choisi pour notre thèse, composé d'expressions tirées de notre communauté :

« code is poetry », but « a documentation resource should be managed »[57].

[57] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000521.html . « Code is Poetry » est un slogan (assez trivial en fait) qui apparaît en bas de la page principale du site de la communauté.

Chapitre 2 : les limites de l'analyse par la littérature de la gouvernance des wikis dans les communautés open-source[58]

2.1. Introduction

Nous explicitons ici dans quelle mesure le savoir accumulé sur les wikis comporte les limites significatives qui suivent. Le terrain des communautés open-source a été négligé, ce qui est cohérent avec des recherches également lacunaires sur les tâches guère prestigieuses dans ces communautés. La littérature rassemblée comporte également des limites méthodologiques significatives, en particulier en matière d'étude de cas et d'analyse longitudinales. Des questionnements en sciences de gestion significatifs sont également marginaux. Enfin, sur ce dernier point, l'analyse des formes de constitution du bien informationnel collectif, de routinisation, de collaboration et d'appropriation, des travaux sont encore nécessaires.

De plus, une revue de littérature sur les wikis en général *qui se concentre bien entendu*, au fur et à mesure par un mouvement de convergence, vers la justification de notre travail de recherche qui en complète les insuffisances dans une certaine mesure est nécessaire. Il ne s'agit pas seulement du rituel de thèse. Faire le point et proposer une première synthèse des travaux scientifiques abordant les wikis se justifie. Il n'existe *aucun « survey »* en la matière pour le moment, comme pour d'autres TIC récents comme les blogs, et traiter le sujet comme il le mérite demanderait *un article de recherche en soi* en étendant davantage ce qui est proposé ici[59].

Ce travail est également nécessaire pour faire contrepoids aux récents ouvrages de vulgarisation[60], qui ont précédé les recensions plus systématiques, et qui ont un objectif plus appliqué et pédagogique. Ensuite, des articles récents affirment « rituellement » qu'il n'y a « pas grand-chose » d'écrit sur le sujet qu'ils abordent, comme Wikipedia. Or, si cela était vrai avant 2005, au fur et à mesure de l'avancement de ce travail force est de constater que cette idée ne tient plus, quantitativement du moins. Il apparaît que notre travail s'inscrit dans

[58] Cette revue de littérature a été commentée par Matthias Den Besten, auteur de plusieurs travaux sur Wikipedia. Nous le remercions ici de ses remarques judicieuses et d'avoir confirmé les conclusions auxquelles nous aboutissons ici.

[59] L'article qui irait au-delà de l'effort de concentration progressive des propos que nous livrons ici est à écrire. On peut renvoyer cependant à RUZE (2009) pour compléter à la marge en matière de gouvernance des communautés.

[60] Tous ne sont pas de bonne qualité. Nous mentionnons EBERSBACH (2005), auquel nous renvoyons les néophytes.

un environnement scientifique dont il est nécessaire de déterminer les contours et surtout la solidité. Enfin, lorsque cette idée d'une « littérature antérieure » est (brièvement) abordée comme « passage obligé », elle suggère à notre sens la perplexité pour des raisons que nous expliciterons. Ailleurs, les premières bibliographies[61] sur le sujet sont manifestement incomplètes et méritent un sévère travail critique, même si elles facilitent le travail du chercheur (VOSS, 2005).

Cette recension permet donc de justifier le *recentrage* de l'attention des chercheurs sur notre terrain, négligé manifestement.

Après avoir exposé la méthodologie de la revue de littérature, nous montrerons que les orientations de la recherche actuelle ont conduit dans une mesure notable à négliger notre terrain de recherche. Nous soulignerons ensuite les limites méthodologiques de la littérature antérieure, en particulier les insuffisances des approches gestionnaires, avant d'examiner dans quelle mesure les thématiques ayant trait à la gouvernance de communautés faisant usage des wikis méritent des approfondissements.

2.2. Orientations de la revue de littératures

Notre choix d'une concentration des propos sur le thème du wiki provient de l'hypothèse d'un pouvoir structurant de l'outil suffisamment fort pour caractériser le phénomène de collaboration massivement distribuée, dont l'idée d'origine provient de Wikipedia (voir chapitre précédent).

Il s'agit d'une littérature émergente pendant la durée de la thèse, de 2005 à 2009. Les wikis ont fait il y a quelques années l'objet de moins d'attention que les blogs (EFIMOVA, 2005) et on leur a peu prêté attention (RAITMAN et AUGAR, 2002). La raison de cela est sans doute leur développement plus lent (SZYBALSKI, 2005). La littérature sur les wikis que nous examinons ici est donc le fruit d'un développement récent, comme le montre la récurrence des conférences « Wikimania » et « Wikisym » (qui rassemblent surtout des informaticiens), conférences de la Fondation Wikimedia de 2005 à 2008, le sommet émergé de l'iceberg de la WikiResearch (sic).

Les recherches effectuées sur les bases de données classiques des chercheurs (sciencedirect, SSRN, SAGE…) montrent que la littérature sur la question n'y est pas d'une taille imposante. Le recoupement des bases de données a permis de créer un effet de

[61] … collaboratives ou non, comme http://blog.lib.umn.edu/kenne329/wikibiblio/ (consulté le 01/09/09).

saturation, qui fait qu'il est possible de faire le tour des références provenant des *revues à comité de lecture*. Cela contraste cependant avec l'essentiel des contributions sur les wikis qui sont des actes de conférences, des « working papers » : GoogleScholar ou « ACM Portal » sont plus prolixes.

Plusieurs problèmes méthodologiques doivent être pris en compte pour mener à bien cette revue de littérature. Il s'agit tout d'abord de la fiabilité des travaux, même parmi les travaux dans des revues significatives ou des colloques « de marque », ce qui demande un travail critique dont il faut avoir conscience. Il s'agit ensuite de l'exhaustivité du travail d'enquête et la détermination d'un seuil d'acceptation raisonnable d'un article dans la catégorie « scientifique »[62] , ce qui ne va pas de soi lorsqu'il est question surtout de publications de travaux « en ligne ». Notre postulat de travail était d'être « indulgent », afin de ne pas manquer des travaux exploratoires utiles. Ensuite, la pertinence de la transposition des conclusions ou problématisations des travaux sur Wikipedia à d'autres terrains ne va pas de soi, mais il est impossible de ne pas prendre en compte les travaux s'y rapportant.

Etant donné que les travaux en gestion ne sont guère nombreux, plusieurs autres disciplines sont susceptibles d'être mobilisées, dans l'ensemble des sciences sociales : sociologie, économie, ethnologie, « études communicationnelles », voire sciences juridiques, etc. . Cela peut poser des problèmes de compatibilité ou de pertinence. A cet égard, nombre d'articles sont *difficiles à situer*, autant parmi les « working papers » que parmi les travaux plus officiels. Où classer par exemple les articles de *First Monday* ? Cela montre aussi la nécessité d'une approche en SI plus « puriste » et d'une façon générale un *problème de lisibilité* en matière d'accumulation du savoir scientifique.

Nous avons seulement pris en compte les travaux « scientifiques » et non les savoirs informels des communautés de pratique[63]. Il y aurait pourtant quelques recherches à mener, car elles pourraient fournir des pistes pertinentes de recherche (sous forme de « benchmarking » peut être). Ne pas avoir été plus loin en la matière est aussi l'une des

[62] Nous n'avons pas tout mentionné en bibliographie pour cette raison.

[63] Signalons en effet qu'un certain nombre de communautés ou de praticiens enthousiastes (comme ceux de CraoWiki pour citer l'une des plus anciennes) partagent en ligne une partie du savoir accumulé en la matière depuis 1994, date de l'apparition des premiers wikis : l'expérience du terrain montre que les communautés qui font usage de nouveaux TIC parlent de ce qui les lie, en particulier l'outil lui-même. Les gourous des wikis aiment aussi à parler de leurs propres pratiques. Ces communautés de pratiques offrent des pistes de réflexion souvent pertinentes et variées, mais la prise en compte de ces savoirs suppose une approche critique attentive plus longue : les praticiens sont fréquemment « intéressés ». Nous pensions traiter ce matériau, mais il a fallu établir des priorités, et nous avons choisi de prendre le risque de le négliger, c'est un parti pris de ce travail. Une comparaison critique entre les apports de ces communautés épistémiques et communautés de pratique est souhaitable, mais nous avons arbitré en décidant qu'elle n'était pas prioritaire dans une perspective académique orientée vers l'étude de cas.

raisons qui nous ont poussé, en contrepartie pourrait-on dire, à « l'indulgence » dont nous venons de parler en matière de sélectivité.

2.3. Pourquoi une telle limite dont nous cernons ici les contours ?

Nous ne faisons pas que mentionner les limites de la littérature, nous tentons de les comprendre et de les mettre en perspective, ce qui est aussi utile pour d'autres recherches qui s'orienteraient dans des directions voisines.

2.3.1. Une littérature pour l'essentiel centrée sur Wikipedia

Wikipedia, projet gigantesque parmi les mieux référencés, mondialement connu, médiatisé et fascinant, a *littéralement hypnotisé* l'attention de chercheurs de plus en plus nombreux et selon des perspectives diverses. Nous pouvons raisonnablement faire l'hypothèse de phénomènes d'imitation et d'entraînement dans le milieu scientifique. Nous renvoyons par exemple à LEVREL (2006) pour une présentation générale, VOSS (2005) pour un travail statistique, KINZLER (2005) pour une méthodologie sémantique, STVILIA (2005) ou HOLLOWAY (2007), et à EMIGH (2005) pour une analyse des processus de certification de la qualité, CIFFOLILLI (2003) pour une analyse du phénomène en termes de coûts de transactions, etc. etc. . Nous en citerons d'autres par la suite sur des points d'analyse plus précis. Un courant scientifique, la « WikiResearch » (sic), dont l'événement-phare est depuis trois ans le WikiSym(posium), abritée par la WikiMedia Foundation, traite quasiment exclusivement du célèbre projet. Il existe certes d'autres systèmes de belle taille tournant sous Mediawiki, mais ce mouvement a le mérite d'être le « point focal » des chercheurs traitant du sujet des wikis[64]. Nous n'avons pas jugé pour le moment utile de quantifier la littérature sur Wikipedia. Il suffit de survoler rapidement les titres de notre bibliographie pour se rendre compte du degré de focalisation sur l'encyclopédie.

Ce type de travaux est tout à fait légitime: une *concentration* sur cette thématique peut permettre un progrès cumulé en matière de techniques d'analyse, ou le raffinement des problématisations et hypothèses à tester. On trouve aussi dans ce sous-courant de recherche

[64] Pour la France, voir en particulier le projet INRP à la frontière des problématiques Wikipédiesques et pédagogiques (http://www.inrp.fr/vst/Dossiers/Wikipedia/bibliographie.htm , consulté le 02/10/09).

les rares travaux proposant une lecture théorique minimale ou des approches quantitatives significatives. Mais il s'agit d'une organisation *particulière* et pas forcément représentative des usages des wikis. Pour ne parler que d'aspects démographiques, la plupart d'entre eux ont un faible nombre de contributeurs, alors que Wikipedia en compte plusieurs millions. Rien ne dit que les modes de régulations de tous les wikis ressemblent à ceux de l'encyclopédie (RUZE, 2009). Quelques critiques suggèrent qu'une telle focalisation de la recherche sur un terrain aussi controversé est téméraire (CORREA et al., 2005).

Lorsque le terrain d'étude diffère de Wikipedia, quelques gros wikis focalisent malheureusement l'attention (ARONSSON, 2002), de même que la littérature sur les projets open-source se focalisent sur les projets les plus voyants et les plus gros, et les démarches se font plus exploratoires. Les wikis sont également utilisés dans le domaine de *l'éducation*, ils font fureur par exemple en Allemagne et la recherche a également souligné de façon significative cet aspect du phénomène (GUZDIAL 2001, WANG et al., 2005, DESILETS et PAQUET, 2005, pour ne citer que quelques exemples). La thématique du journalisme participatif a aussi été étudiée. Il en est de même de l'appropriation par les libraires (FARKAS, 2005, GORMAN, 2005) et des usages politiques (RAYNES-GOLDIES et FONO, s.d. , MAKICE, 2006). En revanche, on observe, c'était attendu, une attention croissante portée aux wikis « corporates », qu'il était hors de question de cerner de façon exhaustive (on renverra par exemple pour une mise en perspective à HASAN et PFAFF, 2006 ou à DANIS et SINGER 2008 pour une enquête conséquente). Mais, même là, des approfondissements sont à prévoir.

2.3.2. Le terrain primordial des communautés open-source a été négligé

Cette attention limitée mais orientée vers le wiki le plus connu a très probablement conduit les chercheurs à négliger les terrains originels des communautés Open-Source (SCHWALL, 2003, BROUDOUX, 2002), alors que les usages des wikis ont été associés depuis le début à ce type de communautés (CUBRANIC, 2003). En effet, l'articulation entre le phénomène Open-Source et celui des wikis, quoique reconnue, n'a pas été sérieusement étudiée : nombreux sont les travaux où il n'y est fait *qu'allusion*, comme UNDP (2005), qui cite les wikis comme outil de coordination des projets, de même que BAC et al. (2005), JULLIEN et al. (2002), GUNNAR (2005), FROOMKIN (2003), CHRISTEN (2003), LIH (2004), HALLORAN et al. (2005), et enfin KOLBITCH et MAURER (2006). Si l'articulation entre le phénomène wiki et le phénomène open-source a donc été remarquée, il n'est pas

apparu que leur articulation soit comprise (COFFIN, 2000). Les rares approches critiques (qui restent, elles encore, centrées sur Wikipedia) du phénomène « wiki » rappellent cependant à juste titre qu'il s'agit avant tout d'un « outil de geeks » (pour un curieux pamphlet, MONTESSUIT, s.d.) :

> « Wikis became popular in the free and open-source software (FOSS) community, where they were ideal for collaboratively discussing and documenting software, particularly given the loose structure of the projects» but « *these early software-focused wikis did not attract widespread attention.*», (SZYBALSKI, 2005, c'est nous qui soulignons, voir de même BARTON, 2005).

Le fait que les wikis soient utilisés seulement dans une minorité de projets il y a quelques années (ROBBINS, 2003) où l'outil était moins connu pourrait aussi expliquer une telle négligence[65]. En conséquence, la diversité des usages dans ces communautés, influencée ou non par les pratiques de Wikipedia, influence dont aucun travail de recherche ne fait mention, a été oubliée.

Nous évoquons ici ponctuellement quelques éléments plus concrets qui servent de point d'appui à notre argumentation pour souligner l'importance de la thématique. Tout d'abord, nous avons personnellement constitué entre la fin de l'année 2005 et le début de l'année 2006 une base de données de plus de 300 communautés faisant usage des wikis. En voici un :

Figure 3: le wiki de la communauté Linux « Ubuntu » (capture d'écran du 5/07/09)

[65]Ainsi, les membres de la célèbre suite Open Office ont manifestement refusé au départ l'usage du wiki, même si ultérieurement un moteur MediaWiki a finalement été lancé.

Certaines communautés relevées par nous, comme KDE, Apache, Mozilla, Open-Office, sont connues. Il en existe bien d'autres, les wikis pouvant être utiles pour les petits projets qui comptent moins de contributeurs potentiels et où toutes les bonnes volontés sont bienvenues. *Nous n'avons pas exploité cette base de données en thèse*[66]. Nous le rappelons, les communautés open-source sont le terrain originel de développement des wikis ; le wiki C2 qui date de 1994 était un wiki de programmation extrême (« extreme programing »). L'une des conséquences tout à fait révélatrice de cela est que certains wikis sont même *conçus spécifiquement* pour les communautés open-source (Trac, WikiDev, XPSwiki, Scaki, etc.). Ensuite, la seule statistique dont on dispose sur le sujet vient de GLOTT et al. (2007). Elle est cruciale : si 94% des communautés disposent d'un forum et d'une documentation ou "knowledgebase", dans 51 % des cas un wiki a été utilisé pour supporter la fonction en question. C'est donc conséquent, et nous avons observé *effectivement* une telle profusion.

Nous reprenons la recension des travaux, qui montre *l'écart* entre la précédente énumération et la littérature actuelle. Quelques travaux « voisins », pourrait-on dire, tentent cependant d'aborder la thématique dans le domaine de l'informatique, mais avec des limites de poids, dont la principale est le caractère exploratoire, ce qui est souvent un euphémisme... CHAU et MAURER (2005) est un travail exploratoire évoquant les wikis dans le secteur IT. SERAKIOTOU (2008) est une courte description d'un wiki ayant eu un rôle majeur pour rendre la gestion du savoir plus efficiente dans un département IT, ce qui montre qu'il n'y a pas que des échecs. Des références que nous ne dénombrerons pas, comme ETO et al (2005), ont immédiatement des visées appliquées et proposent des améliorations par un nouveau *produit* informatique. Dans le secteur du logiciel, ANDERSEN (2004) mentionne, dans un contexte propriétaire, le cas STATA (le logiciel bien connu des lecteurs statisticiens) dont la mémoire repose sur un wiki dont l'usage aurait été, d'après l'entreprise en question, un réel succès. D'autres études sont souvent un éloge de TWiki (RAYGAN et GREEN, 2003, RADZIWILL et SHELTON, 2005, et BUFFA, 2006) qui n'est pas le meilleur outil «communautaire », mais plutôt un outil *d'entreprise*[67]. MARTLAND (2003), propose une étude exploratoire de leurs usages dans *l'entreprise* Platypus Group. Plus avant, XIAO et al., (2007) abordent le sujet mais n'évoquent que la programmation, et pas la documentation.

[66] Elle fera l'objet d'autres travaux plus quantitatifs par la suite, en particulier pour comprendre les contours d'un phénomène dont les descriptions sont souvent trop optimistes : contrairement au cas WordPress, les wikis sont le plus souvent des échecs (« villes-fantômes ») d'après nos premières recherches.

[67] Les utilisateurs sont par exemple Disney, Motorola, Cetelem, Google, mais aussi l'ENST, Université de Nice, etc. .

CRAIG (2005) propose quelques suggestions sur les wikis comme outil de management de projet dans les communautés Open-Source, mais le travail est exploratoire. CHENG et al. (2005) bien qu'abordant le rôle des wikis dans les activités de conception, est un travail qui relève plutôt du génie logiciel que des sciences sociales. Bref, le cadre original au sens strict n'a pas du tout été traité de façon approfondie. On peut penser que privilégier l'étude de « communautés internes » aux entreprises est plus rentable, en particulier dans ce contexte récent où des outils wikis « propriétaires » sont apparus. Enfin, un seul workshop en 2007[68] a brièvement abordé la question pour la programmation en général, mais il n'évoque pas spécifiquement le cas des communautés open-source et encore moins les activités humbles de support, comme la documentation. Il n'est en rien suffisant pour comprendre de façon exhaustive l'étendue des usages, ou la profondeur des problématiques en SI. D'une façon générale, un nombre significatif d'auteurs sont des informaticiens (en particulier ceux intervenants aux « WikiSyms ») ; or, l'expérience montre que les interrogations de ces spécialistes sont souvent différentes de celles des sciences sociales.

Pour donner quelques précisions sur ce dont il s'agit dans notre thèse, on peut résumer quelques points de AGUIAR (2005), qui est pertinent sur ce point, bien qu'il ne s'agisse pas d'un travail en sciences sociales. Il insiste sur la « web-based documentation », qui a supplanté les précédentes méthodes (« litterate programming », « multiple-source approach », « XML-based documentation »...) même si le format papier comporte toujours des avantages (confort visuel). Cette méthode est fondée sur l'usage du wiki ; il est l'un des rares auteurs à expliciter leur popularité dans le domaine du logiciel open-source. Il rappelle le rôle fondamental de la documentation dans le processus de développement du logiciel, et non plus seulement la dimension stratégique de la documentation pour comprendre le succès du logiciel auprès des utilisateurs. Il souligne plusieurs problèmes ayant certainement des conséquences pour comprendre la viabilité d'un projet organisé sur un wiki : la mise en cohérence d'éléments hétérogènes (…), l'écart (« gap ») entre documentation et développement (rythme, inciter les développeurs à composer de la documentation, ce qui ne va pas de soi, même lorsque des formes de « litterate programming » sont mises en place…), le processus de formalisation de l'information, i.e. le problème de cohérence sémantique tout au long du cycle de vie du logiciel, son caractère difficile, coûteux et fatiguant. Il s'intéresse aussi à la combinaison entre code et autres formes de textes.

[68] http://ws2007.wikisym.org/space/Wikis+for+Software+Engineering .

2.3.3. La question des tâches humbles dans les communautés open-source a été négligée, ce qui permet de donner une perspective plus générale à ces limites de la littérature

On le sait, il existe à présent une littérature en développement sur le rôle des utilisateurs dans le processus d'innovation en matière de produits (VON HIPPEL, 2005, pour l'ouvrage fondamental, CHAN et LEE (2004) pour une discussion sur l'efficience de ces processus). De ce fait, ce n'est pas à l'innovation seule que nous nous intéressons ; une autre raison est que le rôle des utilisateurs ne se réduit pas seulement à cela.

Dans les communautés open-source spécifiquement, l'étude des parties plus marginales, plus périphériques, de ce qu'on appelle le « code » est balbutiante (FELLER et FITZGERALD, 2003, BROWN, 2004). La prendre en compte en tant que telle est nécessaire pour montrer les limites des travaux qui nous concernent. En effet, de rares chercheurs éminents ont signalé leur importance et entamé leur étude, mais ces travaux *négligent* cette interface cruciale et complexe qu'est la documentation (LERNER et TIROLE, 2002, LAKHANI et VON HIPPEL, 2003, mais aussi ROBLES et GONZALEZ-BARAHONA, 2004). Le processus de documentation des logiciels est mal connu.

La littérature sur les communautés s'intéresse surtout à l'activité noble des informaticiens ainsi qu'aux projets « prestigieux », double orientation qui est cohérente. La littérature a pourtant reconnu l'importance fondamentale d'une « bonne » communauté (SENYARD et MICHLMAYR, 2004, NICHOLS et TWIDALE, 2006, ou LEE et al., 2009 pour prendre des dates éloignées), i.e. une population qui ne comporte pas que des développeurs. Notre étude de cas confirmera au besoin cette idée. Il est fait parfois mention dans la littérature à ces activités annexes, auxiliaires, ancillaires, ou secondaires: « open source software (OSS) communities are internet-based networks of individual who collaboratively write software and perform many related activities like writing documentation and solving user problems" (VAN WENDEL DE JOODE, 2006). Il faut mentionner cependant KRISHNAMURTHY et TRIPATHI (2006) qui évoquent le problème d'incitation à l'écriture de documentation.

La littérature abordant ce type de thématique n'est pas aisée à dégager du fait de son caractère lacunaire. On signalera cependant pour nuancer quelques exceptions, comme ROBLES et GONZALEZ-BARAHONA (2004), SOWE et al. (2006), qui est un travail pertinent à citer en matière de gestion du savoir et d'étude des formes d'expertises ailleurs que parmi les développeurs, ainsi que, parmi les thèses récentes BERDOU (2007). Dans une perspective voisine, on peut mentionner aussi une littérature pertinente émergente, qui

appartient à ce qu'on appelle les « gender studies » ; elle étudie la question des femmes dans ces projets (LIN, 2006, METIU et OBODARU, 2008), Ce sont en effet elles qui sont concernées de façon significative par les tâches humbles dans ces communautés masculines, et ces travaux mentionnent la documentation, mais sans aller plus avant dans l'étude de l'activité en question.

Bref, il n'est pas exagéré de dire que la littérature s'est intéressée à d'autres aspects que les tâches humbles dans les communautés open-source. Des travaux supplémentaires sont nécessaires, et *pas seulement* en ce qui concerne notre thématique sur les wikis.

2.3.4. Le cas général de la question du document numérique mériterait aussi de plus amples développements

La question de l'étude des documents numériques ou des processus de documentation numérique font l'objet de quelques travaux (SALAUN, 2004, NEWELL, 2005, PEDAUQUE, 2006)[69] mais sont mis en œuvre principalement parmi les linguistes cogniticiens ou les informaticiens. Leur rattachement ou leur application au terrain d'étude de l'open-souce est parfaitement envisageable, mais elle ne va pas de soi du fait de l'importance de la « culture organisationnelle » originale de ces communautés, et l'outil wiki, négligé par ces chercheurs, comporte par ailleurs des caractéristiques techniques fortement originales (« hypertexte sous stéroïde », ARONSSON, 2002). La notion de « documentation dynamique » ne semble pas non plus avoir attiré l'attention, sauf parmi les informaticiens soucieux de promouvoir leurs applications (AGUIAR, 2005). Pourtant, pour prendre un exemple concret, le cas du « manuel de PhP » (qui n'est pas un wiki), interface qui a permis le succès de la diffusion du fameux langage dynamique, montre l'importance stratégique (ce n'est pas un abus de langage) du processus de documentation numérique pour comprendre la diffusion d'innovations *majeures* dans le domaine de l'Internet, par exemple les sites web dynamiques[70].

[69] Nous renvoyons aussi à la petite revue Document Numérique (http://dn.revuesonline.com/), qui ne propose malheureusement aucun texte pouvant être utilisable pour nos travaux.

[70] Nous devons ces informations à l'informaticien et économiste Bruno Chaves, expert dont nous reparlerons dans un chapitre ultérieur lors de la justification du choix de notre cas. Le manuel de PhP aurait pu faire aussi l'objet d'une thèse selon lui, c'est la raison pour laquelle nous ne pouvons citer de référence sur ce point.

2.4. Les limites méthodologiques et techniques de la littérature

2.4.1. De nombreux travaux déroutants

Il est attendu dans toute revue de littérature sérieuse de faire preuve d'esprit critique, de ne pas prendre les travaux pour argent comptant et d'en tester la solidité[71]. *Trois exemples révélateurs* laissent songeur (pour employer un euphémisme) et révèlent par exemple un manque de recul de ceux qui font mention des travaux ou outils antérieurs :

- L'article de Ciffollili de 2003 sur le concept de « phantom authority » dit fonder l'analyse des wikis sur la théorie des coûts de transaction. C'est l'un des articles les plus cités. Or, la lecture de l'article peut laisser sceptique le lecteur: on peut douter en effet qu'il s'agisse d'une analyse en termes de coûts de transaction, au sens *traditionnel* (« coût d'usage du *marché* »). Dire qu'un wiki facilite les contributions est pertinent sans mobiliser ce cadre théorique.
- Le travail des chercheurs d'IBM, VIEGAS et al. (2004), décrit les usages d'un outil d'analyse des wikis appelé « History Flow ». Or, l'outil, à l'état de version « bêta », testé par nous, n'est pas toujours pertinent pour la recherche scientifique: d'un usage guère commode[72], il ne peut servir qu'à des visualisations. Pourtant, il est aussi cité (ORTEGA et GONZALEZ-BARAHONA, 2007), voire réutilisé (par exemple CHENG et al., 2005).
- Enfin, un récent rapport de l'OCDE (2008) sur le Web participatif cite « wiki » dans son titre sans jamais aller au-delà d'une définition dans le texte lui-même, ce qui ne l'empêche pas d'être pertinent sur d'autres thématiques. On aurait pourtant apprécié une synthèse à portée historique minimale sur cette thématique (elle fait défaut pour le moment).

Il n'est guère utile d'aller plus loin en matière d'insuffisances, à moins de raffiner par quantification de la littérature rassemblée.

De plus, la recension des sources précises et critiques utilisées pour toutes les études mentionnées laisse deviner des limites significatives. Le lecteur ne sait pas forcément l'origine des données travaillées. Il est alors impossible de suivre le processus de recherche, sauf en ce qui concerne les travaux sur Wikipedia[73], mais aussi d'avoir l'assurance que des résultats solides ont été établis. Parler d'accumulation du savoir sur les wikis est donc parfois

[71] On peut se demander s'il n'y a pas là des inconvénients significatifs de l'open-science.
[72] Il est impossible de copier-coller les « url » des pages étudiées par exemple.
[73] La raison en est que Wikipedia et des outils proposés par la Wikimedia foundation permettent d'étudier des entrepôts de données téléchargeables.

périlleux. Des travaux où le processus de traitement des données est plus clair sont souhaitables.

2.4.2. Des résultats lacunaires en sciences de gestion, ou en matière de questionnements gestionnaires dans la littérature qui n'est pas identifiée comme telle

Dans quelle mesure la littérature en management, en particulier en management des systèmes d'information a-t-elle abordé de façon significative la problématique des wikis[74]? Nous constatons que la littérature émergente, positionnée ou non en gestion, comporte des lacunes, en particulier en MSI. D'une façon générale, peu de travaux se sont intéressés aux questions de processus et d'efficacité, i.e. une approche gestionnaire indispensable, dans les projets virtuels qu'ils supportent. Ce que nous allons relever ici montre d'un autre côté *l'étendue* des thématiques à étudier.

On trouve en gestion quelques travaux de qualité différenciée. KING (2007), ALBORS et al. (2008), GLASER et EBERSBACH (2004) en MSI, et BEAN (2005) dans les autres disciplines de gestion, sont exploratoires. MAJCHRZAK et al., (2006), PFEIL et al. (2006), WAGNER (2006), WAGNER et MAJCHRZAK (2007), GOLDSPINK (2008) en MSI sont en revanche plus élaborés, ainsi qu'en gestion LEE et LAN (2007). Au-delà, on peut difficilement classer en science de gestion des textes « limites » comme FUCH-KITTOWSKI et KOHLER (2005) qui est une recette d'intégration d'un wiki dans les processus de travail et pas du tout un article analytique. HENRIKSSON (2008) pourrait plus aisément s'y rapporter : c'est une enquête générale par questionnaires, mais sur les wikis d'entreprise. GEORGE (2007) est plus pertinent, examinant de façon exploratoire ce qu'on pourrait aisément rebaptiser des « facteurs-clef de succès ». Certains rares articles sociologiques comme JACQUEMIN et al. (2008) qui étudient de façon descriptive le « management des conflits » sur Wikipedia ont en revanche des implications en gestion. Il en est de même de KITTUR (2007), exploratoire sur les problèmes de coordination dans Wikipedia et les outils de résolution de conflits. KROWNE et BAZAZ (2004) abordent la notion d'autorité pour problématiser le « wiki style mode of collaboration ». LAWLER (2005) traite de la question de l'apprentissage organisationnel. La question de la gestion de l'hétérogénéité des contributeurs est aussi marquée, mais on se contente de constater cette diversité, pas de savoir comment elle est gérée pour aboutir à un résultat collectif, à part PFEIL et al. (2006)

[74] Un wiki fait partie des outils de Content *Management* System (CMS).

Quelques travaux abordent aussi indirectement des thématiques de « théorie des organisations au sens large » (MOULTHROP, 2002, pour l'auto-organisation, ou KROWNE & BAZAZ pour le phénomène original de l'autorité dans ces « free-form models of production», ou enfin SABEL et al. (2005) pour l'étude des mécanismes de réputation. ANDRUS (2005) approche les questions d'organisation sous l'angle *légitime* de la complexité, mais les chercheurs travaillant de façon approfondie sur cette thématique pourraient aller plus loin…

2.4.3. Les quelques thèses émergentes

On a pu mettre en évidence le caractère exploratoire et dispersé des travaux se rapportant à notre sujet. Les travaux approfondis, c'est à dire les travaux sur plusieurs années, ont longtemps été rares. Les wikis ont été des *objets de curiosité*, même jusqu'à une date avancée dans la littérature grise (KING, 2007). Cependant, deux thèses sont en cours de rédaction ; nous les présentons, bien qu'il soit délicat de décrire des travaux en cours[75].

La plus pertinente des thèses en cours est la thèse de Derek Hansen qui traite des wikis dans une communauté d'informaticiens, sur laquelle nous nous devons de nous pencher. A la différence de la nôtre, elle porte sur l'amélioration du support utilisateur (par liste de discussion) au moyen d'un wiki, et non sur une forme de documentation officielle. D'après HANSEN et al. (2007), elle porte sur une forme particulière de « mass collaborative authoring » (un concept proche du nôtre qui sous-entend le concept d'édition), l'usage des wikis dans les communautés d'informaticiens, et il a eu la même intuition que nous, à savoir coder les citations du wiki dans les listes de discussions[76]. La question qu'il pose est la suivante : les tensions liées à l'usage du savoir en ligne sont-elles résolues par l'introduction du wiki comme « repository of knowledge » dans une communauté de support utilisateurs ? Son originalité est d'examiner les formes de complémentarités entre liste de discussion et wiki, une thématique que nous retrouverons partiellement. En revanche, à part la thématique du conflit, il ne traite pas de façon approfondie de la question de la régulation de la

[75] Nous pouvons citer aussi la thèse de PAQUET (2002), mais en informatique. Il faut signaler également la thèse de Anja Ebersbach, en allemand, sur les wikis dans l'éducation, que nous ne pouvons comprendre, ainsi que celle en cours d'Andrea Forte, qui est dans la même veine mais comporte une visée très appliquée. Celle-ci a cependant permis l'écriture de communications lors de conférence que nous avons citées au besoin.

[76] Notre lecteur doit nous croire lorsque nous disons que nous avons eu la même intuition : Pierre-Jean Benghozi, notre directeur de thèse, pourra confirmer que nous lui avons montré les premières opérations de codage en septembre 2006, alors que les travaux de Hansen n'étaient pas apparus sur la Toile, sa principale communication datant de 2007.

communauté et ne se focalise pas explicitement sur l'importance du contexte de l'open-source, entre autres.

En France, une thèse est en cours, mais sur Wikipedia. Il s'agit de celle de Julien Levrel, mais nous n'avons pas accès à l'avancement des travaux, si ce n'est par CARDON et LEVREL (2009). C'est l'un des principaux travaux exploratoires en français à avoir tenté de comprendre l'originalité de la gouvernance de telles communautés de grande taille. C'est la raison pour laquelle nous citons pour une fois intégralement un passage de leurs conclusions en la matière :

> « En séparant les sanctions locales contre les contenus et les sanctions centrales contre les personnes, Wikipédia a inscrit dans son architecture même une tension interne qui oblige constamment les participants à débarrasser la discussion de l'évaluation des personnes, de leurs compétences et de leurs intentions. Le très ingénieux système de règles, à la fois incorporé dans l'interface de Wikipédia et porté par la vigilance exercée par la communauté pour assurer le respect de ses propres procédures, permet d'opérationnaliser une vertu cognitive des plus difficile dans la construction des éthiques de la discussion […]. Ce que montre la réussite du système de gouvernance de Wikipédia, c'est la force du dispositif de discussion pour encadrer, socialiser et réguler les comportements en faisant de la vigilance de la communauté le vecteur commun et la recherche raisonnable d'un consensus. »

2.4.4. Autres limites méthodologiques en matière d'étude de cas, d'analyse qualitatives, d'études longitudinales, et d'articulation entre approche empirique et outils théoriques

Nous avons regroupé ces limites ensemble uniquement parce que ces thématiques se caractérisaient par des limites.

L'une des caractéristiques de la littérature rassemblée est qu'on trouve rarement des travaux articulant données empiriques et théories existantes et des travaux permettant une théorisation (comme le propose la « grounded theory » par exemple). Quelques références seulement peuvent être citées en la matière, CIFFOLILI, 2003, pour les « coûts de transaction », PENTZOLD et SEIDENGLANZ (2006), pour une approche foucaldienne du phénomène, et FROOMKIN (2003), pour une approche habermassienne. En particulier, les approches sérieuses et affichées comme économiques du phénomène wiki sont inexistantes.

Ensuite, un cadre d'analyse est fréquemment retenu et privilégié dans les travaux de sciences sociales, à savoir l'étude de cas, qui a inspiré quelques travaux abordant les wikis (RAITMAN et AUGAR, 2005, RAYNES-GOLDIES et FONO (2005) , DA LIO et al. (2005). Cependant, en ce qui concerne le degré de profondeur des travaux rassemblés, à part les

thèses mentionnées, il n'existe guère d'études de cas respectant ce qui fait consensus en matière méthodologique, l'étude approfondie d'un terrain en prenant en compte sa complexité et celle de son contexte. Une tentative qu'il est bon de mentionner à la rigueur dans la littérature académique est WAGNER et MAJCHRZAK (2007). Leur problématique est a priori semblable à la nôtre i.e. celle de la « peer production of web content » ; mais il s'agit davantage du degré d'implication des clients d'une organisation. La perspective institutionnelle est donc différente des communautés open-source. La portée des résultats est exploratoire, et, malheureusement, l'approche comparatiste, la première du genre, rend les études de cas *individuelles* trop exploratoires, alors qu'il est préférable de commencer par là, au moins une fois. Malgré ses limites, il était impossible de ne pas citer ce travail qui se différencie nettement du reste de la littérature.

Quelques mots sur les caractéristiques des méthodologies observées. L'évaluation des travaux qualitatifs peut se faire sur plusieurs critères, comme le travail approfondi des données, en particulier le « travail du sens », la méthodologie d'élaboration des preuves et le sérieux des tests d'hypothèses, ainsi que, bien sûr, la présence de conclusions fiables. Il est d'importance car nous serons nous-même confrontés lors de plusieurs sous-questions à des approches qualitatives. La plupart des travaux sont qualitatifs mais sans une méthodologie qualitative *approfondie*. Deux récents travaux de codage des historiques d'articles méritent cependant d'être soulignés, pour Wikipedia seulement, à savoir PFEIL et al. (2006) et JONES (2008). On trouve également peu de travaux tentant une articulation qualitative et quantitative. Nous évoquerons deux exceptions, l'une centrée sur Wikipedia, EMIGH et HERRING (2005), l'autre, CABY-GUILLET et al. (2009), sur le cas plus original des wikis professionnels.

Par ailleurs, nous n'avons jamais trouvé de travaux effectuant une analyse longitudinale sérieuse. Aucun travail ne propose une approche historique ou même une perspective historique sur les wikis.

En revanche, les travaux quantitatifs sont significatifs. Mais cette thématique recoupe totalement un autre point que nous développerons un peu plus loin.

2.5. La littérature sur la gouvernance des wikis

Nous avons vu un certain nombre de limites générales sur les wikis en ce qui concerne les préoccupations des sciences de gestion. Nous allons à présent continuer sur cette lancée en

resserrant nos propos sur les questionnements abordés ici, les questions de gouvernance, introduite dans le précédent chapitre. Nous montrerons de ce fait en quoi notre travail est original.

2.5.1. Les routines : une approche exploratoire centrée encore sur Wikipedia

La question de la gouvernance informelle par des routines n'a été abordée que récemment, à la marge et d'une façon exploratoire, et la littérature sur les wikis n'a fait que très récemment allusion à une approche « évolutionniste » (au sens large). Le travail exploratoire de DEN BESTEN et DALLE (2008), l'un des premiers travaux à évoquer une approche « évolutionniste » de type « dynamic capabilities » à propos du cas des wikis, porte sur Simple Wikipedia, une version simplifiée pour les « jeunes » et qui comporte des *spécificités éditoriales*. Il évoque en creux les pratiques éditoriales et le management d'une telle communauté, les routines nécessaires à sa viabilité et souligne l'importance de notre objet de recherche : « We believe that the management and maintenance of collaborative knowledge production is soon to become a very important but difficult issue »[77]. C'est effectivement fort pertinent et c'est ce que nous observerons. Cependant, la notion de routine n'est envisagée que sous l'angle de l'automatisation, ce qui est tout de même pertinent pour renforcer l'idée d'étudier ce sujet. Ailleurs, LOUBSER et DEN BESTEN (2008) proposent une lecture moins naïve du cas de Wikipedia, et montrent les formes d'adaptations des pratiques, une perspective proche de la nôtre.

En ce qui concerne *même Wikipedia*, force est de constater que la question est sujette à *débats*, qui sont loin d'être terminés au vu du caractère *exploratoire* de ces approches. Le travail pionnier en la matière est sans doute EMIGH et HERRING (2005). Ils proposent sur Wikipedia une analyse de la relation entre degré de contrôle éditorial et modes de mise en forme du contenu : ils mettent en évidence sur cet aspect de la gouvernance une corrélation au moyen d'une analyse linguistique qualitative et quantitative. Ce travail montre qu'une étude des régularités (ici du genre et du style) et de leur potentielle diversité peut être fructueuse. VIEGAS et al. (2007) s'intéressent directement aux processus et à l'ordre caché derrière la composition des articles[78]. Il constate une rigidification des pratiques, à l'encontre de ce que

[77] DEN BESTEN et DALLE (2008).
[78] "Analyzing the organizational principles behind these emerging processes can help us better understand the inner workings of Wikipedia. The emergence of these processes is, we believe, just as interesting—and "magical"—as the emergence of high-quality articles."

l'on pense habituellement de la communauté. Les limites de l'article sont apparentes. Il n'analyse qu'*une* pratique « phare » de Wikipedia (les « featured articles »). L'analyse de données est légère : *une* interview, et une seule routine, alors qu'on peut penser que les routines font système, ce qui impose d'aller plus loin. FORTE et BRUCKMAN (2008), qui sont précisément partis de VIEGAS et al. (2007) vont plus loin, et proposent un point de vue complémentaire en montrant la difficulté à faire suivre les pratiques et insistent sur les formes d'adaptation au point de parler de décentralisation et de « self governance ». Ils inventent sans doute le concept qui est à retenir. Ce travail est seulement valable pour Wikipedia, et ils n'ont pas examiné les pratiques *effectives* au sens strict, étant donné que leur méthodologie est celle de l'entretien (dix seulement en fait). Sur ce point, leurs conclusions sont renforcées par GOLDSPINK (2008), qui montre l'importance des aspects informels et des conventions (au sens de régularité de comportement), le rôle mineur des règles, et les aspects informels d'influence. *Les communautés dont le système d'information repose sur des moteurs wikis sont largement régulées par des régularités informelles, i.e. des routines*. DI LORIO et ZACCHIROLI (2006) rappelaient d'ailleurs que l'émergence de pratiques standardisées mais informelles de rédaction des articles est à présent un fait admis. Chez certains informaticiens, dont les travaux sur les wikis sont pertinents, BESCHASTNIKH et al. (2008) étudient la problématique de l'émergence « bottom-up » de ces pratiques qui se stabilisent par consensus.

Notre conclusion ponctuelle sur cette littérature exploratoire est la présence avérée de régularités informelles de comportements. En constater nous-même ne saurait être désormais considéré comme un résultat. Cependant, le caractère exploratoire de l'étude de la thématique et la diversité potentielle des pratiques appellent à coup sûr d'autres développements, dans une perspective gestionnaire assumée… et ailleurs que sur Wikipedia. Nous y viendrons.

2.5.2. Les questions démographiques et de motivations des contributeurs

Que sait-on de l'analyse des trajectoires de contributions des participants à des wikis et de la démographie d'une communauté de ce type: qui sont-ils, que font-ils et pourquoi ?

Les articles sur la question ne sont pas nombreux, mais un peu plus aboutis, peut-être du fait du caractère quantitatif de la thématique ; nous mentionnons ceux à retenir. ORTEGA et al. (2008) montrent que les contributions aux principales encyclopédies Wikipedia (dans différentes langues) se caractérisent par une loi *constante* non plus seulement d'un point de vue linguistique, mais aussi dans le temps: 10% des contributeurs écrivent 90% des éditions

quels que soient la langue et le moment de la mesure en question. Dans la même perspective, MAKICE (2006) trouve des distributions des contributions en forme de loi de puissance. Sur ce point, les conclusions ne sont pas surprenantes pour un chercheur connaissant les caractéristiques des communautés sur Internet. Par comparaison avec ces travaux, il faut mentionner le travail plus critique d'ADLER et al. (2008) qui aborde la question d'une façon différente et technique; c'est un travail rendu accessible trop tard hélas pour que nous le prenions en compte. Il forge le concept de « longévité des éditions », qui prend en compte leur résistance aux changements ultérieurs, et non plus leur seule quantification. Cet indicateur serait plus pertinent que leur nombre pour caractériser la trajectoire de contributions des participants à un wiki. Cependant, son originalité est pertinente par rapport aux biais de Wikipedia, et pas forcément à ceux d'autres wikis.

Cette littérature prend davantage en compte le temps, ce qui méritait d'être explicitement signalé par comparaison avec les autres travaux mentionnés ici. ORTEGA et GONZALEZ-BARAHONA (2007) emploient une méthodologie originale qui va au-delà de la différenciation entre administrateurs et non-administrateurs en regardant des « patterns » d'activité, i.e. des séries chronologiques de contributions de différentes catégories de participants. Malheureusement, - ce qui est vraiment dommage étant donné que l'outil qu'ils proposent, WikiXRay, est disponible -, nous ne pouvons transposer cette méthode à notre cas, ne disposant pas des entrepôts des historiques (« dumps ») du wiki de la communauté WordPress. KITTUR et al. (2007) différencient sur Wikipedia deux populations initiales : l'une, l'élite, se chargerait de l'essentiel du travail, l'autre, la « foule sage » (« wisdom of the crowd »), d'une partie secondaire des éditions. Ce travail montre que la charge de travail passe d'une population à l'autre au fil de l'histoire de la communauté. La mise en évidence de cette évolution, dite « montée de la bourgeoisie », par sa prise en compte de la dynamique d'une communauté est a priori originale. Ensuite, DEKEL (2007) un des rares articles ne portant pas sur Wikipedia, est à souligner. Il s'attaque à un aspect du problème voisin du nôtre, le comportement direct des utilisateurs des wikis ; il propose une technique permettant de reconstituer à partir de plusieurs types de traces un profil utilisateur (il ne donne que la technique, et non des résultats).

Pour le moment, la littérature ci-dessus montre que l'analyse des trajectoires des contributeurs à un wiki n'a jamais été sérieusement effectuée ailleurs que Wikipedia. Même là, des travaux supplémentaires seraient bienvenus. Dans d'autres communautés, le nombre de clics pour certains gros contributeurs est dissuasif. Analyser des historiques suppose aussi de réfléchir aux variables à étudier pour simplifier une telle masse de données.

Que sait-on des motivations des contributeurs ? La question des motivations des contributeurs fait l'objet d'une littérature en ce qui concerne les communautés open-source. Elles se posent aussi pour les communautés de contenu ouvert qui sont apparentées, mais pas de la même façon d'après les premiers travaux qui n'ont cependant qu'une portée exploratoire[79].

Pourquoi contribuer à un wiki ? Plusieurs articles s'intéressent aux motivations des contributeurs aux Wikis, surtout à Wikipedia. Lors d'une étude encore exploratoire, WAGNER et PRASARNPHANICH (2007) montrent que pour le cas de Wikipedia, les motivations altruistes domineraient *et se différencieraient nettement de ce qu'on observe dans le modèle open-source*[80]. SCHROER et HERTEL (2009) tout en mettant en évidence la pluralité des motivations montrent que les caractéristiques des tâches effectuées (autonomie, sens du travail effectué, variété des compétences requises) sont aussi pertinentes ; or, les caractéristiques des tâches sur un wiki sont *différentes* de celles de la programmation. Le problème de ces travaux est qu'ils ne sont valables que pour des contextes institutionnels localisés et ne sont pas forcément transposables ailleurs (pluralité des cités évaluant les actions d'autrui). De plus, les contributeurs sont différenciés, et les tâches et motivations d'un administrateur et d'un contributeur occasionnel pour prendre des cas polaires ne sont pas les mêmes.

CROWSTON et FAGNOT (2008), qui proposent une analyse plus large qui va au-delà des wikis, distinguent trois niveaux de contributions, « initial », « sustained » et « meta », et explicitent un certain nombre de propositions quant aux motivations effectives de ces différentes catégories de participants. Mais elles restent à tester. MUNSON (2008) est aussi à signaler, car c'est l'un des rares articles portent sur un terrain autre que Wikipedia : il discute de façon exploratoire l'effet motivationnel du wiki lui-même. Encore une fois, l'outil technique comporte un effet structurant qui justifie son originalité.

Enfin, signalons qu'une petite littérature commence à problématiser des formes de motivations surprenantes à la constitution de bien collectifs de cet ordre. POLBORN (2007) montre que cette constitution, y compris sur Wikipedia, serait due à la compétition entre groupes (« clubs ») pour la reconnaissance dans une communauté, reconnaissance qui se fait par la contribution. GEORGE (2007) évoque non pas l'absence d'incitation mais montre que

[79] Il y aurait aussi certainement à dire pour comparer les méthodologies d'étude de l'open source et des communautés open-media, mais nous ne pouvons développer ce point ici.

[80] Ce point mériterait d'être creusé, la question des motivations des développeurs, quoique traitée, mériterait des travaux plus systématiques. Nous pensons qu'il faut prendre ces conclusions avec prudence.

l'un des problèmes est d'arriver à maintenir un degré satisfaisant de contributions tout en ne *dissuadant* pas de façon rédhibitoire les principaux gros contributeurs bénévoles. Il trouve le même résultat que REAGLE (2007) qui montre que des individus peuvent aussi contribuer pour gagner un statut dans la communauté Wikipedia. BRABHAM (2008) aborde ainsi les motivations des participants au mouvement « crowdsourcing » (qui n'est pas notre terrain exactement) qui relèveraient de la reconnaissance par les pairs. ZHANG et ZHU (2009) aboutissent à une relation totalement opposée à ce qu'on savait de coutume sur la formation de biens collectifs, à savoir que la motivation à contribuer diminuait avec la taille de la population. Ce type de modèle contre-intuitif en économie, s'il est vérifié, serait sans doute une avancée significative en sciences sociales[81].

2.5.3. L'analyse quantitative des formes de collaboration au sens large et de la structuration du fonctionnement des wikis

Pour cette sous-question de recherche, *la prise en compte de la littérature antérieure n'est pas triviale*.

D'une part, elle relève de deux littératures différentes, l'une sur la structuration des wikis, l'autre sur la structuration des listes de discussions, ce qui n'était pas prévu au départ et qui contraste avec l'orientation générale de la présente revue de littérature. D'autre part, il s'agit d'une littérature au caractère technique[82]. Il est possible que les résultats soient de ce fait plus solides scientifiquement en matière quantitative qu'en matière qualitative.

Le wiki est souvent classé parmi les « outils relationnels », ce qui justifie l'analyse de formes de collaboration. CROWSTON et HOWISON (2004), insistent sur la diversité des formes de centralisation/décentralisation dans les projets open-source, et on peut penser que des formes de diversité se retrouvent dans les parties des projets comportant un wiki. Comment qualifier et étudier les modes d'interaction entre agents sur ces outils, par ces outils, ou au-delà de ces outils ? La question des interactions entre contributeurs à un wiki, mais pas forcément uniquement *sur* le wiki, a-t-elle fait l'objet de travaux sérieux ? KLEIN et IANDOLI (2008) soulignent l'idée d'échec des wikis en matière de délibération, précisément.

[81] Sur ces questions « politiques », on peut renvoyer aux travaux sur le langage de Jean-Louis Dessalles (DESSALLES, 2006, 2007), en particulier sur le langage comme « costly signaling ».

[82] Sur ces questions, il est possible que nous ayons été surclassés techniquement, peut-être du fait de la dispersion des méthodologies adoptées ailleurs dans cette thèse. Mais ce n'est pas forcément rédhibitoire

La thématique des réseaux et de la collaboration comporte deux aspects, les tentatives de modéliser les relations à l'intérieur des wikis, mais aussi à l'intérieur des listes de discussions. D'un autre côté, on le sait, l'analyse de réseaux, en particulier la SNA (« social network analysis ») fait l'objet d'un intérêt pour bien des chercheurs. En sciences de gestion, l'approche devient à la mode (THIETART et al., 2003)[83].

Une première littérature porte sur l'analyse quantitative de la structuration des wikis.

En matière d'analyse de la collaboration *sur* un wiki, plusieurs questions peuvent être posées : de quels résultats disposons-nous ? Quelles sont les limites de telles approches pour aborder la question des « relations » ?

Les travaux en la matière comportent des limites. Ceux de HARRER et ZEINI (2006) ou de PEREIRA et SOARES (2007), sont exploratoires. Ce dernier est directement appliqué (préconisation de l'outil SnetCol). On trouve quelques travaux sur les modes de croissance des wikis et de fait leur structuration. Ainsi, CAPOCCI et al. (2006), montrent le rôle du mécanisme d'attachement préférentiel pour comprendre la croissance de Wikipedia. La croissance et la structure du célèbre wiki ressemblent à celle du reste de la Toile (structure en « nœud-papillon »), bien que le mécanisme qui le sous-tend soit différent. BELLOMI et BONATO (2005) étudient la structure globale et l'effet des biais culturels dans la structuration de la communauté (tropisme occidental, valable au minimum pour avant 2005). Signalons que plusieurs travaux sont écrits par des physiciens (par exemple ZLATIC, 2006, pour un travail en physique statistique). Ces travaux sont cependant frustrants pour un chercheur en sciences sociales, qui ne voit pas toujours en quoi la recherche de propriétés structurelles universelles aide à comprendre les spécificités gestionnaires de tels phénomènes. On observe aussi des recherches strictement quantitatives (à l'exclusion de l'analyse de réseau) dans la lignée de VOSS (2005): les chercheurs sont par exemple à la recherche sur Wikipedia de régularités de type loi de puissance et scale-free, comme sur le Web en général, tout en tentant de conceptualiser les caractéristiques des phases de croissance et de maturité de la communauté et de son produit (BURIOL et al., 2006). Il s'agit des recherches parmi les plus abouties sur la question des wikis, avec des méthodologies lourdes. On peut cependant rester dubitatif devant ces travaux : repérer des distributions en forme de loi de puissance est-

[83] Les notions de complexité et de réseaux ne posent pas de problème en management des systèmes d'information. Cependant, ces questions soulèvent des difficultés spécifiques, et les résultats pour le gestionnaire en particulier ne vont pas toujours de soi. Nous renvoyons par exemple à une présentation de vulgarisation de BENKLER (2009).

il toujours significatif ? Repérage de régularités structurelles n'équivaut pas forcément à étude de la gouvernance.

Quelques travaux sont plus proches d'une analyse des formes de collaboration. KORFIATIS et al. (2006), un travail d'informaticiens, permet aussi d'évaluer la crédibilité de chaque contributeur en fonction de sa connectivité. De même, KORFIARTIS et NAEVE (2006) proposent une analyse de l'activité et des contributions par une approche réseau, et tentent de conceptualiser la crédibilité de chaque auteur en mesurant la centralité de celui-ci dans le réseau des contributeurs. Cette approche reste exploratoire et localisée, et la crédibilité mesurée par une approche de cette sorte trouve ses limites[84]. HARRER (2008) aborde la construction collective du savoir. TANG (2008) est probablement l'un des travaux les plus aboutis d'un point de vue technique. Il propose un algorithme de mise en évidence des co-auteurs d'un article, ce qui est une façon pertinente d'envisager des modes de collaboration. Malheureusement pour nous, cet article arrive trop tard pour que nous puissions nous former à ce type d'approche, et il n'est pas sûr de toute façon qu'on puisse l'adapter à notre cas. Enfin, il ne s'agit que d'un outil, et les auteurs ne proposent pas de résultats. Mais ce travail méritait d'être souligné malgré ses limites. La portée de l'analyse de réseau pour les wikis a fait l'objet de discussions (VOSS, 2005) et un programme de recherche (AutoGraph) montre la difficulté d'une telle approche.

Là encore, les résultats de ces travaux ne sont valables que pour Wikipedia.

Une deuxième littérature porte sur la structuration des listes de discussions.

Notre étude de la structuration de la gouvernance par une approche quantitative se focalise sur l'étude d'une liste de discussions. Curieusement, contrairement à nos premières attentes, il n'existe pas autant de travaux pertinents que prévu sur ce type de matériaux[85].

De nombreux articles sur la question sont exploratoires, comme CONEIN (2004) qui propose une approche de la structuration des fils de discussions qui repose sur des théories de cognition sociale, CHEN et al. (2006) qui tentent un classement des participants à une liste de discussions, ou enfin KROGH et al. (2003) pour tout ce qui touche l'analyse des fils de discussions, analysée de façon quantitative mais exploratoire.

KARINBEGG (1999) quoique ancien, est plus pertinent : il porte sur deux caractéristiques significatives des listes de discussions, à savoir la participation et le

[84] Dans notre communauté, une approche plus qualitative des compliments, du nombre des contributions, du nombre de relations et du nombre de thématiques abordées serait plus complète.

[85] Traiter ce type de littérature nous avait aussi été recommandé oralement par Paul David, lors d'un séminaire informel de la Chaire Innovation et Régulation des services numériques en 2008.

management. Ce type d'approche gestionnaire a été par la suite négligé d'après nos recherches, à part MATSUMURA et SASAKI (2005), sur le double rôle du leader (attraper les messages et en poster) et ses implications managériales dans ces listes de discussions et SOWE et al. (2007), qui étudient la question du transfert de connaissances entre développeurs et non-développeurs par une approche plutôt quantitative[86].

D'autres travaux sont plus quantitatifs et modélisateurs. WEISS et al. (2006) proposent des tests d'hypothèse du mécanisme d'attachement préférentiel dans la liste de discussion du projet Apache. DORAT et al. (2007), l'un des travaux les plus poussés en la matière, est une analyse raffinée des propriétés topologiques du réseau associées à la liste de discussions de Debian (loi de puissance, forme de centralité, etc.). BARCELLINI et al. (2005) croisent un mélange d'analyse de citations et de structure des fils de discussion.

Nous signalons en particulier que l'une des limites de ces littératures, qui nous concerne au premier chef est le manque d'analyses longitudinales, et donc historiques : c'est la voie vers laquelle nous allons nous orienter, tout en nous inspirant de ce qui précède. Pour nous, l'analyse quantitative au sens large n'est qu'un outil avec une finalité appliquée locale au chapitre 7, et pas une fin en soi comme dans certains des articles mentionnés, que ce soit sur le plan technique ou en matière d'analyse fondamentale des réseaux. Elle ne nécessite pas forcément une élaboration poussée, et la nature des données ne le permet pas forcément[87].

2.5.4. Appropriation, usage et efficacité d'un wiki

C'est là l'une des questions les moins aisées à propos des wikis, qui occasionne des débats parfois passionnés. C'est ce que suggère par exemple le caractère exploratoire de certains travaux (HESTER, 2008, NEUS, 2001). De quelle façon la littérature a-t-elle de près ou de loin conceptualisé l'appropriation, l'efficacité, la qualité, etc. du contenu d'un wiki ?

Là encore, la littérature scientifique s'est focalisée en majorité sur Wikipedia (CHESNEY, 2006, ADLER et al., 2008, ARAZY et al, 2006, STVILIA et al., 2005, WILKINSON et HUBERMAN, 2007, ZHANG, 2006). La préoccupation de ces études est de

[86] Cet article a aussi le mérite de découvrir d'autres phénomènes (de nature fractale) plutôt que des lois de puissance.

[87] La forme d'un réseau (présence de boucles ou de fortes connections entre certains utilisateurs) interdit d'effectuer certaines mesures élaborées d'après notre expérience en matière de logiciel et des discussions avec Carla Taramasco, doctorante au CREA.

tester la fiabilité du produit de la collaboration massivement distribuée : par exemple, des chercheurs évaluent-ils favorablement le contenu de l'encyclopédie (CHESNEY, 2006) ? Cependant, leur différenciation en fonction d'un « lectorat » pluriel n'a pas été envisagée (à part ZHANG, 2006, qui montre la spécificité des modes de lecture de Wikipedia), que ce soit dans le domaine des wikis de la création open-source ou non.

D'après ce que nous avons pu rassembler, plusieurs approches peuvent être distinguées pour aborder, même de façon incomplète, la problématique de la « réception » d'un wiki.

Certains auteurs tentent d'expliquer le niveau de qualité. L'optique est en général quantitative. Pour WILKINSON et HUBERMAN (2007), la structure d'éditions de Wikipedia suit un processus stochastique de type « boule de neige », ce qui a des implications en matière de qualité des articles. ADLER et al. (2008-2) tentent une approche de l'évaluation du contenu en fonction de deux variables (caractéristiques de l'historique de l'article et réputation des contributeurs.), que nous ne mettrons pas en place dans cette thèse par manque de temps, l'adaptation de la méthodologie au contexte de l'open source supposant une réflexion que nous n'avons pas le temps de mener. WILKINSON et HUBERMAN (2007) met en évidence une corrélation entre le nombre d'éditions, nombre de participants différents et niveau de collaboration et qualité de l'article[88]. ARAZY et al. (2006) examinent l'origine du niveau de qualité de Wikipedia, et plus précisément le lien entre nombre et diversité des participants et qualité des articles. KITTUR et al. (2008), propose de mettre en évidence les effets (positifs) d'indicateurs de qualité d'articles de Wikipedia. KRAMER et al., (2008) proposent de définir des indicateurs de qualité généraux pour les wikis à partir des caractéristiques du texte et des auteurs, et des façons de les mettre en place.

Une autre voie est l'évaluation des articles par des spécialistes des domaines concernés. Dans la lignée de l'article remarqué de *Nature* qui compare l'encyclopédie collaborative à une encyclopédie sur papier, CHESNEY (2006) par exemple propose une évaluation de la crédibilité des articles de Wikipedia en mobilisant des experts du domaine des articles évalués qui confirme que le résultat de la collaboration massivement distribuée peut être évalué positivement.

[88] Cela dit, nous avons des doutes sur la validité et la généralité de ce modèle : certains articles avec nombre d'éditions, contributeurs, et collaboration peuvent être sur Wikipedia des articles controversés (par exemple ceux sur le 11 septembre 2001). Il est très probable que ces travaux doivent être nuancés pour le cas de Wikipedia, et que leurs conclusions ne sont pas transposables à d'autres contextes institutionnels.

KRAMER et al., (2008), cités précédemment, mentionnaient les limites de leur approche[89], ce qui permet d'introduire le point suivant. Une dernière voie de recherche est en effet celle de *la problématisation de la réception par les utilisateurs eux-mêmes*. Paradoxalement, c'est celle qui a été le moins traitée. Sur ce point, le principal travail à retenir (et mentionné ici plus haut) est HANSEN et al. (2007), qui a eu en même temps que nous l'idée d'étudier les traces d'usage à partir de mots-clefs. Nous irons nous aussi dans cette voie.

2.6. Conclusions

Nous avons donc montré que notre travail de recherche était justifié par la limitation de la littérature sur les points que nous avons dégagés, au-delà des biais que nous avons rassemblés (focalisation sur Wikipedia, manque de perspectives gestionnaires, négligences des parties périphériques des communautés open-source…). En effet, pour ce qui concerne nos sous-questions de recherche et notre thématique de thèse plus généralement, il reste des recherches à mener malgré le développement récent de la littérature dont nous avons souligné certains contours : la littérature concernant les caractéristiques et motivations des contributeurs est peu fournie, celle concernant les formes de routinisation est exploratoire, et celles concernant l'analyse des formes de collaboration et de qualification du produit informationnel plus ou moins dispersées en différentes approches. D'une façon plus générale, il existe nombre d'inconnues sur la gouvernance des wikis, et pas seulement ceux appartenant aux communautés open-source. Sur les autres insuffisances, nous avons lancé des pistes de travail (il était utile de le mentionner en passant). Le seul point sur lequel la littérature rassemblée a sans doute abouti à une maturité significative durant notre travail de thèse était sur une partie des aspects quantitatifs, mais cela ne rendra pas inutile notre travail.

Nous présentons enfin les limites de ce chapitre. Dans une perspective qui étendrait ce travail à la totalité de ce qui a été écrit sur les wikis, une approche par codage et catégorisation systématique des articles serait utile, afin de tracer plus systématiquement nos conclusions. Cela permettrait aussi de traiter la littérature selon une perspective dynamique et quantitative,

[89] "Another point we have emphasized is that automatic analysis of history cannot totally substitute for human rating. Only humans can determine if the article is complete, consistent, neutral, and has adequate citations, among other quality aspects. Since readers typically far outnumber authors, capturing and publishing their feedback can potentially enhance trustworthiness."

afin de tester telle ou telle hypothèse sur les insuffisances du corpus rassemblé, ce qui peut être utile à d'autres chercheurs qui auraient besoin d'effectuer des « requêtes », ou un historique de l'évolution des thématiques étudiées. Mais le présent chapitre exploratoire et orienté a une utilité pour donner des directions de travail, y compris en faisant allusion à des thématiques significatives, mais négligées dans cette thèse. Par ailleurs, nous ne prétendons pas avoir traité toutes les limites de la littérature. Ainsi, nous n'avons pas examiné par exemple dans quelle mesure la littérature a abordé la diversité des moteurs wikis[90] et les limites associées. Or, le type de moteur peut comporter un rôle structurant dans les communautés.

[90] Les principaux moteurs cités ici sont MediaWiki et Twiki, mais il y en a d'autres (MoinMoin, TikiWiki, DokuWiki, etc.)

Chapitre 3 : une approche historique des archives d'une communauté

3.1. Introduction

Notre présent objectif, au-delà de la présentation attendue d'archives pouvant exciter la curiosité de lecteurs néophytes en la matière, est ici d'ordre méthodologique et technique. *Nous souhaitons montrer que la démarche historique est ici indispensable*. Nous voulons souligner l'intérêt d'une telle approche spécifiquement historique, plus « ancienne » pourrait-on dire, sur des archives d'une certaine nouveauté où un chercheur n'y aurait pas forcément pensé[91]. Plusieurs problèmes essentiels de nature historique peuvent être abordés: le traitement des archives (listes de discussion, historiques de wiki...) comme matériau brut, le repérage des sources pertinentes, les problèmes associés aux traitements qualitatifs des "traces" qui sont le seul matériau disponible pour répondre aux questions que nous nous posons[92].

Nous allons donc aborder la question en profondeur, mais nous ne descendrons pas dans les détails techniques et les *protocoles* de traitements de données, l'ingénierie de la thèse : des parties méthodologiques plus spécifiques seront développées brièvement au début de chaque sous-question de recherche (saisie, codage, comment analyser telle liste de discussion, les protocoles d'analyse qualitative, l'analyse de réseau, l'analyse quantitative, etc.).

Nous allons d'abord présenter notre cadre d'analyse, *l'étude de cas*, compatible avec une approche historique, tout en justifiant le choix de notre cas.

Nous expliciterons nos choix analytiques, en montrant qu'au-delà des approches habituelles, une démarche de type historique est requise, et permet de caractériser plus correctement nos pratiques de recherche. Nous mettrons en évidence l'originalité d'une telle démarche ainsi que le choix du genre « micro-histoire ».

Nous expliquerons le choix de nos données, nous les présenterons, et mettrons en évidence les biais que nous pouvons attendre lors de leurs traitements.

[91] Nous renvoyons à RUZE (2009) pour se faire une idée de la portée de la prise en compte de ce genre de matériau primaire par les historiens et les chercheurs en gestion intéressés par l'approche historique.

[92] Nous précisons donc ici à l'avance que notre démarche a été en tout point « non-intrusive », contrairement à d'autres travaux en sciences de gestion, et la communauté WordPress ignore qu'elle fait l'objet d'une recherche.

3.2. Les fondements du choix de notre étude de cas

3.2.1. Justification du choix de l'étude de cas

Bien que des études de cas aient été fortement demandées dans le cadre du projet PERSI sur l'analyse des réseaux sociaux de l'Internet[93] auquel nous avions participé, mener une étude de cas n'était pas ce qui était prévu au commencement de notre thèse. Nous avons ainsi restructuré en totalité notre sujet à notre arrivée au CRG en septembre 2005 ; mais cette orientation est aussi due à la prise de conscience de la difficulté de le traiter.

Bien que valorisée par plusieurs ouvrages de référence (YIN, 2003, RAGIN et BECKER, 1992, par exemple) et opérant un retour dans l'ensemble des sciences sociales (PASSERON et REVEL, 2005), cette approche a fait récemment débat lorsqu'elle est employée dans ces disciplines au point de demander des justifications supplémentaires (GEORGE et BENNETT, 2005) ou un niveau d'exigence accru (DUMEZ, 1999).

YIN (2003) affirme que l'étude de cas trouve sa raison d'être si les objets étudiés sont complexes (on le devinera sans difficulté dans le cas d'un gros wiki), si le phénomène observé comporte une dimension dynamique et enfin si le phénomène est étudié « en son contexte ». En ce sens, le cadre de l'étude de cas pour l'étude du wiki de WordPress dans son contexte communautaire et sectoriel a plus de sens que pour Wikipedia. Dans une optique semblable, PASSERON et REVEL (2005) justifient cette approche par « le souci d'une interprétation circonstanciée des réalités » et par l'idée de « prétendre généraliser » à partir de « configurations singulières ». C'est ce que nous tenterons.

Cependant, on peut nous faire quelques objections. YIN (2003, page 3) pense en effet que le cadre de l'étude de cas et la méthodologie historique ne sont pas complémentaires, mais en concurrence. Il est tentant par curiosité scientifique de montrer le contraire. De plus, notre approche de l'étude de cas n'est pas seulement qualitative, ce qui n'est pas le plus fréquent mais ne pose pas de problème décisif. Les exigences de cette approche sont les suivantes : plusieurs sources, croisement de données, plusieurs théories en concurrence ou complémentaires, mise en évidence des limites propres des théories, plusieurs outils d'analyse, possibilité de généralisation à partir des cas, analyses profondes ou fines des phénomènes étudiés.

Concrètement, nous répondons à un manque d'étude de cas sur les wikis (nous renvoyons au chapitre 2 sur ce point).

[93] http://www.ehess.fr/ldh/Atelier_ADR/recherche/projet_PERSI/PERSI.html (consulté le 13/06/09).

3.2.2. Justification du choix du terrain, la communauté WordPress

Le choix d'un terrain lors d'une recherche est toujours, au moins dans une certaine mesure, un pari. Nous défendons ci-dessous qu'il a été ici gagné. Notre choix s'est révélé en effet par la suite judicieux, étant donné qu'il s'agit d'un logiciel qui a connu un grand succès et pour lequel la communauté des utilisateurs était de première importance. De plus, le wiki s'est révélé viable pendant une durée significative.

A l'origine, il a été sélectionné parmi quelques autres cas plus décevants[94] sur les conseils éclairés d'un « marginal-sécant » pertinent à notre connaissance, un expert à la fois en économie et informatique, Bruno Chaves, doctorant à Paris 10[95].

Il s'agit par ailleurs d'une communauté réputée dans le domaine de l'open-source. WordPress est le principal moteur de blog open-source[96]. Framasoft et de nombreuses organisations connues l'utilisent aussi (voir infra). WordPress a été choisie par Business Week comme le moteur de blog de l'année 2006. Le site open-source « CMS Award » l'a primé de la même façon en 2007. C'est aussi une « success story du web », étant donné que la jeune pousse (« start-up ») associée est en pleine croissance. En 2007, le logiciel a été téléchargé 2,845,884 de fois[97]. Une requête avec Google à « WordPress » donne en juin 2009 560 millions de pages référencées. Il s'agit désormais d'une vraie marque, et pas seulement d'une communauté.

Le wiki de WordPress a été seulement cité par de rares chercheurs qui travaillent de façon récurrente sur la question des wikis (DESILETS et al., 2005), même si ce travail ne nous a pas servi d'inspiration directe. Le wiki de WordPress a servi de modèle à (au moins) un autre projet, Gallery 2, ce qui montre que son intérêt est perçu[98]. Le succès du logiciel a abouti à la rédaction de documentations concurrentes sous forme de livres[99].

Ce cas se distingue par sa durée de vie, ainsi que par le nombre de phases par lesquelles la communauté est passée : un premier wiki, une phase de croissance, de crise, de reprise et d'arrivée à une relative maturité, suffisante pour être citée comme source de

---

[94] Le wiki de Tela Botanica, le wiki d'Objectweb…

[95] Ce dernier s'intéresse à une autre forme proche de knowledge management, qui n'est cette fois pas un wiki, celui sous-tendant le « manuel de PhP », qui a permis la diffusion de ce code informatique novateur pour la constitution de sites dynamiques. Nous envisageons de collaborer sur ce projet après la thèse

[96] Un biais francophone peut faire que notre lecteur connaisse davantage Dotclear. Mais la communauté française de WordPress est dynamique.

97 Source : http://wordpress.org/about/ (consulté le 10/06/09)

98 La communauté WordPress le sait (source : http://lists.automattic.com/pipermail/wp-docs/2006-December/001383.html).

99 Source : http://www.packtpub.com/wordpress-books (consulté le 10/06/09).

référence par les manuels de documentation du logiciel publiés ultérieurement sous forme imprimée. Toutes les communautés ne disposent pas d'une histoire « riche ». On rappellera au besoin que la plupart des projets open-source et la plupart des wikis meurent vite. Cette communauté a conscience de sa qualité et de l'importance de celle-ci, comme le montre ce passage d'un administrateur qui signale ce point sur la liste de coordination du forum:

> "I'm sure some have seen this already, and it applies through the WP world but it also applies to us. It's damn good :)
>
> "Different blogware packages feature varying levels of community support and documentation. Because the installation of some types of software requires an advanced knowledge of computer administration, community support and documentation can be very helpful. The web servers and database software can be more difficult to install than the blogware itself. Also a strong and active community surrounding the blogware gives advice on integrating the blogware into a personal site. WordPress community support is generally regarded as one of the best available."
>
> http://en.wikipedia.org/wiki/Blog_software »[100]

Cette communauté offre une richesse de données remarquable pour une étude de cas au point de ne pas nous laisser le loisir de tout exploiter: trois wikis simultanés ou successifs, dont le principal s'appelle le Codex, les archives relativement complètes des listes de discussions, un gros forum, des blogs de développeurs, des archives des logs des « chats » IRC, etc. (voir annexes pour l'ensemble des sources). La multiplicité des sources offre de ce fait la possibilité d'étudier la spécificité des wikis par comparaison ou en complémentarité avec d'autres modes de coordination ou de gestion du savoir, cités ci-dessus. L'hypothèse vraisemblable selon laquelle ces outils ne peuvent être efficaces seuls est aussi plus aisée à évaluer. Nous avons aussi choisi un projet comportant plusieurs systèmes d'information, ce qui permet de souligner l'aspect organisationnel de l'étude des wikis, et de son efficacité spécifique, ainsi que les limites de cette efficacité.

L'un des *critères décisifs* qui a amené au choix de cette communauté a été la présence d'une liste de discussion de coordination de l'activité de documentation. Nous avons pensé, ce qui a été vérifié par la suite, qu'il s'agit, en première approximation, d'un « révélateur » pertinent pour identifier les problèmes de gestion dans cette communauté. La présence de cette source absolument essentielle a été *déterminante* pour effectuer le choix du terrain d'étude : en effet, d'après nos recherche, l'usage d'outils de coordination extérieure au wiki

[100] Source : http://lists.automattic.com/pipermail/wp-forums/2006-April/002133.html .

n'est pas si courante, quoique mentionné dans la littérature (ROTH, 2007). Des recherches sur une telle question seraient *une priorité à coup sûr*[101].

D'autre part, parmi les études de cas possibles, il a été jugé bon de choisir un cas de wiki supporté par MediaWiki. Ce moteur est utilisé pour sa clarté et sa facilité d'utilisation pour les novices, et c'est un bon moyen de voir les différences par rapport à Wikipedia. De plus, son usage se répand, de plus en plus de projets passent d'un moteur donné à MediaWiki, entre autre à cause du modèle de Wikipedia, l'archétype de la collaboration massivement distribuée. Nous avons aussi choisi le moteur de Wikipedia, afin de neutraliser la variable technique dans notre problématique et mieux différencier les pratiques par rapport à ce cas limite.

Ensuite, il était nécessaire d'éviter les projets caricaturaux, centralisés, et gigantesques, comme Linux, Apache, Debian, Mozilla, qui ne sont pas forcément représentatifs de toutes les formes organisationnelles observées qui se répartissent plutôt sur un continuum entre « cave » et « communauté » (KRISHNAMURTHY, 2002), et qui ont été fréquemment étudiés[102]. Il fallait d'un autre côté que la communauté étudiée ait une taille satisfaisante pour faire l'objet d'une thèse, ce qui élimine nombre de petits projets, ainsi qu'une durée de vie conséquente. Nous nous différencions ainsi en étudiant un projet qui supporte un logiciel de taille moyenne de type logiciel de gestion de contenu (Content Management system), dit CMS, un type de logiciel répandu[103], mais qui n'a quasiment pas fait l'objet de travaux de recherche. Le seul cas que nous connaissons est le cas du logiciel SPIP, étudié par DEMAZIERES et al. (2007). De ce fait, la difficulté technique du projet est moindre que dans d'autres projet avec des développeurs de très haut niveau, comme les grands « debianistes », ce qui est bienvenu sur des terrains d'études déjà suffisamment complexes.

Enfin, le choix du logiciel produit par la communauté, à savoir un moteur de blog, n'est pas anodin non plus, même parmi les logiciels libres, dont on connaît la philosophie sociétale : les blogs forment *plus de la moitié* des sites web, mais sont également l'un des vecteurs essentiels de *la libre expression dans nos sociétés démocratiques mais aussi moins*

---

[101] Nous pensions aussi avoir un avantage en la matière en nous différenciant de Wikipedia, mais cette hypothèse s'est révélée fausse par la suite. Wikipedia est aussi coordonnée par des listes de discussions, même si elles ne sont *pas du tout de la même nature*, dans la mesure où il ne s'agit aucunement de la coordination de la mise en forme du contenu. Cela n'a aucunement été gênant, mais il était honnête de le mentionner.

[102] Nous renvoyons pour cela par exemple à la base de données du MIT qui répertorie les principaux textes scientifiques qui s'intéressent au logiciel libre (source : http://opensource.mit.edu/online_papers.php?lim=1000 , consulté le 10/06/09)

[103] Les logiciels de gestion de contenu, ou « content management system » (CMS) sont l'une des grandes catégories de logiciels proposées par Framasoft (source : http://www.framasoft.net/, consulté le 10/06/09).

démocratiques, en particulier la Chine, l'Egypte, l'Iran[104], et des contributeurs à la communauté que nous étudions ont parfaitement conscience de l'importance sociétale de leur rôle. La tradition du CRG est d'aborder des problématiques sociétales : le laboratoire se positionne dans ses travaux de recherches à la frontière des sciences de l'ingénieur (ici les nouvelles technologies), des problématiques sociétales, et des sciences sociales (pour nous, histoire et économie). Le choix d'un tel terrain (peut être inconscient) n'est que *la continuation en un sens d'une tradition de recherche*. Si nous étudions une forme moderne « d'atelier d'imprimerie », nous en examinons le système d'information. La problématique de constitution d'un bien collectif, d'un savoir public, est conforme aussi à cette tradition. On notera aussi l'importance contemporaine des wikis pour supporter l'innovation technologique et sociétale (QUIGGIN, 2006).

3.2.3. Justification et conséquence du choix d'un cas unique avec plusieurs items, sélection des unités d'analyse et du découpage temporel.

Nous avons choisi de n'effectuer qu'une étude de cas dans le cadre de cette thèse. Plusieurs raisons nous y ont amené : selon YIN (2003), l'accès à des données privilégiées *justifie* à elle seule une étude de cas. Il s'agit de ce qu'il appelle un « revelatory case », c'est-à-dire un cas où le chercheur a l'opportunité de ce fait d'observer des phénomènes auparavant inaccessibles ; nous nous sommes déjà justifié sur ce point plus haut. D'autre part, une analyse approfondie suppose le rassemblement de données conséquentes et pénibles à traiter ou interpréter, ce qui est *conséquent* pour un thésard isolé sur une période limitée. L'histoire de cette recherche et les contraintes associées ont fait que, dans le temps imparti, il était ardu de faire plus. Notre idée est d'étudier la question de la gouvernance selon plusieurs optiques, comme on le verra, et cela imposait en retour de ne considérer qu'un ensemble singulier. Etudier sérieusement en profondeur une communauté, et en particulier un de ses « organes », la documentation, en interaction avec d'autres « organes » lors d'une thèse autorisait à n'étudier qu'un cas. Notre revue de littérature a aussi montré que l'étude *sérieuse* d'un seul cas, hors Wikipedia, qui est un cas unique à lui tout seul, est rare. La tradition du CRG où des études de cas uniques ont été effectuées et encouragées est une autre justification.

[104] Nous ne développons pas ces faits connus, nous supposons que notre lecteur lit la presse quotidienne généraliste à grand tirage.

Ensuite, nous avons plus particulièrement fait le choix de ce que Yin appelle « single case embedded design » avec « multiple units of analysis ». Ces unités d'analyse peuvent être soit nos sous-questions de recherche, soit nos terrains, le wiki, les listes de discussions, les archivages des interactions par IRC, le forum.

Enfin, quelques mots sur le choix du découpage temporel. La question est d'importance, quoique ordinairement négligée. Ici, elle est simplifiée. Le choix du terrain s'est effectué définitivement début 2006. Notre étude porte sur la période qui va de mars 2003, date de naissance approximative de la communauté, à décembre 2008. Nous avons étudié la communauté sur la durée la plus longue possible sur la durée de notre thèse. Mais nous devions prendre en compte la nécessité de rédiger notre thèse en laissant de côté les matériaux étudiés. D'autre part, en ce qui concerne la documentation, l'observation de l'évolution de la liste de discussions montre qu'une nouvelle phase pourrait s'ouvrir à partir de janvier 2009, comme on le verra dans le chapitre d'histoire de la communauté, ce qui clôturerait la phase d'analyse.

Nous n'avons cependant pas pu aller forcément jusqu'à fin 2008 sur certains aspects de certaines sous-questions de recherche, en particulier la dernière. Mais un complément de dépouillement et codage sur ce point n'a qu'un coût marginal d'un mois et une utilité marginale incertaine en thèse. Notre travail peut être complété après soutenance, si besoin est.

3.3. « Rendre à César ce qui est à César », l'intérêt d'une approche essentiellement historique[105]

Nous présentons ci-dessous ce qui est l'évolution de notre pensée en matière de méthodologie en thèse. Nous montrons les insuffisances des approches traditionnelles en méthodologie qualitative au minimum dans le cas particulier de notre travail, à savoir l'ethnographie virtuelle et la « grounded theory », même si nous observons des recoupements

[105] « Rendre à César ce qui est à César » se justifie. Les premiers chercheurs en sciences sociales qui ont fait des travaux qualitatifs étaient dans l'état actuel de nos connaissances des historiens (la critique de la véracité de la donation de Constantin par Valla ou *La sorcière* de Michelet sont des travaux qualitatifs). Nous sommes particulièrement étonnés de constater que dans les manuels d'analyse qualitative ou autres cours et articles dédiés à la question (sur papier ou sur la Toile) il n'y ait jamais la moindre référence au modèle d'analyse historique, alors qu'on sait que les historiens ont toujours travaillé des données qualitatives, et que nombre d'opérations intellectuelles existaient déjà avant en histoire (triangulation des sources existait déjà sous forme de « croisement des sources »). Seul MILES et HUBERMAN (1994) à notre connaissance leur rend justice en montrant, par une référence à *Rashomon*, l'importance du principe de l'enquête (mais sans parler de la problématique de l'établissement des faits et des archives).

ou des formes d'inspirations. Elles ne sont pas sur le même plan, mais leurs insuffisances convergent vers le même « point de fuite », la nécessité progressive d'une approche de type historique[106], qui comporte des avantages décisifs pour traiter des matériaux d'archives et des traces. Ce n'est pas pour autant que cette méthode a été fréquemment employée et que les historiens eux-mêmes se sont intéressés au sujet. Enfin, nous expliquerons pourquoi nous avons choisi la « micro-histoire » comme spécification méthodologique.

Nous présentons à l'avance, de façon résumée, en quoi cette approche historique permet d'obtenir des résultats. Ils permettent par traitement d'archives de qualités diverses de comprendre les dynamiques comparées des différentes parties de la communauté, de faire l'archéologie du processus de formation du bien collectif, de comprendre par une source indirecte la spécificité des routines, de comprendre l'évolution dynamique des relations entre contributeurs, et enfin à partir de traces la question des usages efficaces du wiki.

3.3.1. Les limites des autres approches, pourtant envisagées, et leur utilité à la marge

3.3.1.1. Apports à la marge et limites de « l'ethnographie virtuelle »

Nous n'avons pas ici l'intention d'opposer les méthodes de façon caricaturale, mais de nuancer nos propos sur le passage de l'ethnographie à l'histoire, même si nous maintenons nos conclusions. Si notre approche est historique, nous ne prétendons cependant pas à une originalité absolue, les différentes sciences sociales se recoupant ici à la marge[107].

Une simple enquête rapide sur « l'Internet des chercheurs », par exemple sur le « Forum of Qualitative Social Research » montre que le nombre de références au mot « virtual ethnography » est conséquent. C'est là une évolution de plus en plus dominante, probablement depuis HINE (2000). En particulier, cette approche se répand pour analyser les communautés de pratiques traitant de nouvelles technologies (par exemple MADANMOHAN et NAVELKAR, 2004), ou les communautés open-source, par exemple

[106] Bien entendu, nous n'abordons ici que les aspects qualitatifs, une telle question ne se pose pas pour les aspects quantitatifs, principalement des séries chronologiques qui se rapportent plus aisément à l'histoire qu'à l'ethnographie ou la « grounded theory ».

[107] La frontière entre histoire et ethnographie ne va pas de soi dans ce terrain si particulier, même si on le sait la philosophie des deux approches est opposée (terrain contre archives). La nature du matériau choisi pourrait inciter à des développements méthodologiques pour définir la frontière exacte entre les différentes approches, mais de tels raffinements n'entrent pas dans notre sujet. C'est ce qu'a révélé par exemple le séminaire AEGIS de mars 2007 où intervenait Anca Metiu sur l'ethnographie virtuelle (associée à la « grounded theory ») : nous lui avions alors dit que ce qui se faisait ressemblait par certains aspects à de l'histoire, elle a acquiescé (pour le compte-rendu, voir DUMEZ, 2008).

HUYSMANS et YUWEIN (2005) ou DEMAZIERES et al. (2007). Il ne serait guère utile de citer d'autres travaux. L'usage d'une telle approche pour les communautés de conception de logiciel n'est plus à démontrer, et nous n'insistons pas là-dessus. L'ethnographie trouve sa raison d'être entre autres dans l'étude d'un milieu où les *relations* sont essentielles (BEAUD et WEBER, 2003). La question des relations, i.e. des réseaux peut être d'ailleurs abordée autrement que de façon ethnographique.

Nous pensons qu'une telle approche, *dans une certaine mesure pertinente* (connaissance de la communauté, contextualisation, qui ne lui sont en fait pas spécifiques), comporte des limites significatives. Certes, le travail de nos données est exigeant, et cela rejoint le travail de l'ethnologue : il s'agit d'un matériau « local » technique ardu tout particulièrement pour le non informaticien[108]. Cependant, il est impossible de suivre en détail l'activité intense quotidienne dans la communauté, comme un ethnologue : le coût quotidien en nombre de « clics » serait trop conséquent. L'approche historique qui suppose de sélectionner les sources est plus efficiente. Elle a aussi davantage vocation à permettre de tester des hypothèses, ce qui n'est pas le cas de l'ethnographie, ce qui fait consensus chez les ethnologues (ULICHNY, 1991). A la marge, si l'ethnographie virtuelle est cependant possible dans une partie de la communauté[109], ce n'est pas le cas dans la plupart des situations (modifications du wiki, surveillance des listes de discussions à haut débit). Sur les listes de discussions à haut débit, nous avons adopté la démarche suivante pour les listes de discussion : nous avons choisi parfois une *stratégie de recherche intermédiaire*, dépouiller en début de mois les données *récentes* de chaque liste de discussions.

Par ailleurs, une méthode de type intrusif comme l'ethnographie participante aurait à coup sûr perturbé la communauté lors des phases de « régime faible » où nos interventions auraient certainement provoqué des discussions supplémentaires, aurait donné à ses membres le sentiment d'être observés, et nous ne voulions pas prendre ce risque[110]. Nous pouvons sur ce point justifier une approche non-intrusive, l'éthique de la recherche autorise l'analyse d'archivages clairement publics, dont la vocation est d'inciter à la lecture et à la participation.

---

[108] Le travail des listes de discussions et du Codex amène à prendre en compte les spécificités du langage et des formes d'énonciation des informaticiens, et celui de ce type de communautés en particulier.

[109] La surveillance de la liste de discussions « wp-doc » peut par exemple se faire sans problème lors des périodes d'activité réduite, même si notre travail a consisté en dépouillement d'archives correspondant à des événements éloignés, et non en une surveillance au jour le jour. La raison de cela est simple : la prise en compte des matériaux doit se faire après réflexion sur la façon de l'aborder, ce qui occasionne retards et surtout dépouillements et lectures multiples, *ce qui n'est pas une approche ethnographique*.

[110] Nous nous différencions donc des méthodologies de « recherche-intervention » du CRG, parfaitement légitime dans d'autres contextes, et peut-être pour les wikis d'entreprise. Nous ne nous avancerons pas là-dessus, et nous *n'excluons nullement* d'y recourir par la suite.

Etudier une communauté ainsi, c'est d'une certaine façon lui rendre hommage, ce qui n'a rien d'hostile.

Enfin, l'intérêt de l'histoire pour examiner les comportements réels et les processus est compréhensible, même si une telle orientation est moins aisée ici que les approches sociologiques. En effet, de même que la fouille des poubelles est plus instructive que les interviews pour comprendre les habitudes de consommation (DOUGLAS, 1999), de même l'examen des entrepôts des historiques des wikis sous format XML (« dumps ») ou des archives en dit probablement plus long que les affirmations orales ou les enquêtes par questionnaires électroniques pour comprendre par exemple la nature des contributions et leur quantité, ainsi que les aspects dynamiques dans la communauté dont des interviewés ne se souviennent pas forcément.

3.3.1.2. Apports à la marge et limites de la « grounded theory »

Nous étions partis, en suivant l'une des approches prônées au CRG (DUMEZ, 2006) de l'idée de mettre en œuvre les principes de la « théorie ancrée » (« grounded theory »), une approche dominante en méthodologie qualitative. Elle a servi d'inspiration. Mais les limites d'une telle approche sont vite apparues. Si nous avons conservé l'idée du codage et d'un travail inductif, la nature du résultat ressemble bien davantage à un travail de type historique qu'aux travaux de codage à partir d'entretiens (POINT et VOYNNET-FOURBOUL, 2006).

Par ailleurs, notre approche, qui comporte de nombreuses données et sources, suppose de considérer en priorité les faits et l'établissement des faits, à partir de traces ou de données excessives, ce qui n'est pas la vocation de la « grounded theory ». Dans notre perspective, si codage il y a, sa finesse importe moins que le fait d'extraire tous les fragments pertinents pour arriver à tester des hypothèses concurrentes. Nous n'étions finalement pas intéressé par la théorisation (« codage à visée théorique » mentionné dans le précédent article) ; et « travail inductif fin » ne signifie pas nécessairement « grounded theory ». Elle ne permet pas de prendre en compte de façon décisive la dimension chronologique et dynamique des phénomènes, de tester des hypothèses[111]. Pour de plus amples développements sur ces questions et les limites multiples de cette approche, nous renvoyons à la synthèse de SUDDABY (2006), qui a confirmé nos vues issues de l'expérience.

[111] Cette question d'ordre méthodologique en sciences sociales mériterait évidemment de plus amples développements, et a des implications cruciales sur la façon de composer des analyses de type qualitatif en économie et gestion.

3.3.1.3. Quelques précisions sur l'absence d'interviews ou de questionnaires électroniques dans le cadre de cette thèse

Il est fait souvent usage de ces outils pour l'analyse des phénomènes de l'économie du numérique[112]. Plusieurs raisons permettent de nous en tenir à l'analyse d'archives. Certes, *un historien n'a pas à s'interdire d'effectuer des entretiens*[113]. Nous avons d'abord manqué de temps, et nous avions l'accord de notre directeur de thèse de nous en tenir là. Nous tenions aussi, dans le cadre d'une thèse, à une cohérence d'ensemble satisfaisante et à tester la solidité d'une méthodologie, ce qui permet d'en cerner les limites. Comprendre les avantages et les limites d'un outil est un apport utile d'une thèse, surtout si cet outil a été auparavant négligé. Le risque de perte d'archives rendait aussi leur étude *prioritaire*. Les questionnaires électroniques sont par comparaison plus aisés et nous voulions en priorité nous différencier en nous attaquant à ce qui était ardu, les traces, les comportements effectifs, les données primaires. Il est enfin utile de rappeler que nous voulions effectuer une enquête non-intrusive, le phénomène de documentation était fragile. Cependant, il est toujours possible de mettre en œuvre ce type d'approche *par la suite* au besoin avant de délaisser définitivement notre cas.

3.3.2. Au-delà de l'analyse longitudinale, la nature des données appelle une analyse de type historique

Les méthodes d'analyse qualitatives et quantitatives sont nombreuses : nous avons choisi la méthode historique, qui s'adapte aux données rassemblées, des traces et au mode d'entreposage des données (« archives »). Les archives des contributions, des textes, et des interactions sont particulièrement ardues à étudier, et le croisement des sources de chaque élément de l'étude de cas n'est pas un détail. Cependant, si notre formation personnelle d'historien nous rend sensible à ces questions[114], elle est également due aux contraintes spécifiques de notre sujet et du matériau rassemblé, qu'il s'agisse « d'archives » de listes de discussions ou « d'historiques » de wiki : nous sommes confronté à des traces. Nous évoquerons après présentation des archives leurs biais, ce qui est une autre façon de

[112] On pourrait par exemple renvoyer à la thèse de Kevin Mellet sur le recrutement par Internet.
[113] L'ouvrage d'histoire des idées *L'empire du sens* de François Dosse est un bon exemple d'application de cette méthode. Mais elle reste marginale.
[114] On peut douter aussi que ce soit à cause de notre formation d'historien, ce à quoi pensera évidemment notre lecteur : nous avions effectué un DEA d'économie pour *partir* de cette discipline, et y revenir ne correspond pas à la trajectoire que nous avions choisie. Il s'agit d'un curieux hasard dû à la recherche « chemin faisant ».

poursuivre la justification de la nécessité de l'approche historique. Enfin, l'originalité de l'approche historique se retrouve aussi dans la spécificité des données statistiques et de leur analyse : les données historiques appartiennent à un monde « parétien » (avec des distributions en loi de puissance, dont nous verrons de nombreux exemples) et non à un monde « gaussien » caractérisé par des courbes « en cloche ». Sur ce point, nous renvoyons à GUERREAU (2004), qui intervient en la matière à l'Ecole des chartes.

Quel est l'avantage significatif de l'approche historique ? L'une des difficultés majeures de l'analyse d'un projet dans une optique longitudinale est de prendre en compte le temps : en effet, notre documentation se modifie au cours du temps, et passe d'un état embryonnaire à des formes de stabilisations plus ou moins satisfaisantes ; de même notre communauté, comme d'autres, passe par un *cycle de vie*, à caractériser avant éventuelle généralisation. Une approche historique est un bon moyen de traiter ce type de problématique. Ensuite, elle permet de prendre en compte l'établissement des faits de façon sérieuse, le contexte (faits, énonciation) et sa complexité. L'articulation entre le quantitatif et le qualitatif est aussi une des caractéristiques de l'histoire. C'est exactement ce à quoi nous sommes confrontés. Nous attendons aussi de cette approche qu'elle nous permette de tester des hypothèses (un des corrélats de la question de l'établissement des faits), une question qui nous tarabustait en matière d'approches qualitatives, les précédentes méthodes ne nous ayant pas donné satisfaction sur ce point.

Cette approche a cependant un coût, reconnu par les auteurs en méthodes qualitatives : un travail exigeant sur les matériaux. L'approche historique, au-delà de ses spécificités, comporte une grande part de bricolages et d'improvisation, en fonction de ce que peuvent révéler les sources. Elle suit le modèle de l'enquête, qui n'est pas incompatible avec l'analyse des mécanismes économiques. L'approche historique, l'enquête, est d'autant plus cruciale que nous n'avons pas eu la possibilité, contrairement à l'étude d'autres projets de pouvoir extraire les « dumps » des projets de la Wikimedia foundations, et de disposer de ce fait d'outils d'extraction et de traitement de données spécifiques.

3.3.3. Les approches historiques en sciences de gestion et en management des systèmes d'information sont appréciées, mais des travaux supplémentaires sont bienvenus[115]

On n'associe plus guère aujourd'hui la démarche historique à une visée expansionniste en sciences sociales, par comparaison avec la période du développement de l'Ecole des Annales[116]. Cependant, l'approche historique en sciences de gestion comporte une légitimité solide, mais qui ne suscite peut-être pas assez de vocations. On observe cependant en prenant du recul en théories des organisations un retour *cyclique et ponctuel* du besoin d'analyse historique, comme le remarquait KIESER (1994)[117]. Certains auteurs isolés travaillant la notion de « rationalité limitée » préconisent aussi cette approche (LORENZ, 2001). En ce qui concerne notre axe d'analyse économique, SIMON (1999), un auteur entre autres économiste dont la légitimité en gestion n'est plus à discuter, demandait clairement à ce que la spécificité de la démarche historienne soit davantage reconnue pour aborder des problématisations économiques comme l'économie du savoir, précisément. DUMEZ et JEUNEMAITRE (2005) rappellent l'utilité d'une démarche narrative en économie. En France, Anne Pezet a donné un cours au CEFAG sur cette question[118]. Une telle approche est parfois employée au CRG: nous pouvons mentionner la thèse d'Eric Maton sur un phénomène de longue durée, l'usage des représentations graphiques en management (MATON, 2006), ce qui est rare en gestion. Cependant, ce n'est pas par « mimétisme communautaire » que cette approche a été choisie, même si une convergence est observée.

En revanche, seules quelques recherches éparses en systèmes d'information montrent déjà l'intérêt d'une approche historique, qui peut aller au-delà d'une focalisation sur les aspects économiques. On se rappellera par exemple DE VAUJANY (2006) qui a étudié la notion de diachronie lors d'une thèse originale sur les systèmes d'information de la Curie romaine. BANISTER (2002), qui est certainement l'article méthodologique de référence en la matière, soulignait déjà les raisons de la rareté des approches historiques dans les travaux de recherche en système d'information, malgré l'importance du champ. Il dégage les apports

[115] La question ne se pose pas uniquement en science de gestion. Le récent master sur l'étude interdisciplinaire de l'Internet ne mentionne pas l'histoire parmi les nombreuses disciplines pertinentes qu'il cartographie (http://masterwebscience.org/, consulté le 14/09/09).

[116] La discipline qui imposerait, de façon discutée, sa « rigueur méthodologique » aux autres sciences sociales serait plutôt l'économie.

[117] On peut signaler au lecteur la récente reprise de la revue *Management & Organizational History* qui propose d'aborder et problématiser les thématiques d'organisation et management à partir d'une approche historique. De telles hésitations montrent encore une fois que des travaux réguliers suivant une telle approche ne vont pas de soi.

[118] En ce qui concerne sa légitimité, J-M. Saussois est l'un des rares à mentionner sa légitimité en gestion dans un ouvrage universitaire à grand tirage (SAUSSOIS, 2004).

possibles de cette méthodologie[119], dont la liste n'est pas exhaustive; mais il ne mentionne pas son utilité pour aborder des problématiques économiques, notre préoccupation ici. Sur une question proche de la nôtre en SI, ALEXY (2008) met en évidence le rôle crucial du temps dans les processus de décision[120].

La seule tentative qui s'approcherait de ce genre pour l'analyse du phénomène open-source que nous ayons trouvée sous la forme d'article ou « working paper » est BENUSSI (2006)[121]. Lorsqu'il est question ailleurs d' « histoire » dans ces supports, il s'agit uniquement d'introductions au sujet du travail en question[122]. Cela peut expliquer que, d'après nos fructueuses discussions de mai 2009 avec Jean-Michel Dalle, il a été confirmé que l'approche historique des communautés open-source est l'un des aspects originaux de notre travail.

3.3.4. L'analyse limitée par les historiens de l'avènement du numérique néglige la question des archives

Il était indispensable d'évoquer ce qui se faisait dans cette discipline voisine, qui intéresse aussi la gestion[123], ce qui donne aussi une perspective plus large à notre sujet. Et c'est à l'avantage de cette dernière discipline si elle peut contribuer au passage à l'avancement de la science dans les disciplines voisines.

La question du *traitement* des matériaux spécifiquement numériques, sous forme d'archives ou non, n'a été soulevée que de façon marginale, en particulier chez les historiens (RYGIEL, 2005). En revanche, elle est davantage prise en compte par la recherche en ingénierie logicielle (CROWSTON et HOWISON, 2004, ROBLES et al., 2005) ou en informatique (MASSEY, 2005). Ces chercheurs analysent les changements dans les entrepôts publics de code source (« concurrent version systems »), mais c'est une approche trop éloignée des formes de questionnements spécifiques des sciences sociales. Une préoccupation émergente chez les historiens concerne la question de l'archivage des données de l'administration électronique (DHERENT, 2002). Pour le type d'archives que nous

[119] Analyse poussée du rôle structurant des prises de décision, visualisation de régularités historiques, prise en compte poussée de toutes les sources disponibles, possibilité de tester des hypothèses et des théories, prise en compte des phénomènes dans leur contexte, etc. .

[120] Sur la question plus générale de l'histoire de l'informatique et des SSII, on renverra le lecteur intéressé au numéro spécial n°40 d'*Entreprises et Histoire* de 2005, mais aussi au musée de l'informatique récemment ouvert à La Défense. Sur cette histoire et celle des SI, la recherche en est à ses balbutiements, encore une fois.

[121] Parmi les ouvrages adoptant une perspective longue, on peut aussi renvoyer à (sic) NOISETTE et NOISETTE (2004).

[122] Par exemple, parmi les travaux les plus prestigieux, VON HIPPEL et VON KROGH (2003).

[123] Nous avons sur ce point suivi les recommandations utiles d'Hervé Dumez.

mentionnons ici, les références sont également rares d'après nos recherches (LANGNER, 1997), et la différence de traitement de données est marquée par comparaison avec les approches sociologiques, comme CONEIN et LATAPY (2008) et BOULLIER (1989). On peut cependant observer une *prise de conscience* à la marge par les communautés de leur propre histoire indépendamment des historiens (YAMANADA, 2005, FLIN, 2007). Nous renvoyons sur ce sujet au projet « Community archives and identities: documenting and sustaining community heritage » à l'UCL[124].

Nous préférons préciser à propos de ces insuffisance, pour éviter d'éventuelles confusions, que cette question est différente d'autres problématiques voisines qui intéressent les historiens, à savoir du processus institutionnalisé ou non de numérisation d'archives ou de documents existants, de l'usage de l'Internet par la profession historienne pour écrire l'histoire et traiter des données de façon plus collaborative, ou encore de la question de la conservation, de la politique d'archivage ou de l'organisation des matériaux spécifiquement numériques.

Toutes ces questions ayant trait au numérique, y compris la nôtre, sont pourtant de première importance pour la profession historienne, ou pour tous ceux qui sont contraints d'adopter une démarche historique, par exemple en gestion. L'émergence de la prise en compte de la numérisation des archives ne peut être précisément datée et qualifiée, et mériterait un travail historique en soi. Signalons seulement que le service « Web archives » est né en 1996[125], alors que la Toile naît en 1989. L'histoire des réseaux informatiques a plus de quarante ans, fait l'objet de débats[126], et l'histoire contemporaine ne saurait faire l'impasse sur un phénomène aussi majeur. Il est sans doute révélateur que le numéro spécial d'*Entreprises et Histoire* de 2006 sur le déploiement du numérique néglige la question du traitement des sources et des archives. P. Fridenson est par ailleurs l'un des rares historiens à s'être intéressé ponctuellement à la question du numérique selon une perspective d'histoire immédiate (FRIDENSON, 2006).

[124] Source : http://www.ucl.ac.uk/infostudies/research/icarus/community-archives/ (consulté le 09/06/09)
[125] Source : http://www.archive.org/about/about.php, (consulté le 28/11/2008.)
[126] Nous citerons le travail exploratoire de CAMPBELL-KELLY et GARCIA-SWARTZ (2005), qui montre l'utilité d'une approche historique sur ces questions.

3.3.5. Une tentative de micro-histoire dans un sujet d'histoire économique « immédiate » où sont traitées des problématiques de gestion

Il est nécessaire de *spécifier* notre méthodologie, de même qu'il est bon de spécifier les théories mobilisées.

Parler d'approche historique demande des précisions, car ce champ disciplinaire a une certaine largeur, et plusieurs « genres » historiques pourraient être mobilisés ici de façon plus ou moins pertinente. F.-X. De Vaujany ou E. Maton ont fait, eux, de l'histoire sur des formes de plus longue durée. Notre approche relève plus précisément, au-delà d'une catégorisation peut-être désuète en « histoire économique et sociale », de *deux genres historiques*. D'une part, nous avons effectué comme P. Fridenson de « l'histoire immédiate »[127], au vu de la périodisation choisie, d'autre part, nous avons mobilisé un genre historique original, la « micro-histoire ». C'est sur ce dernier point que nous allons insister. Nommer ce que nous avons fait contribue à donner une légitimité.

Bien qu'il recoupe cet autre genre qu'est l'étude de cas, à notre connaissance, ce genre historique n'a pas été mobilisé en gestion, à l'exception de BORZEIX et al. (2006) ; une expérience de recherche supplémentaire en la matière comporte donc une utilité marginale forte. Son retour d'expérience proposée dans le *Libellio d'AEGIS* est pertinent pour éclairer l'emploi de ce genre scientifique :

> "La micro-historia se refuse à opposer global et local ou même à les hiérarchiser. L'expérience la plus élémentaire, individuelle ou d'un groupe, est éclairante parce que plus complexe ; elle permet d'analyser des phénomènes dynamiques, d'émergence, de circulation, de négociation, d'appropriation, des processus, et non des états." (BORZEIX, 2008)

C'est ce que nous allons mener à bien par exemple en étudiant au chapitre 6 les processus microanalytiques que sont les routines dans la communauté, et au chapitre 7 en tentant de comprendre la structuration dynamique des échanges permettant entre autres la création et discussion de ces routines. C'est là un état d'esprit de recherche que nous allons assumer, et pas seulement en matière qualitative, les approches quantitatives permettant des apports pour comprendre justement une communauté au niveau plus global. Par ailleurs, nos matériaux fins, disponibles, se prêtent tout à fait à ce type d'analyse, alors que…

> « rares sont les cas où la documentation à caractère dialogique (paroles proférées, actes de langage) est disponible, ce qui permet de rapprocher la micro-historia de l'expérience ethnographique et de la microsociologie des interactions verbales. » (*Ibidem*)

[127] Le groupe de recherche en histoire immédiate (source : http://w3.grhi.univ-tlse2.fr/, consulté le 12 mai 2009) s'est par exemple spécialisé en la matière.

Les archives que nous allons présenter dans la prochaine partie sont des matériaux à caractère dialogique marqué.

Ensuite, ce genre historique se prête bien à une articulation avec d'autres sciences sociales (comme les sciences de gestion) examinant le niveau fin des décisions, des comportements individuels, les interactions, en particulier dans des groupes oubliés[128] ou restreints, comme un village[129].

> « C'est la complexité de l'analyse, selon les micro-historiens, qui requiert le rétrécissement du champ de l'observation. Mais le principe de base demeure : c'est à partir de comportements individuels que l'on peut construire les modalités de l'agrégation sociale. Il s'agit donc pour eux (comme pour nous) de dé-naturaliser, de dé-banaliser ces mécanismes d'agrégation dont les entités sociales sont le fruit, en insistant sur les modalités ou mécanismes relationnels, organisationnels, communicationnels, pragmatiques qui les rendent possibles. » (*Ibidem*).

Ici, au-delà de la ressemblance entre une communauté et d'autres groupes marginaux étudiés par la micro-histoire, au-delà de l'analyse organisationnelle que nous allons mener et qui est attendue en Système d'information, l'articulation avec l'économie qui part du principe de l'individualisme méthodologique rend pertinente une approche micro-historique.

Enfin, nous précisons que contrairement à une opinion répandue, micro-histoire et quantification ne sont nullement incompatibles, malgré des débats dans la profession historienne. Des micro-historiens font usage par exemple de l'économétrie et de l'analyse de réseau (LEMERCIER et ZALC, 2008), comme nous le proposons dans notre chapitre 7. Plus généralement, la micro-histoire devient une perspective pionnière du renouvellement de l'histoire économique et sociale en France, au-delà des perspectives « labroussiennes » que la micro-histoire a critiqués[130]. Sa pertinence en science de gestion peut également en découler.

3.4. Présentation, critique des sources et formes de triangulations

L'étude de la collaboration massivement distribuée fait que les données sont aussi distribuées (une communauté comporte plusieurs sous-projets de tailles différentes). Ce point

[128] On peut étudier ainsi le groupe peul baalinkobe, « ceux des moutons », en Afrique BOURGEOT André (1999), *Horizons nomades en Afrique sahélienne: sociétés, développement et démocratie*, KARTHALA Editions.

[129] DEMADE Julien,"Transiger, transmettre, lier :micro-histoire économique d'un village franconien au XVe siècle", http://hal-paris1.archives-ouvertes.fr/halshs-00008940/en/ .

[130] Nous renvoyons sur ces questions méthodologiques à LEMERCIER et ZALC (2008) ou à la journée d'étude 10 juin 2009 de l'équipe « Économie et société : territoires, activités et relations sociales » du laboratoire IRHIS (source : http://irhis.recherche.univ-lille3.fr/JE-10juin2009-Galvez.html, consulté le 12/06/09).

fait que l'exhaustivité des sources et la réduction satisfaisante des risques d'incomplétude des faits doivent être traitées de façon sérieuse.

3.4.1. Nos choix des sources parmi celles qui étaient disponibles

Dans le cadre d'une thèse de doctorat, et sans aide et compétences en traitement automatisé de textes et extraction automatique de données, il est impossible de lire l'ensemble des listes de discussions. Il n'est pas dit non plus que l'utilité marginale d'une telle démarche aurait été systématiquement positive.

L'importance des sources oblige le chercheur à des *arbitrages*, d'autres sources sont trop conséquentes pour être traitées « à la main ». Par exemple, le forum des utilisateurs comporte plus de 100.000 fils de discussions et l'extraction de toutes les données pertinentes est impossible sans compétences poussées en informatique. Nous n'avons extrait des informations qu'à la marge à partir des mots-clefs proposés sur le forum lui-même. Une consultation par l'intermédiaire de google montre que le mot « codex » y apparaît des milliers de fois. Un examen exploratoire par ce moyen montre que des milliers de fils de discussions du forum comportent des liens vers le wiki pour aider l'utilisateur (ce qui a été confirmé par l'analyse qualitative de la liste de discussions de régulation du forum). De plus, les résultats des requêtes par moteurs de recherche à l'intérieur de cette sous-partie de la communauté sont hautement sujets à caution[131], rien n'est dit de la répartition temporelle des citations, de leur nature, etc. . Il était au-delà de nos forces, sans support informatique adéquat, de dégager et analyser rigoureusement tout cela… Le forum comporte enfin une limite majeure : au-delà d'un an, la datation des fils est floue (« one year ago », « two years ago », etc.) ce qui interdit toute analyse dynamique et rendrait les conclusions a-temporelles pour l'essentiel. C'est là un des biais occasionnés hélas par l'outil de forum « bbpress », qui limite la portée d'une analyse dynamique rigoureuse.

Nous avons par ailleurs négligé faute de temps d'étudier le wiki sous moteur « Trac » spécialement dédié aux développeurs et au suivi des défauts informatiques « bugs ». Nous allons donc négliger la dimension « management de projet » dans sa partie programmation et « production », alors qu'il s'agit d'un wiki, certes particulier. Nous avons enfin négligé les phénomènes de traductions dans d'autres wikis (français, espagnols). Pourtant, il y aurait certainement des problèmes de coordination en matière de documentation à creuser : cette

[131] On pensera en particulier à l'exagération des chiffres par Google.

documentation se fait sur le wiki, comporte des contributeurs d'origines hétérogènes, et doit suivre la documentation en anglais et ses modifications nombreuses.

3.4.2. Présentation des sources sélectionnées et contextualisation

Dans la communauté WordPress, chaque sous-projet est coordonné par une liste de discussions (« mailing-list »), et celles-ci sont les principales archives disponibles, mais non les seules. Certaines listes sont de grande taille, d'autres sont de petits projets annexes (« wiki-tech ») ou avortés. La datation du début des fils de discussions montre clairement que les projets associés à chaque liste se sont multipliés au fur et à mesure du succès du logiciel. Il était impossible de tout prendre en compte avec la même finesse.

3.4.2.1. Le site de référence « Web archives » est indispensable mais insuffisant pour une analyse fine, que ce soit en micro-histoire ou en gestion.

Voici un entrepôt de captures d'écran de la communauté collectées par http://web.archive.org/ :

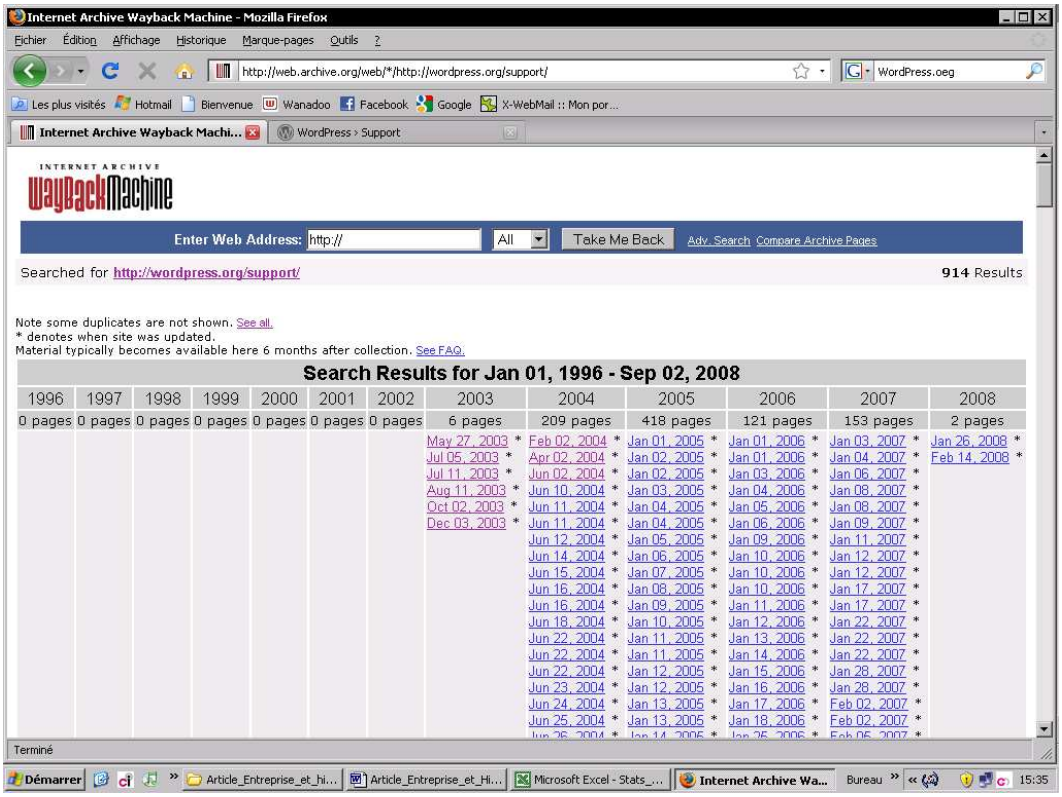


Figure 4 : archives du forum de la communauté capturé par Internet Archives (http://web.archive.org/web//http://wordpress.org/support/ , consulté le 02/09/08)*

L'intérêt de ces archives est de permettre de remonter dans le temps, ici, jusqu'aux débuts de la communauté, dans la mesure où les archivistes repèrent les parties du site essentielles. Chaque date ci-dessus correspond à une capture de la page du site à la date donnée ; elle se trouve derrière le lien de la date. Nous avons ici présenté ce dont nous disposons pour le forum de la communauté, ce qui est fort utile pour établir des séries chronologiques du nombre de requêtes des utilisateurs, un indicateur pertinent du succès de la communauté. Nous pouvons aussi avoir accès à d'anciens systèmes d'information, comme l'ancien wiki qui a été ôté des serveurs de la communauté, et que l'on retrouve ci-dessous :

Figure 5: capture d'écran de l'ancien wiki de la communauté. (source : http://web.archive.org/web/20050224095754/wiki.wordpress.org/ ; date : 24 février 2005)

Ces archives sont indispensables pour permettre de dater certains événements. On ne saurait cependant trop insister sur *l'insuffisance* de ce type d'outil de collecte systématique et

« officielle » pour aborder des problématiques fines. En effet, cette pratique d'archivage ne capture que la « surface » d'activité des sites, et non pas toutes les pages de tous les sites archivés. Tout collecter demanderait en effet un travail « astronomique », au sens mathématique du terme. Archiver oblige l'institution à sélectionner, ici autant qu'ailleurs. Les archives sont par ailleurs collectées de façon plus intensive au fur et à mesure que le temps passe et que la communauté devient connue, ce qui fait que les débuts sont plus lacunaires. Cette insuffisance montre donc la portée des analyses que nous proposons ici à partir *d'autres* archives internes à la communauté. Nous y venons.

3.4.2.2. L'entrepôt principal des listes de discussions

Voici à quoi ressemble le répertoire des archives de la communauté :

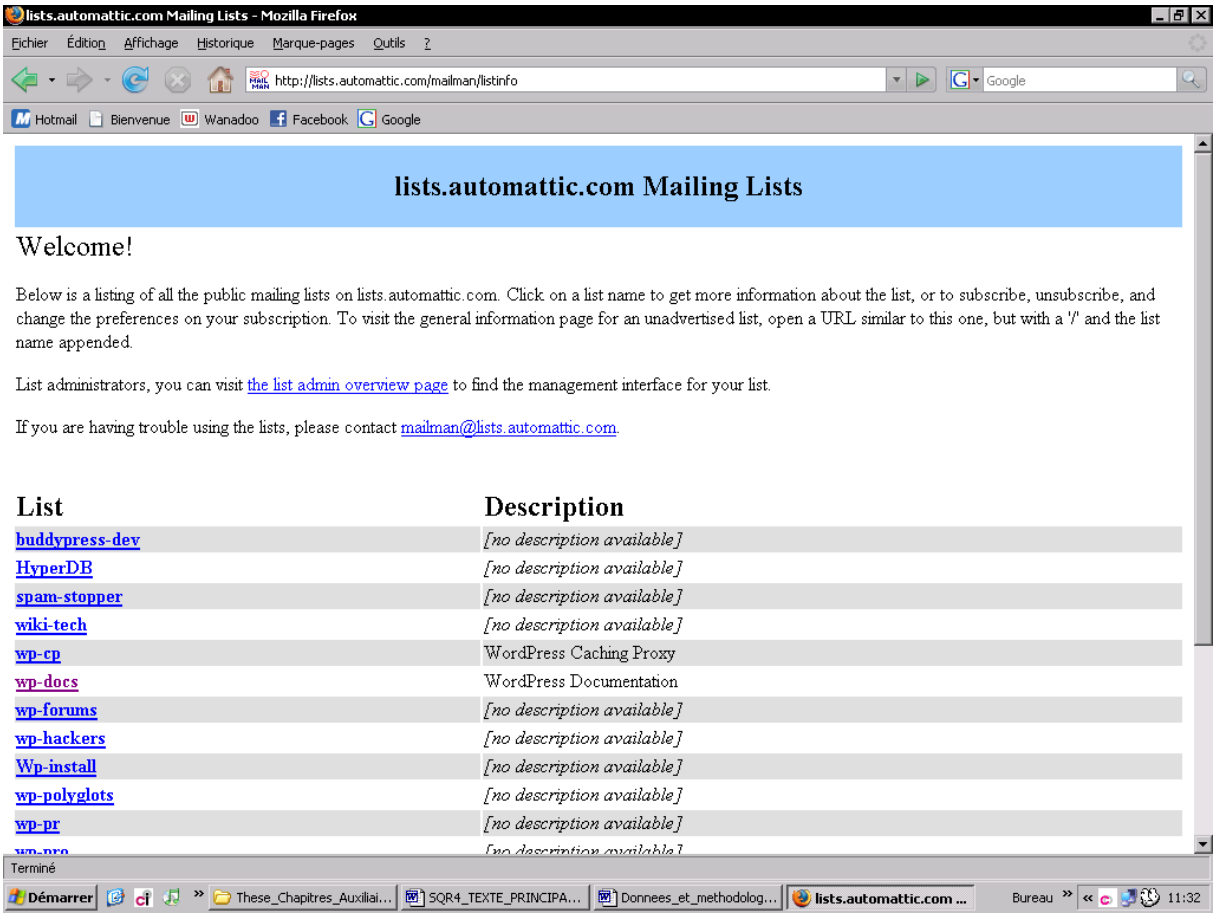


Figure 6 : capture d'écran du système d'archivage officiel de la communauté WordPress (source : http://lists.automattic.com/mailman/listinfo)

Chaque archive dont le modèle sera présenté plus loin se trouve derrière les liens en colonne à gauche. Les listes ont des fonctions diverses et ne sont pas classées : on pourrait les diviser en deux catégories en première approximation, celles coordonnant des projets, celles coordonnant des activités. Toutes ces archives ne sont pas prioritaires, en particulier les différents projets de programmation qui n'ont pas de rapport avec la question des wikis[132], par exemple « HyperDB » qui est un projet de constitution de base de données. A contrario, d'autres archives pertinentes n'y apparaissent pas (plus ?), un point développé plus loin, ce qui fait que les présentations de ce genre sont trompeuses.

3.4.2.3. Comment analyser de façon satisfaisante un wiki ?

Nous allons aborder la question encore une fois à partir d'une visualisation, ici celle de la page d'un contributeur et administrateur reconnu.

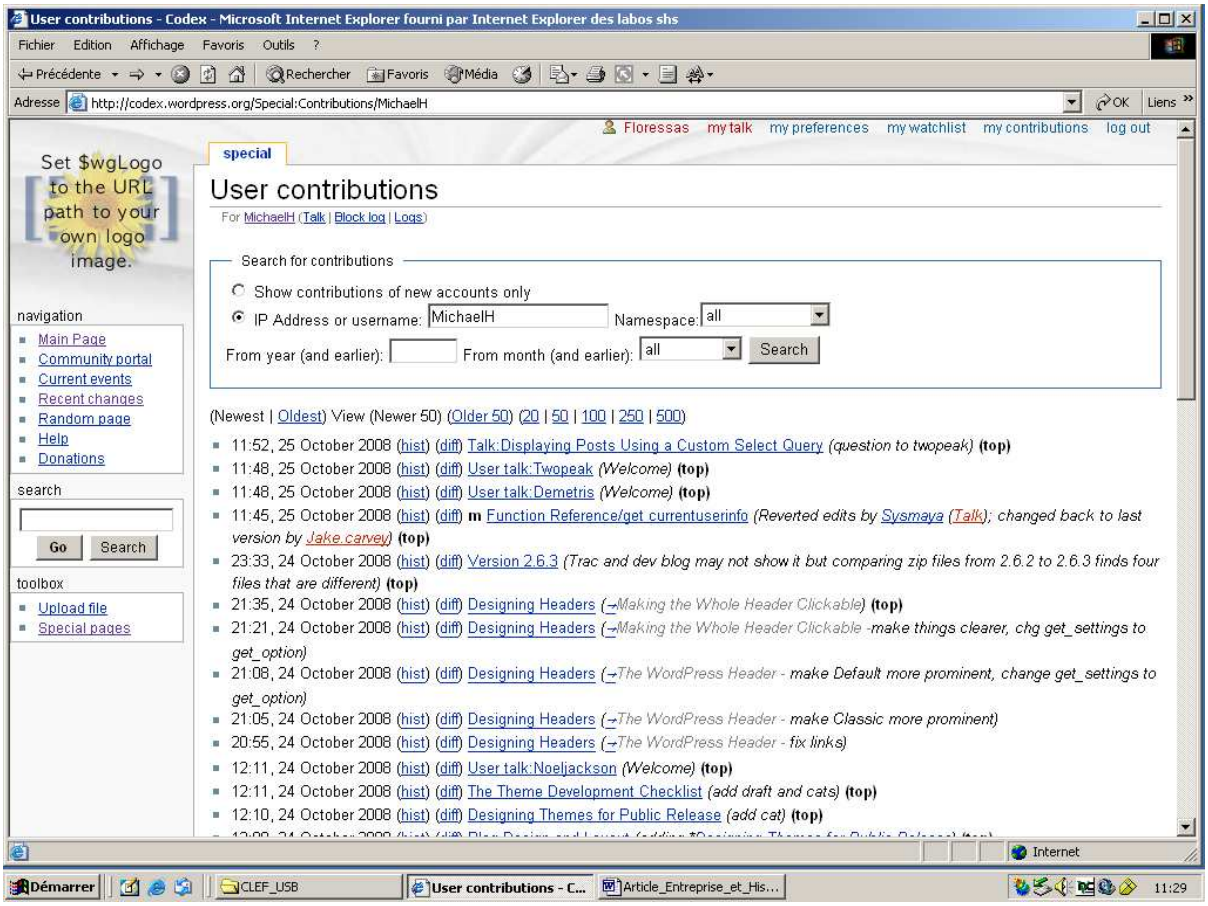


Figure 7: historique des contributions de l'administrateur Michael Hancock (le 25/10/08)

[132] Pour savoir cela au-delà de la fonction de la liste négligée, il suffit d'examiner les fichiers « .txt » récapitulant l'activité de celle-ci et de tenter d'en extraire des mots-clefs liés au wiki avec la commande « ctrl + F ».

Chacune des modifications effectuées ici ou là dans le wiki par ce contributeur notable correspond à une ligne, i.e. un titre souligné (en bleu sur la Toile). A la gauche de chaque ligne la datation exacte, des liens « diff » permettent de dégager à chaque fois par différenciation la nature exacte de la contribution. A droite, en italique, apparaissent souvent des commentaires allusifs sur la nature de la contribution en question. On peut à partir des liens « diff » faire défiler tous les ajouts et trouver ceux qui sont les plus pertinents. Ce type de travail demande du temps. Pour chercher de façon rationnelle, le dépouillement des onglets de discussion des articles (« talk ») et des pages utilisateurs (« user :talk ») sont certainement *prioritaires*. C'est la démarche que nous avons adoptée.

Les historiques comportent les spécificités du moteur MediaWiki, en particulier les menus déroulants permettant de sélectionner et compter visuellement le type d'éditions[133]. Cependant, l'étude d'un wiki est plus ardue que l'étude d'une liste de discussion, qui est chronologique et *centralisée*. L'étude des profondeurs d'un wiki comporte la caractéristique inverse d'une fouille archéologique sur un terrain uniforme, où, en tout point, la distance par rapport au sol permet la datation. Les « couches » d'un wiki sont au contraire de taille non uniforme, le contenu est d'importance inégale, un travail critique est nécessaire car les « éditions mineures » ne sont pas forcément secondaires[134]. Il faut regarder en détail sur chaque « carottage » le détail des dates, et les rapporter au reste des événements ailleurs sur la communauté, tout en conservant à l'esprit que chaque page peut avoir sa vie et son histoire propre, sa durée propre. Une telle temporalité « en accordéon » est particulièrement exigeante pour l'observateur. Le dépouillement des historiques des contributions individuelles comporte la même caractéristique ; de plus, à chaque édition du contenu correspond au moins deux « clic » de souris, ce qui pose des problèmes de traitement des données lorsqu'il est question d'étudier la « carrière » d'un administrateur dont le nombre d'intervention dépasse la centaine d'éditions. Une étude méticuleuse d'un terrain aussi aride est particulièrement peu ergonomique.

[133] Nous signalons que nous disposons d'une liste des éventuelles éditions détruites (onglet « deletion log ») ce qui permet de se faire une idée des biais éventuels sur ce point.
[134] Ainsi, un simple examen de l'historique des contributions de Matt Mullenweg, le « chef du projet », montre que la plupart des éditions mineures ne le sont pas du tout.

3.4.2.4. Les listes de discussions des programmateurs

Le système d'archivage des listes de discussions est standardisé. Nous le présentons pour l'une d'elles seulement. Une capture d'écran, prise en août 2008 est utile pour comprendre son fonctionnement :

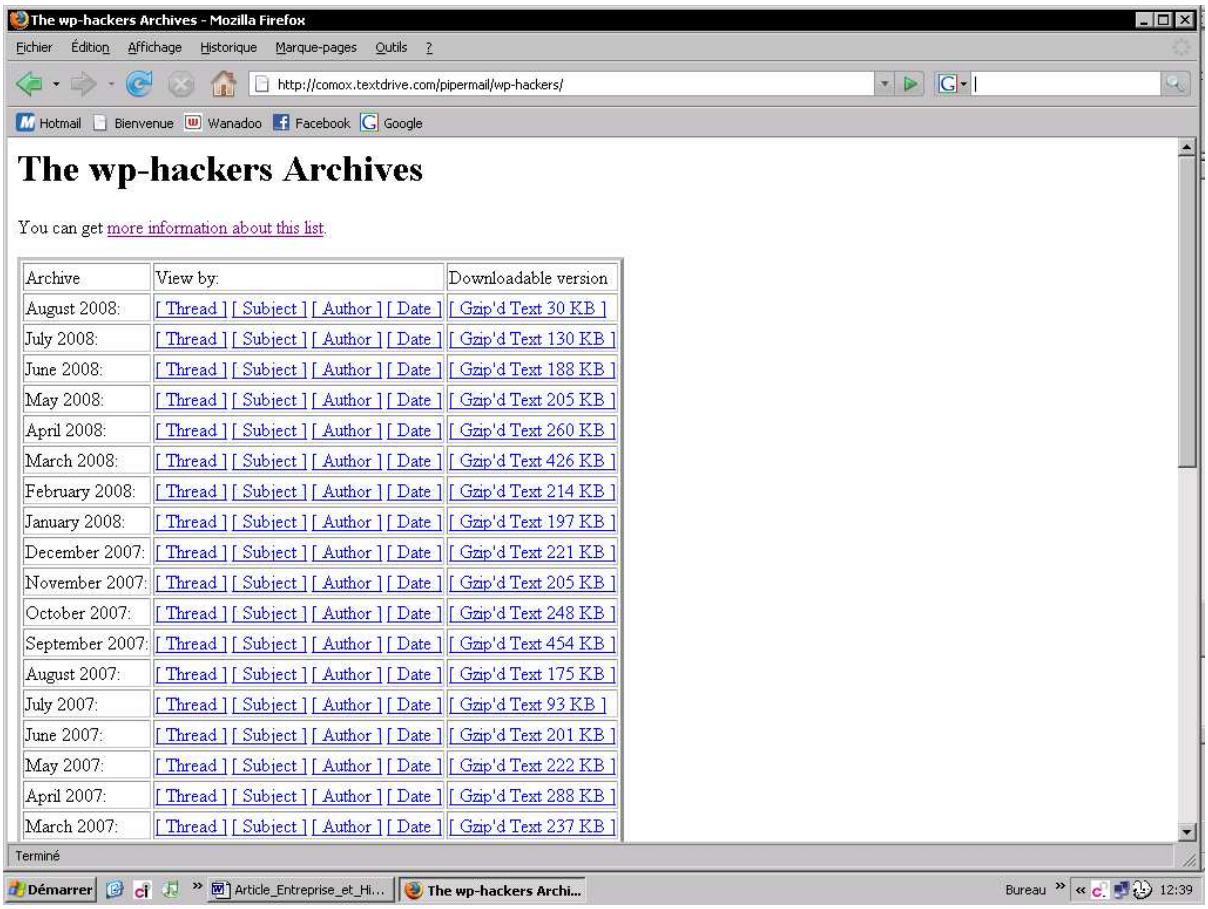


Figure 8 : système d'archivage des listes de discussion de « wp-hackers » (le 29/08/08)

La capture d'écran ci-dessus est un des entrepôts d'archives, celui de l'activité de coordination de la programmation. Les courriels de chaque liste de discussion se trouvent dans ces dépôts, composés de fils de discussions sous format HTML. Chaque dépôt peut être structuré par fil de discussion (l'url « thread » sur la colonne de gauche), sujet, auteur, date. Ces archives proposent en effet divers outils permettant de les classer (par date, auteur, fil, thème). Pour le chercheur, ce format est le plus commode pour étudier les textes pertinents.

Un fichier sous format « .txt » masqué derrière « Gzip'd Text » (« texte sous format Zip ») à droite récapitule l'ensemble des courriels du mois, avec une indication en kilobits de la quantité d'information contenue. Le format « txt. » est utile pour rechercher des courriels par mots-clefs ou pour effectuer d'éventuelles sauvegardes par téléchargement. Ce format permet au chercheur de conserver les discussions dans l'ordinateur et de ne plus être dépendant des données archivées sur le web en cas de panne du serveur, provisoire ou définitive...Tous les dépôts d'archives comportent cette structure d'archivage automatique et la mise en ligne se fait automatiquement sous 24 heures d'après nos observations.

La capture d'écran suivante présente l'activité mensuelle d'une liste de discussion, présentée « par fil » :

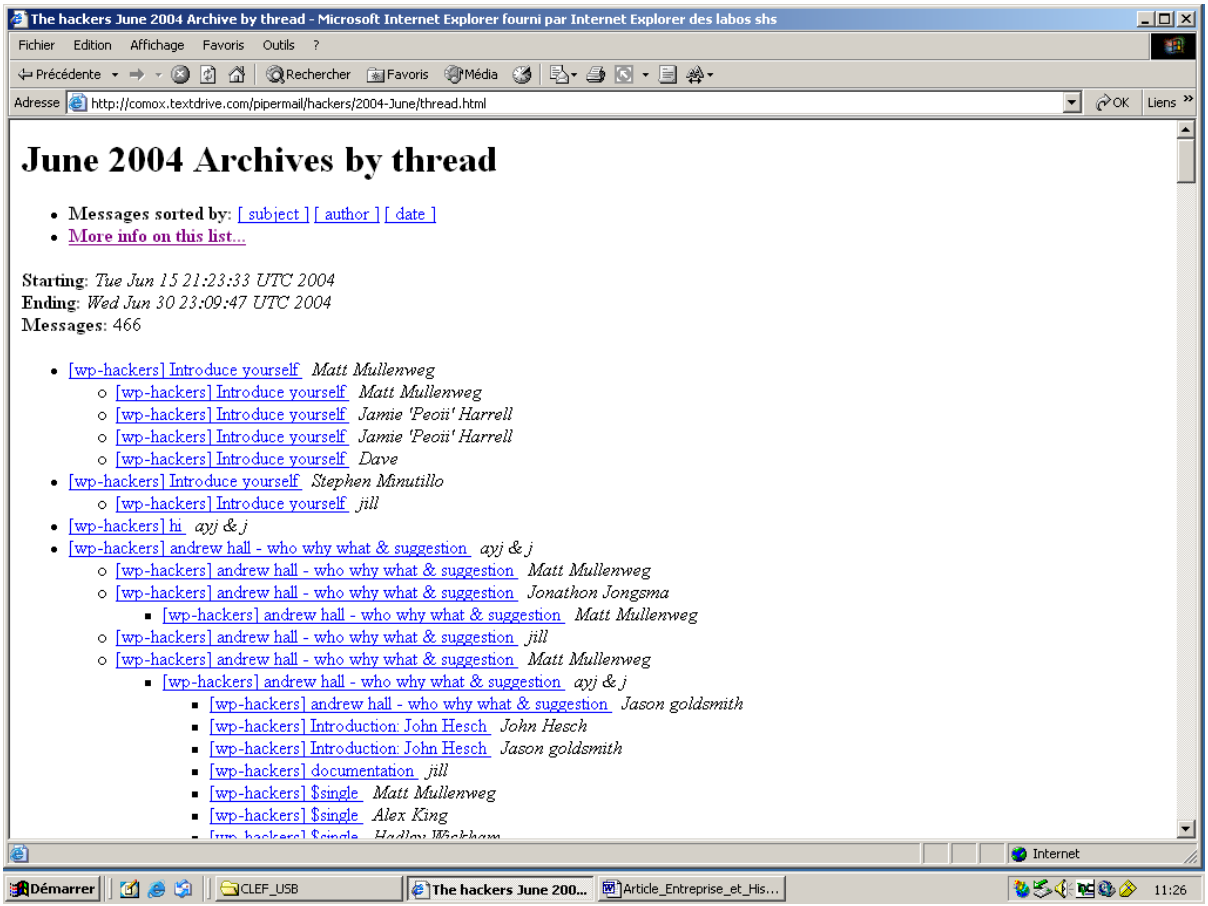


Figure 9 : archives de « hackers », mois de juin 2004, présentées par fils de discussion

Chaque ligne représente un lien vers un courriel. Les courriels se trouvant du côté gauche sont ceux qui ont initié un fil de discussion, ceux qui se décalent vers la droite représentent des réponses successives aux premiers courriels. Chaque courriel est associé à une adresse « url »

permettant de le localiser précisément dans une archive. Ainsi, « http://comox.textdrive.com/pipermail/wp-docs/2005-March/000309.html » désigne, dans l'entrepôt de la liste « wp-docs » sur serveur Apache, le courriel n° 309, parmi ceux du mois de mars 2005. Les courriels sont numérotés dans l'ordre de leur émission[135]. Sur cette capture d'écran, on remarque des fils de longueurs différentes, avec différents niveaux de discussions « en escalier », mais aussi l'intervention fréquente de Matt Mullenweg, à l'origine du logiciel et administrateurs du Codex.

Les listes de discussions des informaticiens sont multiples, et ont différentes fonctions, que nous explicitons ici brièvement pour celles que nous avons dépouillées partiellement d'un point de vue qualitatif (voir infra) et intégralement d'un point de vue quantitatif. C'est *indispensable*.

Celles intitulées « hackers » puis « wp-hackers » servent à la coordination de l'activité de programmation (modification du code ou développement de « plugins ») et à la détermination des orientations générales de conception du logiciel et de la communauté[136]. Elle est donc de première importance par comparaison avec les autres listes mobilisées.

La liste des testeurs (« testers ») sert à coordonner les informaticiens qui testent les nouvelles versions du logiciel (« beta versions », dites aussi « nightlies »[137]) qui comportent des défauts, traités par ces bénévoles.

Celle intitulée « trac » sert à la partie de la communauté qui, comme son nom l'indique, traque justement les défauts du code (ou « bugs »), signalés sous forme de « tickets » (numérotés sur la liste). Le nom vient de l'autre wiki de la communauté dédié à cette fin[138]. Il est spécialement conçu pour les informaticiens, contrairement à MediaWiki. Il s'agit d'une activité de contrôle de la qualité de première importance, sur les versions du logiciel *publiées* (« released ») avec un numéro de sortie, contrairement aux « nightlies ».

Enfin, les modifications du code sont-elles-mêmes signalées sur la liste de discussion de coordination du « concurrent version system » (CVS), repris par la suite sous système « subversion » (SVN).

---

[135] Nous citerons *l'intégralité* de l'url à chaque fois que nous ferons référence à l'un d'eux, contrairement à ce que nous pensions faire. Citer uniquement le numéro obligerait le chercheur à retrouver le mois de référence puis l'url précise dans les archives par tâtonnements, *ce qui occasionnerait d'expérience une perte de temps considérable*. C'est certes inélégant mais hélas indispensable. Le procédé a été retrouvé ailleurs (IANNACCI, 2003).

[136] Source : http://codex.wordpress.org/Mailing_Lists#Hackers .

[137] Il s'agit des versions « nocturnes », i.e. « non visibles à la lumière ». Il arrive aux informaticiens de rester travailler par addiction à leur code la nuit.

[138] Source : http://core.trac.wordpress.org/ .

D'autres listes ont été dépouillées de façon uniquement quantitative, mais elles sont plus secondaires, et nous ne voyons pas l'intérêt d'insister ici sur un sujet aride et de moindre importance.

3.4.2.5. La liste de discussions de la documentation du wiki

La liste de discussions du projet de documentation de WordPress est le lieu d'intervention d'une sous-communauté d'utilisateurs avancés[139] et de toutes les personnes concernées même de façon minimale par la contribution à ce bien collectif, et ce, même sous forme de critiques, de questions, ou d'idées[140]. Il est question aussi de la régulation (comment compose-t-on la documentation ?), du suivi de l'activité (ce qu'on peut appeler « workflow », avec parfois des détails fins), avec des débats d'idées créatifs (« brainstorming »), de suggestions pas forcément prises en compte, des demandes d'expertise et de conseil, des formes de sociabilisation légères, des remarques sur le niveau de qualité de la documentation, etc. . La taille des textes est variable, du simple émoticon approbateur (« smiley »)[141] à des dissertations juridiques complexes de deux pages. L'éventail d'activité et de propos est large. Le rythme d'activité sera étudié par la suite, mais signalons que cette liste est un lieu de réaction très rapide[142]. Enfin, la spécialisation peut prêter à discussions. L'organisation du forum y a été discutée d'une façon significative, avant la création d'une liste spécialisée de régulation. L'importance de la liste de documentation est suggérée par le fait qu'on y discute de certaines orientations stratégiques de la communauté, peut-être de façon incongrue[143]. A contrario, officiellement, il s'agit *du* lieu de prise de décision en la matière, même si on observe des décisions prises ailleurs (externalisation des plugins par exemple).

3.4.2.6. Les archives de chat IRC

Nous avons également pris en compte les « logs » des archives des « chats » « IRC » à notre disposition sur la page d'archivage du Codex : http://codex.wordpress.org/IRC_Meetups C'est là un matériau brut et mal dégrossi, comme le montre la capture d'écran ci-dessous. Il

[139] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-April/001516.html .
[140] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-May/001533.html .
[141] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-January/001208.html .
[142] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-November/001121.html.
[143] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000975.html .

comporte la spontanéité et les conventions et codes associés des conversations en ligne. L'intérêt du matériau est relatif (redondance d'information par comparaison avec la liste de discussions) mais il a offert quelques passages pertinents permettant de mieux comprendre l'organisation du Codex et des traces de décisions non-triviales.

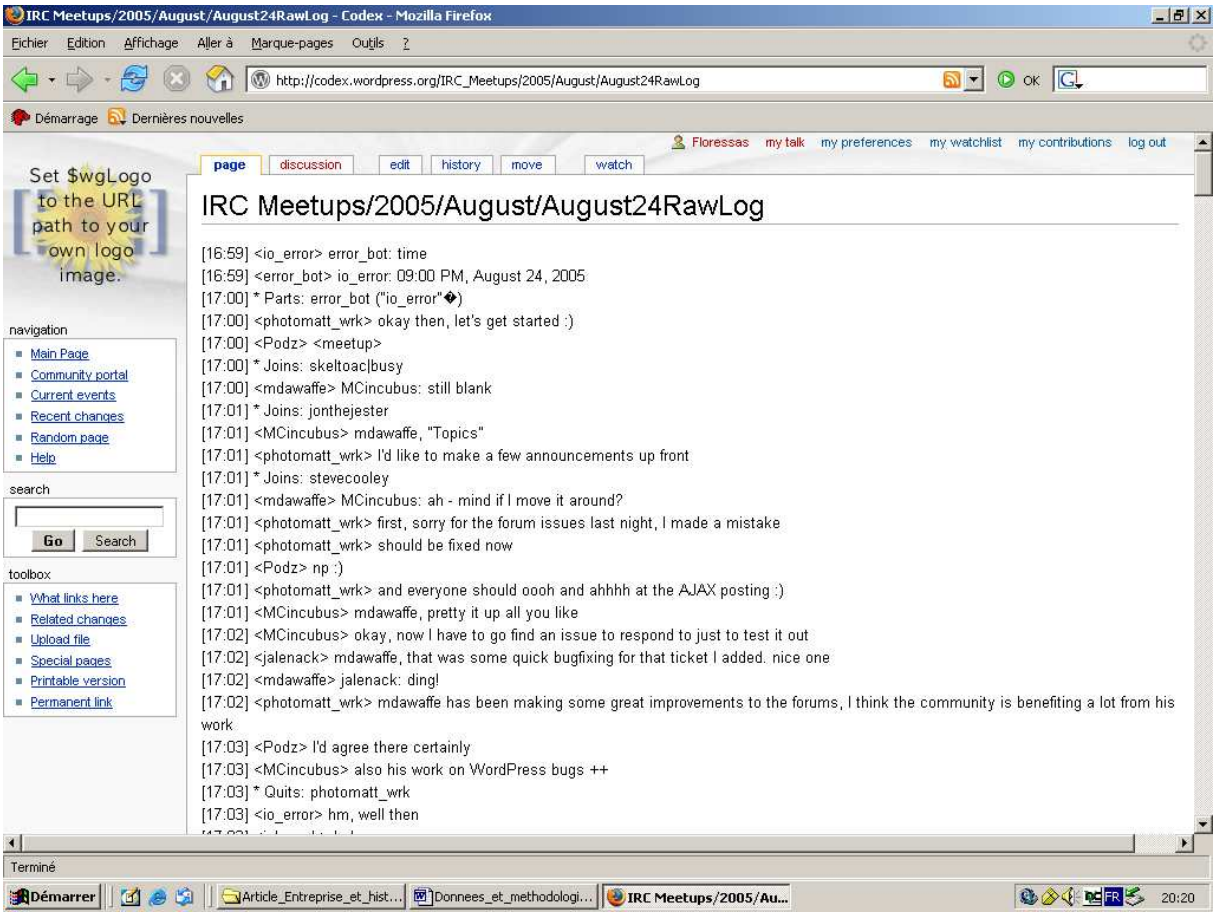


Figure 10 : capture d'écran d'un archivage de « minute » de chat IRC[144]

On observe un écart entre la pratique et la réalité des « réunions ». Officiellement, elles se passent ainsi: "Wednesday at 19:00 UTC we have an IRC "meetup" where everyone gets together in the #wordpress-dev IRC channel to chat about all things WordPress."[145]. Or, des représentations graphiques de l'historique des réunions montrent que le principe des réunions régulières n'a pas été respecté longtemps (voir le détail infra). L'hypothèse de l'abandon de l'archivage régulier n'est pas non plus à exclure même si *rien* ne vient à l'appui de cette idée. La structure des « chats » est informelle, mais elle comporte un « ordre du jour » avec un certain nombre de questions à aborder et proposées à l'avance sur le wiki par les

[144] Source : http://codex.wordpress.org/IRC_Meetups/2005/August/August24RawLog .
[145] Source : http://codex.wordpress.org/IRC_Meetups.

participants[146]. Elles ne sont pas toujours traitées à fond, étant donné que les thématiques reviennent d'une semaine à l'autre, ce qui souligne l'aspect informel des opérations.

3.4.3. Problèmes généraux de critique des sources et biais détectés

Les précédentes études n'ont pas vraiment pris en compte la dimension archivistique de ce type de données. Ce qui différencie une approche historique d'une simple analyse longitudinale est l'attention à la critique des sources, au croisement des données, et une analyse plus fine de la complexité des processus, des rythmes, et des dynamiques. A quel type de questions par exemple telle ou telle donnée permet-elle de répondre de façon satisfaisante ?

3.4.3.1. Biais matériels

Ils sont pertinents pour comprendre certaines analyses mais aussi pour d'autres chercheurs confrontés à ce type de données : on ne sait pas a priori ce que l'on ne sait pas.

Nous entendons par là les biais associés à des problèmes de conservation imparfaite des données, des problèmes d'accès ardu, ou des problèmes liés à des données existantes mais inaccessibles.

Les données ou matériaux peuvent être inaccessibles matériellement, ou pour cause de risque de violation de la vie privée, la correspondance en faisant partie. Par définition, nous manquons toute la correspondance « privé », bilatérale et non « connaissance commune », alors qu'elle peut avoir un impact décisif en matière de prise de décision. *La principale limite de notre travail est qu'il est évidemment impossible de la connaître*. On possède à cet égard de nombreuses preuves de correspondance privée[147]. En effet, lorsque les participants *anticipent* que leurs propos seront entreposés en public, ils peuvent tout à fait éviter d'écrire à tous les membres d'une communauté et interagir par correspondance privée. Remarquons

[146] Par exemple : http://codex.wordpress.org/index.php?title=IRC_Meetups&diff=next&oldid=11629.

[147] Nous mentionnons plusieurs cas qui montrent l'étendue significative du phénomène : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000760.html, http://comox.textdrive.com/pipermail/docs/2004-April/000241.html : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000211.html http://comox.textdrive.com/pipermail/wp-docs/2005-July/000760.html . A contrario, voici un exemple de courriel qui n'aurait pas dû être public : http://comox.textdrive.com/pipermail/docs/2004-May/000278.html .

cependant qu'écrire sur la liste de discussion publique est valorisé (en particulier sur la liste de discussion des hackers, par des statistiques de contributions individuelles publiées en début de mois), ce qui limite aussi l'étendue des interactions privées.

La nécessité de l'analyse de type historique se comprend aussi à cause du risque permanent de pertes de données. C'est le cas par exemple pour d'autres formes d'interactions encore plus informelles, comme les interactions sous forme de dialogue par canal IRC (« chat ») qui ne sont archivées que partiellement. Si les réunions hebdomadaires « en ligne » planifiées semblent complètes et sérieusement archivées, l'analyse de ces mêmes archives révèle l'existence d'un canal de discussion spécifique à l'organisation du Codex. Il n'est pas du tout sûr qu'il s'agisse de celui dont nous disposons des minutes, nous n'avons pu établir de conclusions définitives en cette matière [148].

Si ces sources sont archivées automatiquement, ce qui est la politique de la communauté, on observe aussi que certains courriels disparaissent épisodiquement. On sait que certains courriels ont été perdus[149], d'autres apparaissent à l'état de tronçons[150]. Ces manques demandent un travail d'interprétation supplémentaire. Nous avons une preuve que l'absence de certains courriels dans les archives est due à l'usage intempestif d'HTML pour correspondre[151]. On peut aussi observer des lacunes, comme le « trou » de janvier-février 2005, étant donné qu'il y a eu perte d'une partie des archives entre les deux dépôts[152]. Ces insuffisances ne sont cependant présentes que lors des débuts de la communauté, qui apprend à communiquer et archiver correctement. Que le passé apparaisse à l'état de traces lorsqu'on remonte le temps justifie une approche historique. En revanche, on sait qu'il y a conservation *intégrale* des comptes utilisateurs, alors que la communauté a pensé supprimer ceux des « lurkers »[153]. Cela permet de déduire *l'absence de biais* dû à la conservation intégrale. En revanche, les mises à jour d'un logiciel, comme MediaWiki, provoquent la perte des premières strates des contributions. Les comptes supprimés (ceux des vandales) sont même répertoriés en un lieu adéquat, mais ils ne sont guère passionnants.

L'analyse rigoureuse d'une telle communauté suppose un repérage de toutes les sources disponibles et pertinentes, ce qui ne va pas de soi, surtout sans guide des archives

[148] Source: http://codex.wordpress.org/IRC_Meetups/2006/January/January04RawLog et http://comox.textdrive.com/pipermail/wp-docs/2005-June/000590.html .
[149] Source : http://comox.textdrive.com/pipermail/docs/2004-May/000260.html.
[150] Par exemple : http://comox.textdrive.com/pipermail/docs/2004-August/000469.html .
[151] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-October/001072.html.
[152] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000081.html.
[153] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000308.html.

complet, guide qui est aussi un produit significatif d'une recherche (nous fournissons quelques éléments en annexes). Pour prendre un exemple parlant, trois des premières archives (« docs », « hackers », « CVS ») n'apparaissent plus dans l'entrepôt officiel de la communauté (présenté plus haut) et ont dû être découvertes par déduction et croisement de données à partir de la connaissance des archives précédentes[154]. Si elles ne sont plus d'importance pour la communauté, elles sont indispensables en revanche pour le chercheur en sciences sociales travaillant sur la genèse de la communauté. Cela montre l'importance d'une démarche d'historien de *recherche des sources*, au-delà de l'examen des archives « officielles ». Cette nécessité de trouver les sources indispensables cachées doit être la priorité de chercheurs confrontés à une analyse du « déploiement du numérique ». D'un autre côté, le premier wiki de la communauté, accessible au début de notre recherche, n'est plus accessible sur les serveurs actuellement, une désagréable surprise impossible à anticiper... La seule façon à présent de comprendre les débuts du déploiement d'une solution wiki comme système d'information est, à l'exception du service web.archives.org, d'extraire et analyser les traces présentes dans les listes de discussions correspondant à cette période de l'histoire… dans les archives cachées mentionnées ci-dessus.

3.4.3.2. Biais sémantiques

Ces archives ont un statut particulier dans notre contexte, qui n'est certainement pas un cas particulier dans les archives des communautés: *une pratique d'archivage n'est pas neutre*, même lorsque l'archivage est automatique. Les archives publiques internes comme les listes de discussions sont considérées comme la mémoire *vivante* de la communauté, de ses choix et de ses erreurs, et les nouveaux venus peuvent être invités à parcourir le passé d'une communauté pour *apprendre* (MASAFUMI, 2004), ce qui aurait une influence sur les modes d'interactions et les contenus ajoutés. Cela ne veut pas dire qu'ils le font toujours, même si cela a été observé.

Au-delà, la liste de discussion de la documentation est de nature particulière : elle est la mémoire sélective de ce qui se passe sur le wiki, et donc diffère dans une certaine mesure

[154] Pour donner un exemple de l'importance de la structuration des fonds, il nous semblait hautement improbable que l'activité de programmation n'ait pas été coordonnée par une liste de discussion un an après le début de l'histoire de la communauté. En conséquence, une décision intuitive de fouille de l'historique de la page de présentation des listes de discussions sur le wiki, « pour voir », a permis de mettre à jour le dépôt d'archives désaffecté.

de ce qui y a lieu ; mais elles sont lues par ceux qui y participent, et ce qui s'y dit a certainement une influence sur les processus. Le principal enjeu pour notre recherche est que *cette source d'analyse riche a l'inconvénient de modifier la problématique d'étude du wiki*, car le wiki ne supporte *plus alors un phénomène isolé* ; une liste de discussion modifie son mode de fonctionnement. Ce qui est pour *l'observateur* une archive est pour *le participant* à la communauté un outil et un mode de coordination. Ce problème n'est pas un détail ; cependant, une telle modification du fonctionnement du wiki peut être révélatrice de certaines caractéristiques de son fonctionnement ou son non-fonctionnement réel et général (par comparaison avec ce qu'on croit en savoir), ou sur les complémentarités possibles avec d'autres outils.

Le nombre d'interventions d'un contributeur sur une liste comporte une signification politique : plus un participant parle sur cette agora électronique, plus il est susceptible d'avoir du pouvoir. Ce processus demande examen, qu'il soit explicitement quantifié, comme sur la liste des « hackers », ou non, comme sur les autres listes. Même imprécis, ce phénomène doit être pris en compte. Ainsi, l'un des biais possible de ces données serait leur biais élitiste[155] ardu à identifier, celui-ci pouvant varier au fil du temps : ce genre de considérations serait plutôt le *résultat* d'une recherche. Nous ne connaissons, par ailleurs, que l'opinion de ceux qui prennent la parole dans la liste de discussion, et pas forcément des « lurkers ».

Si des biais d'énonciation sont observés, ce qui nécessite pour la compréhension de certains passages une mise en série des « prises de parole » de chacun, en revanche le nombre conséquent de personnes s'exprimant sur la liste de discussion permet de multiplier les points de vue et de limiter les biais en la matière. Par ailleurs, les listes de discussions « tamisent » les aspects significatifs d'une activité qu'elles coordonnent, ce qui comporte aussi des avantages pour le chercheur. Cela ne veut pas dire que les problèmes économiques auxquels la communauté a été confrontée apparaissent forcément, la prudence s'impose. De même, qu'un des participants évoque un problème de gestion de la communauté ne signifie pas que celle-ci en a clairement conscience ou que le problème a été effectivement traité.

L'une des spécificités des matériaux numériques est que nous n'avons pas connaissance de problèmes dus à l'authenticité des sources : il n'y a pas de « faux ». C'est l'un des aspects de la critique interne.

Cependant, celle-ci conserve toute sa pertinence pour l'interprétation des textes. A ce propos, de même que les historiens de l'Antiquité se doivent de connaître les langues

[155] MILES et HUBERMAN (1994) demandent que ce type de questions soit examiné avec soin.

anciennes, des données aussi volumineuses (plus de 17000 courriels pour la seule liste des « hackers », des millions d'éditions pour le wiki, près de 100.000 fils de discussions sur le forum…) et techniques, comprendre ce qui se passe dans la communauté suppose une *connaissance minimale* en informatique[156], mais pas forcément la maîtrise d'un langage de programmation. Ce point n'est pas un détail, car la compréhension de l'activité n'est pas du tout immédiate. Pour le lecteur, nous avons composé un index minimal placé en annexes. D'autres problèmes, signalés oralement par Paul David dans le cadre de la chaire « Innovation et Régulation des services numériques », comme la gestion des identités multiples, ont été identifiés mais n'ont pas paru insurmontables, et ont été traités par familiarisation avec le contexte. Ce fut le cas lors de l'analyse de la dynamique démographique de la liste de discussion. L'existence potentielle de ce genre de problème est un rappel cependant utile.

3.4.4. Formes et limites des triangulations

Tout travail de recherche doit tenter, autant que faire se peut, une triple triangulation : des théories[157], des outils d'analyses, des sources et données mobilisées. Ce principe au service d'un seul questionnement problématique dont la nature doit évoluer au moins partiellement entre l'introduction et la conclusion. Tout n'est cependant pas faisable ou pertinent.

3.4.4.1. Le croisement des sources

Une pratique fructueuse est de *croiser les sources* : par exemple, il est possible de comparer ce que disent les informaticiens et les contributeurs à la documentation pour comprendre les débuts de celle-ci. Ce type d'approche est utile, comme on le verra, pour montrer par exemple l'idée que l'échec du premier wiki est dû à une problématique d'organisation et non d'incitation, un problème qui a persisté par la suite. Trianguler les sources et les données n'était pas immédiat. Au fur et à mesure de notre recherche, nous nous sommes aperçu que des éléments des sources d'une sous-question de recherche devenaient en fait pertinents pour une autre. Nous y sommes parvenus dans certaines sous-questions de

[156] La connaissance minimale mais systématique des aspects techniques est prônée par BLOCH (1993).
[157] Pour la triangulation des théories, qui est voisine, nous renvoyons au chapitre précédent traitant de la théorie.

recherche, selon diverses modalités : pour l'analyse des routines (chapitre 6), nous avons croisé les données de la liste de discussion de la documentation, les onglets de discussion du wiki, et les fragments issus de la liste des « hackers » et les minutes de chat IRC. Au chapitre 7, nous avons croisé les données des caractéristiques des contributeurs provenant du wiki et celles des autres parties de la communauté (liste de discussion, leurs blogs). Au chapitre 8, nous avons croisé les listes de discussions et un peu analysé le forum. Dans le chapitre 4 de contextualisation et d'histoire de la communauté, nous avons croisé plusieurs sources pour analyser les dynamiques du wiki, pour ne parler que de cela.

3.4.4.2. La triangulation des méthodes d'analyse

Les formes de triangulations des méthodes d'analyse *ne sont pas fréquentes* dans notre champ de recherche (à l'exception par exemple de SACK et al., 2006). Les archives présentées précédemment peuvent être traitées de plusieurs façons, qualitative, quantitative, ou au moyen d'une analyse de réseaux. Une *même source* peut subir au besoin plusieurs traitements différents de cette sorte. C'est le cas pour la liste qui coordonne l'activité auxiliaire de la documentation, qui est traitée sous ces trois modalités. Mais rien n'empêche d'étendre ce principe davantage après la thèse sur des sources où nous n'aurions employé qu'une méthodologie. Notre thèse comporte ainsi une tentative de *triangulation des méthodes d'analyse* : narration, analyses qualitatives guidées par une théorie ou non, quantitative, analyse des réseaux sociaux. Nous nous retrouvons donc dans l'approche proposée par DUMEZ et JEUNEMAITRE (2006), c'est à dire ce qu'ils appellent le « tripode », narration, modélisation, et analyse quantitative au sens large.

Nous n'avons cependant pu effectuer de triangulation des chercheurs, si ce n'est à la fin par de nombreuses relectures bénévoles. La seule intervention technique notable a été l'aide de doctorants du CREA. Nous avons enfin tenté une triangulation des *traditions* de recherche (économie hétérodoxe, tradition du CREA, tradition du CRG, analyse historique), correspondant volontairement ou involontairement à nos parcours antérieurs.

3.5. Des déceptions récurrentes en matière d'outils informatiques de traitement de données

Les insuffisances des programmes informatiques d'extraction automatisée de données des wikis sont une déception pour le chercheur. Les outils d'analyse des wikis sont en version « bêta », ardus à manipuler, en particulier pour un non-informaticien, et adaptés seulement aux projets associés à la fondation MediaWiki, ce qui ne correspond pas aux projets open-source. Des projets comme Autograph se sont spécifiquement penchés sur l'analyse de données complexe issues des données d'historique de wikis, mais pour étudier Wikipedia (voir par exemple NUNES et al., 2008, pour l'outil « WikiChanges »).

Une autre question et déception en matière d'outils est celle du traitement informatique des données, indispensable, en particulier qualitative. Les logiciels de codage[158] sont conçus pour les approches qualitatives usuelles des ethnologues, sociologues et chercheurs en gestion sur des approches voisines, et nous y avons renoncé. Nous avons effectué entre autres une démarche de codage, avec pour objectif de pouvoir retrouver les fragments de textes pertinents. Sur ce point, le bricolage avec les outils bureautiques est satisfaisant, en attendant des outils prenant en compte les spécificités de l'approche historique. Le récent colloque « AnaLogiQual2008 »[159] montre que beaucoup reste à faire en la matière. Nous avons d'ailleurs été confrontés à trois difficultés qui y ont été évoquées : l'analyse de données diachroniques, le détournement d'outils d'usage courant et le traitement de matériaux de type numérique. Signalons que des chercheurs travaillant sur l'open-source ont également eu des difficultés notables (CROWSTON et al., 2007) et finalement renoncé aussi. Ces insuffisances expliquent nos orientations de travail.

Bref, sur ces questions, nous n'avons connu que des déceptions alors que nos attentes étaient fortes. *Il y a donc des travaux d'ingénierie de la recherche utiles à prévoir, en particulier à destination des approches longitudinales et historiques*.

La plupart des présentations et des analyses de données dans toutes les sous-questions de recherche ont été effectuées sous Excel et à la marge sous Word. Nous nous sommes servis aussi de logiciels de statistiques et d'économétrie pour les parties quantitatives.

[158] Nous pensons à ceux qu'on appelle « CAQDAS » (« Computer Aided Qualitative Data Analysis Software »).
[159] http://analyses.ishs.ulg.ac.be/analogiqual/presentation.html (consulté le 14/07/09).

3.6. Synthèse des orientations méthodologiques

3.6.1. Retour sur les présupposés épistémologiques de cette thèse

Aborder les présupposés épistémologiques est l'un des points attendus dans une thèse en sciences de gestion ; nous avons préféré mettre cette question en perspective en fin de chapitre avant synthèse finale. Notre posture épistémologique se caractérise par des formes de « mixité » des approches, pour reprendre un concept usité dans cette discipline. Nous expliquons ici brièvement en quoi, en écrivant en italique les concepts-clefs permettant de se repérer.

L'approche *historique* en système d'information est d'un point de vue épistémologique quelque peu paradoxale. Elle comporte une orientation *interprétative* forte : il s'agit de la familiarisation avec le terrain, l'attention à l'énonciation, et à la critique des sources. Elle comporte également une recherche forte de *l'objectivité* et de la vérité, même s'il s'agit pour une part significative des travaux en histoire d'un *horizon*, ce qui la rattacherait à une tradition plus *objectiviste* en sciences de gestion. Sur ce point, MONOD (2002) tente une réconciliation entre positivisme et approches « interprétativistes », ce que permettrait *justement* selon lui entre autres « l'investigation historique », ce qui va dans le sens du dépassement du paradoxe explicité ici. Cependant, les possibilités de l'histoire en la matière pourraient ne pas avoir été examinées suffisamment, en particulier la limitation de l'utilité de l'histoire à l'établissement du contexte et de la singularité, alors que VEYNE (1971) parlait déjà de spécificité[160]. Mobiliser des outils économiques comporte également ce genre d'ambiguïté, l'orientation économique « standard » (au sens de Favereau) étant *positiviste*, alors que l'approche hétérodoxe tente de prendre en compte la question de *l'interprétation* (DOSSE, 1997). Nous faisons donc usage d'une approche *multi-paradigme* (au sens de THIETART, 2007), entre positivisme (recherche du vrai indépendamment du chercheur) et interprétativisme (nécessité d'une approche compréhensive), sans privilégier l'un ou l'autre.

Si on se réfère à d'autres catégorisations proposées par THIETART (2007), notre approche est orientée vers l'étude de processus plutôt que vers celle du contenu et de variables, même si ce n'est pas exclu en particulier dans les chapitres 4 et 7 ; étudier le comportement d'objets dans le temps est cohérent avec l'approche historique. Ensuite, notre

[160] Les questions d'épistémologie en système d'information mériteraient des approfondissements bien au delà de telles considérations sur les possibilités de l'histoire, mais ce n'est pas notre sujet.

orientation est surtout *descriptive*, ce qui est cohérent avec les méthodes d'analyse choisies. Cela se fait au dépend, dans une certaine mesure, de l'approfondissement des questions d'efficacité et d'efficience. Elle consiste à étudier une forme organisationnelle originale[161], comme un entomologiste (mais aussi de tenter d'expliquer certains de ses aspects). Cela signifie que notre travail comporte des limites en matière de généralisation des résultats, même si nous proposons un apport en la matière, et que des régularités sont observées. Nous avons fait choix d'étudier en profondeur un phénomène de gestion donné, de la façon la plus exhaustive possible sur ses dimensions pertinentes.

Enfin, on l'aura compris, nous nous proposons une approche *à la fois qualitative et quantitative* sur des *données primaire*s. Plus précisément, notre approche porte surtout sur le *niveau « micro »* (questionnement économique et micro-histoire), même si nous faisons varier les échelles en passant du qualitatif au quanti par exemple, ou en multipliant les angles de vue. Nous nous proposons également de suivre deux orientations distinguées par THIETART (2007), à la fois *explorer* (par exemple quelles motivations apparaissent dans les traces qualitatives au chapitre 5) et *tester*, autant en matière qualitative (par exemple tester l'hybridation de routines au chapitre 6), comme en matière quantitative, par exemple l'homophilie en matière d'interaction au chapitre 7). Nous ne mentionnons certes pas ici tous les aspects de notre travail.

3.6.2. Tableau de synthèse des usages des sources

Nous proposons ci-dessous une vision d'ensemble des usages ultérieurs des sources :

[161] Rémi Maniak avait tenté d'effectuer de façon informelle une typologie des trois façons de faire de la recherche en gestion au CRG qui incluait celle-ci. Nous la reprenons bien qu'elle n'ait pas été développée par écrit.

Sources	Chapitres où elles sont mobilisées	Approche principale pour chacun des chapitres
Forum	9	qualitative.
Web-archives	5	quantitative.
Listes "docs" puis "wp-docs".	5,6,7,8	qualitative, qualitative, qualitative, quantitative.
Listes des informaticiens	5, 9	quantitative, qualitative
Minutes de "chat IRC"	6	qualitative
Historiques de wiki	6	quantitative et qualitative

Tableau 1: synthèse de l'usage des sources numériques

On constate à nouveau le caractère central des listes de discussions parmi les sources, en particulier la liste servant à la coordination de la documentation, mais aussi la diversité des sources.

3.7. Conclusions

Nous n'avons ici donné que des orientations générales, qui aboutissent à une utilisation dans notre thèse du genre « micro-histoire » contemporaine et économique pour traiter un cas unique sous plusieurs angles, au moyen de plusieurs sources, et par différentes approches qualitatives et quantitatives. Le choix du genre historique fait entre l'autre l'originalité de ce travail, qui a aussi valeur de test pour une méthodologie. La suite de cette thèse confirmera sa pertinence, et la seule limite que nous soulignerons sera peut-être celle de l'analyse des motivations des contributeurs au chapitre 5.

Nous avons donc introduit le fait qu'à partir de tels matériaux à analyser, qu'une approche sérieuse des dynamiques et de l'efficacité des communautés en ligne suppose en fait nécessairement une approche historique par comparaison avec les autres approches (méthodologiques ou techniques) existantes. Une telle approche devait être précisée davantage, les frontières entre sciences sociales s'étant estompées. *Nous développerons naturellement dans le reste de la thèse* : connaissance des événements sur la moyenne durée, mise en évidence de périodes, des ruptures, ou au contraire de permanences structurelles, ainsi que des phénomènes de « dépendance par rapport au chemin », connaissance des archives, importance de la critique des sources, spécificité de l'analyse des traces et des comportements effectifs (motivations, routines), mise en évidence des spécificités des phénomènes étudiés, et

de leur complexité, discussion sérieuse sur ce qu'on peut savoir ou ne pas savoir à partir des données, et enfin attention aux « humbles » au rôle économique essentiel, tout cela fait partie de la tradition et du savoir-faire de l'historien. La question de l'organisation de toute communauté n'est pas triviale, et l'analyse historique comporte donc un avantage absolu certain pour traiter de questions qui se doivent d'être nuancées.

Ce chapitre et ce qui suit a aussi pour but d'inciter par l'exemple d'autres chercheurs à poursuivre dans cette voie pour analyser le « déploiement du numérique » et à prendre au sérieux l'analyse historique qui va au-delà du concept « d'étude longitudinale » usité en sciences de gestion. Le risque de pertes d'archives rappelle aussi qu'il y a urgence.

Chapitre 4 : contextualisation du cas par quelques éléments d'histoire de la communauté

4.1. Introduction

Nous proposons ici, ce qui est attendu dans une thèse, un chapitre de contextualisation de notre travail : c'est essentiellement une présentation historique de notre communauté. Nous rappelons que les études longitudinales des projets open-source ne sont pas fréquentes, et celles correspondant aux tâches « auxiliaires » inexistantes.

Le processus observé, pour synthétiser, est une trajectoire de centralisation et d'acceptation progressive du wiki comme système d'information central enchâssé dans la communauté WordPress. Cette trajectoire d'un SI interne à une communauté, une coévolution entre phases de contributions et d'organisations peut être considérée, pour la nommer autrement, comme un « cycle de vie » d'un wiki[162].

Ce chapitre est donc une tentative de compréhension et contextualisation d'un tel phénomène, par des propositions de périodisations de la communauté. Nous avons conçu ici un objectif limité, présenter les grands repères historiques et tenter de comprendre l'articulation entre l'histoire de la communauté et celle de sa partie qui travaille sur le wiki. Il n'est donc pas question ici de proposer une histoire du projet *complet*, qui demanderait une thèse en soi. Nous ne présentons que deux tentatives d'analyse historique (événementielle, quantitative) sur notre sujet précis. Le but est de se faire une idée des traits saillants de la communauté, de ses grandes tendances[163].

Nous mentionnons quelques éléments de méthodologie. Il s'agit d'un essai de micro-histoire, qui consiste à examiner un phénomène dynamique de près, et à changer d'échelle. Un tel travail suppose un travail conséquent de dépouillement d'archives[164]. Nous insistons, en ce qui concerne le récit linéaire d' « histoire immédiate », l'*appareil critique est ici fort conséquent, comme dans une thèse d'histoire.* Le lecteur peut tout à fait le négliger et n'a pas à s'en effrayer.

Nous proposons donc ici plusieurs éléments :

[162] Nous ne pensons pas à ce stade à un idéal-type général du cycle de vie des wikis.
[163] Nous précisons donc que de nombreuses thématiques essentielles apparaissent de façon dense et allusive, ce qui peut décevoir, frustrer, et appeler des approfondissements. Elles seront effectivement traitées par la suite.
[164] Pour ce faire, les sources mobilisées sont principalement les archives des listes de discussions, le blog de la programmation, quelques données tirées de web.archives.org…

- Une présentation de la communauté, du wiki, de la start-up : cette partie donne quelques éléments généraux pour le lecteur qui ne connaît pas le sujet. C'est l'arrière-plan de notre travail.
- Par opposition, une partie d'approfondissements historiques selon deux optiques, un récit linéaire « événementiel » et des séries chronologiques sur plusieurs parties de la communauté.

4.2. Présentation de la communauté, du produit (le blog WordPress), et de la start-up « Automattic ».

4.2.1. De la communauté WordPress au succès de son produit

La communauté WordPress, étudiée ici sur la totalité de la période 2003-2008, est devenue un acteur incontournable en matière de conception de moteur de blog open-source. La datation des débuts *exacts* de la communauté n'est pas précise au jour près, mais cela n'a guère d'incidence scientifique[165]. La communauté a cependant commencé à émerger à coup sûr entre mi-mars et mi-avril 2003[166] : la mise en ligne du site date du 28 Mars 2003 d'après le site Alexa, la première trace de code spécifique au logiciel date du 16 avril 2003[167]. En revanche, la « jeune pousse » (« start-up ») Automattic est lancée séparément le 20 décembre 2005[168].

[165] Ce n'est pas là la seule incertitude. Les origines du produit WordPress (les origines du code) ne sont pas tout à fait claires : s'agit-il de bb2, comme il est de notoriété publique dans la communauté, ou de Greymatter ? (source : http://comox.textdrive.com/pipermail/docs/2003-November/000033.html).
[166] L'examen seul des archivages par archives.org (http://web.archive.org/web/*/http://wordpress.org/) induit un léger biais. On ne saurait trop insister sur *l'insuffisance* de ce type d'outil de collecte systématique et officielle pour aborder des problématiques fines.
[167] Source : http://web.archive.org/web/20030701214545/wordpress.org/development/ .
[168] Source : http://web.archive.org/web/20051220125353/http://automattic.com/ .

Figure 11 : la première capture d'écran disponible de la communauté[169].

Une requête en décembre 2005 sur Google mentionne 59 000 000 d'entrées à « WordPress », une autre en septembre 2009 306 000 000. Il s'agit donc d'un produit d'une notoriété considérable. WordPress est désormais l'un des « moteurs de blogs » les plus connus dans les comparatifs proposés sur le web[170]. Les utilisateurs ne sont plus seulement des blogueurs du monde de l'Open-Source, bien que cela ait largement contribué au succès du logiciel[171]. Le travail des contributeurs est respecté pour sa qualité, alors que la concurrence dans le domaine des blogs, un outil à présent largement répandu, est apparue[172]. Pour donner une idée du niveau d'aboutissement du produit et tout simplement donner au lecteur un idée

[169] Source : http://web.archive.org/web/20030618021947/http://wordpress.org/
[170] Par exemple : http://www.ojr.org/ojr/images/blog_software_comparison.cfm (consulté le 02/09/09).
[171] Les puristes du logiciel libre considèrent leur travail comme un art, et le fait qu'un logiciel soit libre est souvent un signal évident de qualité. C'est le cas ici.
[172] Sur Framasoft, portail connu des logiciels libres, les deux principaux moteurs sélectionnés sont WordPress et Dotclear, mais il y en a 12 autres. D'autres moteurs de blog sont propriétaire, nous ne les avons pas dénombrés. Et rien n'oblige à choisir un blog comme CMS, des substitutions pouvant s'opérer entre les produits. (Source : http://www.framasoft.net/rubrique371.html , consulté le 02/09/09.). Ainsi, la concurrence entre les produits en question est *conséquente*.

concrète, voici un utilisateur fameux de WordPress, qui a payé la start-up WordPress.com pour obtenir un service « sur mesure » :

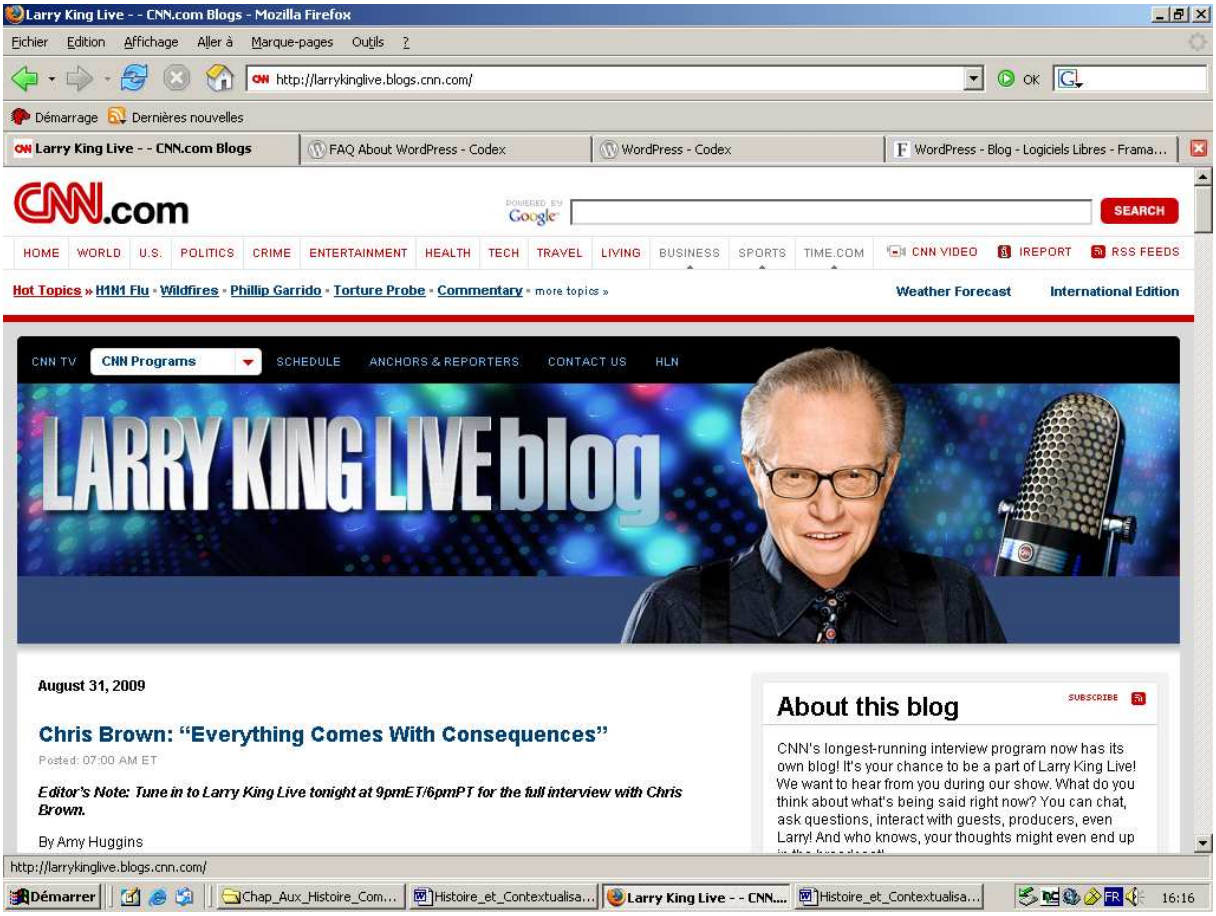


Figure 12: Larry King, célèbre présentateur américain, utilisateur de WordPress[173]

Parmi les autres utilisateurs connus de WordPress nous pouvons citer des organisations connues, comme le New York Times, Second Life, l'université de l'Arkansas, le gourou du web Jimmy Wales, etc[174]. .

Nous présentons brièvement les caractéristiques du produit, qui sont indispensables pour comprendre le succès du logiciel. Sur ce point, nous citons deux fragments complémentaires:

> « WordPress was born out of a desire for an elegant, well-architectured personal publishing system built on PHP and MySQL and licensed under the GPL. It is the official successor of b2/cafelog. WordPress is fresh software, but its roots and development go back to 2001. It is a mature and stable product. We

[173] Source : http://larrykinglive.blogs.cnn.com/ (consulté le 02/09/09).
[174] Le lecteur qui désire obtenir d'autres exemples en trouvera ici : http://wordpress.org/showcase/ (consulté le 02/09/09).

hope by focusing on web standards and user experience we can create a tool different from anything else out there." [175]

"WordPress is publishing software with a focus on ease of use, speed and a great user experience. WordPress is blessed with an active community, which is the heart of open source software."[176]

Le produit comporte de nombreuses dimensions, au-delà des possibilités classiques des blogs : son positionnement est haut de gamme, il est très esthétique, comporte de nombreuses fonctionnalités que permettent le Javascript, la conception légère du code-source permettant d'y ajouter nombre de fonctionnalités comme les thèmes et les plugins. Il faut aussi insister sur la concomitance entre le développement de WordPress avec la montée du langage PhP.

4.2.2. Le système d'information de la communauté WordPress

Nous proposons une vision d'ensemble très succincte et élémentaire du système d'information de la communauté WordPress, telle qu'elle apparaît à la fin du travail de thèse. Elle comporte des utilisateurs, et parmi eux des contributeurs potentiels, qui interagissent avec des entrepôts d'information : ils peuvent y contribuer, les utiliser, ou réguler les activités précédentes :

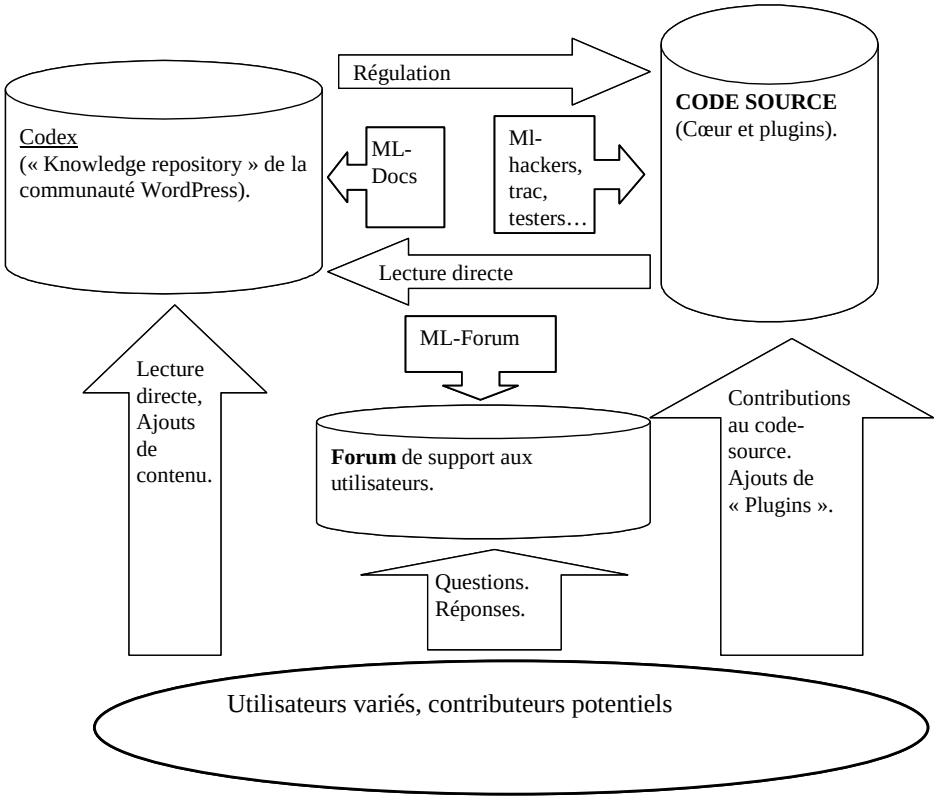


Figure 13 : schéma simplifié du système d'information de WordPress

[175] Source : http://wordpress.org/about/ .
[176] Source : http://codex.wordpress.org/WordPress (consulté le 02/09/09).

Légende :

- Cylindre : entrepôt de données.
- Flèche : activité.
- Carré avec flèche : activité de régulation.
- Ovale : population d'intervenants, qui concerne *toutes* les flèches[177].

Nous commentons ce graphique synthétique. La communauté comporte *trois entrepôts* de données :
- l'un qui regroupe l'ensemble du code-source du logiciel, c'est-à-dire le programme du logiciel ;
- l'autre, le forum (où sont posées les questions des utilisateurs qui ont des problèmes) qui regroupe l'ensemble des fils de discussions formés de posts (questions) et de réponses ;
- enfin, le wiki qui centralise la documentation au sens large concernant le logiciel, et qui nous concerne en premier lieu.

L'activité de chaque entrepôt est coordonnée par une (forum, documentation) ou plusieurs listes de discussions (programmation). Les contributeurs peuvent lire et contribuer au wiki, poser des questions et donner des réponses sur le forum, contribuer au code, soit en améliorant le code-source, soit en proposant des « plugins ».

4.2.3. La start-up

Notre thèse porte sur un aspect du système d'information de la communauté WordPress, et non sur l'entreprise née suite au succès de la communauté en tant que telle. Mais nous devions évoquer son existence et l'évoquer, car il ne s'agit pas seulement d'une « toile de fond » : la présence d'une organisation marchande peut avoir un effet dissuasif sur les contributions bénévoles (« crowding-out effect »). Les sources montrent que la frontière entre la communauté et l'entreprise marchande se doit d'être claire, au point que cela peut affecter sémantiquement la documentation, comme le montre ce commentaire d'un des administrateurs, qui fait donc suffisamment autorité et qui date des débuts de la start-up :

[177] La communauté d'utilisateurs et de contributeur intervient dans l'ensemble des activités, i.e. sur chaque flèche.

> “This is a nice doc (looks like a lot of time was spent on it), but as the initial focus is placed on setting up a wp.com blog, I have a wee problem linking it from the Codex. Brandon, could you split this into two documents, one for WordPress installation and one on wp.com? We constantly try to define where the line for support of WordPress (the software) ends and wp.com begins, and something like this will only help to blur it. -Kaf”[178]

Ainsi, un contributeur bénévole avait écrit un document qui a dû être restructuré à cause de cette politique de séparation entre communauté et entreprise.

Voici à quoi ressemble le site d’Automattic :

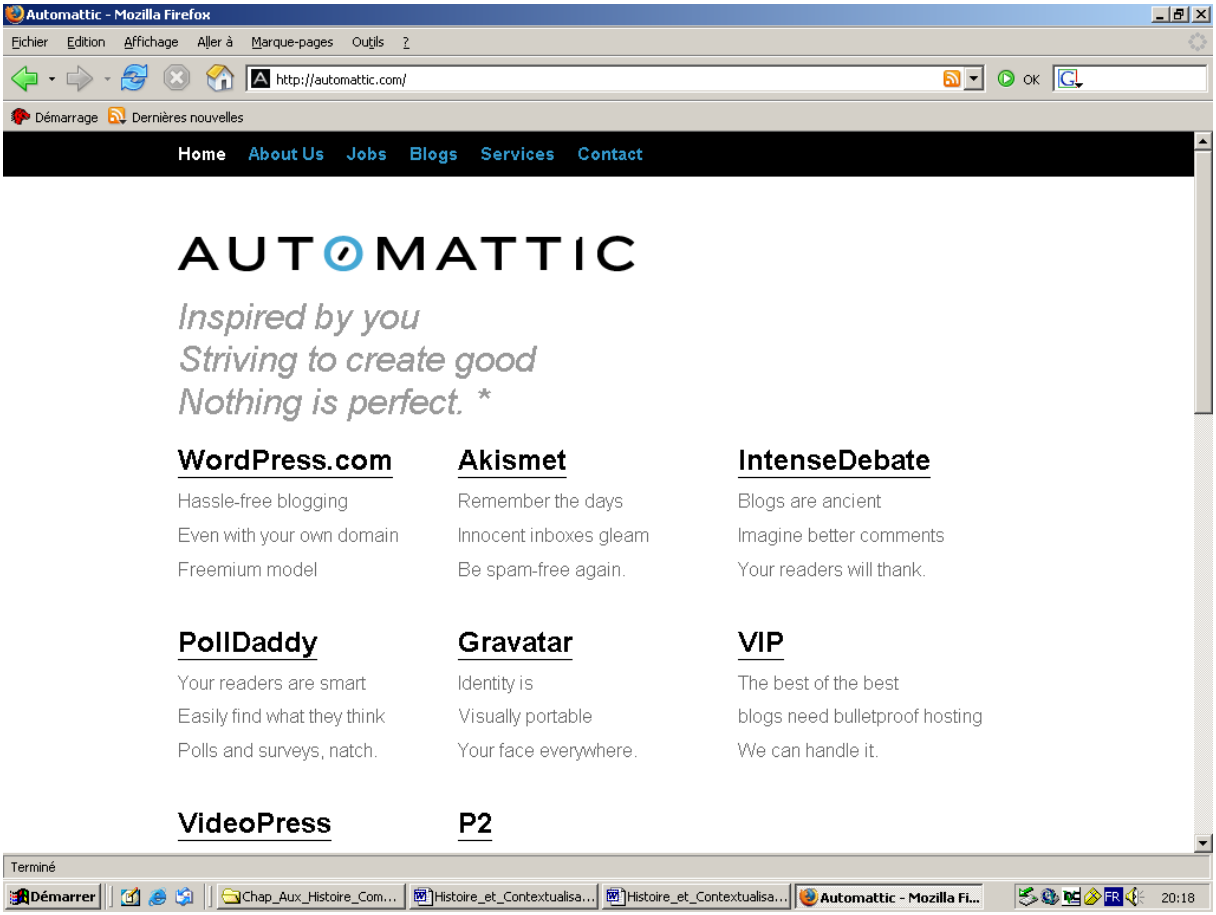


Figure 14: capture d’écran du site d’Automattic (http://automattic.com/, le 07/09/09)

La start-up « Automattic »[179] est lancée le 20 décembre 2005[180]. Elle comprend actuellement plusieurs marques : WordPress.com, qui est une plate-forme d’hébergement de

[178] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-April/001318.html .
[179] Quelques mots sur le nom de la société : « mattic » vient du prénom « Matt », celui du fondateur du logiciel WordPress.
[180] Source : http://web.archive.org/web/20051220125353/http://automattic.com/ .

blogs, PollDaddy, IntenseDebate, les deux dernières étant des acquisitions[181]. Une première tentative, la constitution de « WordPress Inc », envisagée début 2005, n'a pas vu le jour[182]. Le deuxième essai a été gagnant, ce qui montre aussi que l'idée de créer une start-up à des fins lucratives était mûrie. Voici quelques indications sur le modèle d'affaire de l'entreprise:

> "Almost everything on WordPress.com is free, and things that are currently free will remain free in the future, but we do offer paid a la carte upgrades for things like CSS editing and custom domains. How do we pay for everything? WordPress.com is run by Automattic which currently makes money from the aforementioned upgrades, blog services, Akismet anti-spam technology, and hosting partnerships."[183]

L'activité marchande se concentre donc sur le service utilisateur (« support », avec obligation de moyen)[184]. La question sous-jacente à un tel modèle d'affaire qui suppose une articulation entre entreprise et communauté est de savoir comment un tel modèle est viable sans mettre en oeuvre ce que décrit les modèles de l'économie des contrats, à savoir rendre l'offre gratuite peu attractive pour inciter les utilisateurs à payer un service sur mesure. D'un autre côté, l'usage par de futurs clients de WordPress dépend du nombre d'utilisateurs, ce qui n'incite pas à négliger la communauté (même si des défaillances ont été repérées en la matière). Avoir une bonne communauté ne risque-t-il pas de créer un phénomène de cannibalisation en rendant inutiles les services de la start-up ? Des travaux supplémentaires sur ce type de questionnements économiques où les deux incitations jouent en sens contraire sont souhaitables.

La start-up fonctionne probablement sur le mode de l'adhocratie guidée par un esprit communautaire. L'internationalisation de la start-up qui fonctionne selon le principe de l'organisation virtuelle est visible parmi son personnel, ce qui est un aspect d'une organisation virtuelle : nous notons un bulgare, un canadien, etc. Lors de l'apparition de l'organisation, le trombinoscope des quelques membres de l'équipe comprenait un personnage qui montrait au premier plan sur sa photo ses pieds en *chaussettes*, confortablement installés sur son bureau[185]. Cela se passe de commentaires. Elle comprend 20 personnes en avril 2008, et les effectifs n'ont guère changé depuis. Mais l'esprit des « fiches de postes » présentant le personnel est assez amusant et montre clairement que l'esprit communautaire est resté, comme celle-ci :

---

[181] On remarquera aussi des partenariats avec des sociétés, soit pour la publicité, soit pour et l'hébergement, avec Yahoo en particulier.
[182] Source: http://codex.wordpress.org/User:Matt/WordPress_Inc_Story .
[183] Source: http://wordpress.com/about/ .
[184] http://automattic.com/services/support-network/ .
[185] Nous n'avions pas conservé de capture d'écran et hélas pas retrouvé la page dans les archives.

> “Happiness Engineer
> Blogs:MarianneMasculino.com
> Marianne was born in Germany, yet grew up in exotic locales like Elizabethtown, KY and Houston, TX. She now happily resides in San Francisco. A designer at heart and by trade, she’s joined the team to sprinkle happiness among the WordPress community. Marianne has an affinity for useless celebrity facts and makes a mean white chocolate bread pudding.”

Voici celle du fondateur, dans les archives à la date du 15/06/2006 :

> “Matt Mullenweg — CBBQTT
> Blogs: Photo Matt, Matt on WordPress
> As the Chief BBQ Taste Tester of Automattic [CBBQTT!], Matt travels the world sampling cuisine and comparing it to the gold standard of Texas BBQ. Although he originally aspired to be a jazz saxophonist, Matt somehow wound up studying economics which took him to Washington D.C. where he began taking pictures and blogging. The rest, as they say, is mystery. He lives in San Francisco and has a crush on Audrey Hepburn in Breakfast at Tiffany’s.”[186]

Nous ne savons pas dans quelle mesure cette pratique est originale dans les start-up américaines qui suivent le modèle de l’organisation virtuelle. L’étude d’une start-up du Web 2.0. est donc un objet de recherche intéressant et qui mériterait de plus amples approfondissements, en particulier par le dépouillement systématiques des données collectées par archives.org. Mais nous n’irons pas au-delà dans la présente thèse qui ne porte pas dessus, et nous nous concentrons sur la documentation dans ce qui suit.

4.3. L’histoire de la documentation est celle de l’acceptation progressive du wiki

4.3.1. Remarques méthodologiques préalables[187]

Nous proposons ici une tentative de mise en cohérence de la trajectoire de l’activité de coordination du système d’information fondé sur le wiki de la communauté. Il s’agit d’un *récit*, un exercice qui n’est pas facile à partir d’un sujet et d’un matériau inhabituel, la constitution d’un système hypertextuel. Par comparaison, les séries chronologiques posent moins de problèmes. Il s’agit aussi d’une tentative de *périodisation*, à partir du contenu

[186] Source : http://web.archive.org/web/20060615010800/http://automattic.com/about/ .
[187] Nous nous sommes beaucoup inspirés, consciemment ou inconsciemment, de VEYNE (1971), un ouvrage de référence toujours d’actualité selon nous et rarement égalé, pour mener à bien cet exercice.

qualitatif : le découpage proposé sera rendu visible par des titres. C'est aussi un travail de *sélection* de ce qui apparaît être le plus saillant dans les archives sur le plan événementiel. C'est ainsi que nous avons procédé : nous avons repéré les événements les plus saillants, en gardant des références sous formes d'URL, et nous avons effectué une synthèse et mise en cohérence. Lors des périodes dégagées, les événements ne sont pas toujours totalement linéaires, à quelques jours près, ce qui n'est pas grave, et un tel écart ponctuel est facilement repérable à partir des numéros d'URL. Tentant de dégager des régularités, nous avons accentué le désir de rendre les événements compréhensibles, quitte à arrondir les angles. L'arbitrage était délicat mais incontournable.

Les limites de l'exercice sont la conséquence de ce choix narratif: la mise en cohérence peut être excessive ; on peut aussi estimer qu'il y a des biais faute de compréhension complète de ce qui se passe, le contenu n'étant pas facile d'accès sur certains points. Un tel exercice est cependant *indispensable* pour donner une idée concrète de l'activité et de ce qui s'est passé. Un autre choix était possible, à savoir découper intégralement l'activité de coordination par thèmes, dont on ferait la chronologie. Une telle orientation sera en fait proposée sur certains points approfondis par la suite. Mais il a quand même semblé important d'opérer une synthèse plus *dynamique*, ce qui permet aussi de mettre les événements en perspective. Si ce n'est pas là la seule méthodologie adoptée dans cette thèse, l'exercice se défend, ce d'autant plus que nous n'avons pas connaissance d'une telle tentative dans d'autres travaux partant des listes de discussions archivées : après coup, il apparaît que l'exercice est au moins aussi difficile qu'un travail économétrique.

4.3.2. Une première tentative de documentation au moyen d'un wiki

Les premiers échanges publics concernant la documentation datent du 13 novembre 2003[188]. Les principaux coordinateurs de l'activité sont alors Carthik, un thésard canadien en intelligence artificielle, Matt Mullenweg, le chef de projet de WordPress, ancien étudiant en économie, saxophoniste et « geek », et Podz, un homme exerçant le métier de « sage femme ».

[188] Il est toujours possible qu'il y en ait eu avant, mais il est peu probable que cela ait été significatif, il n'est en effet pas dans l'intérêt de la communauté d'agir de façon non collaborative pour concevoir un objet aussi ingrat. Cela ne porte certainement pas à conséquence.

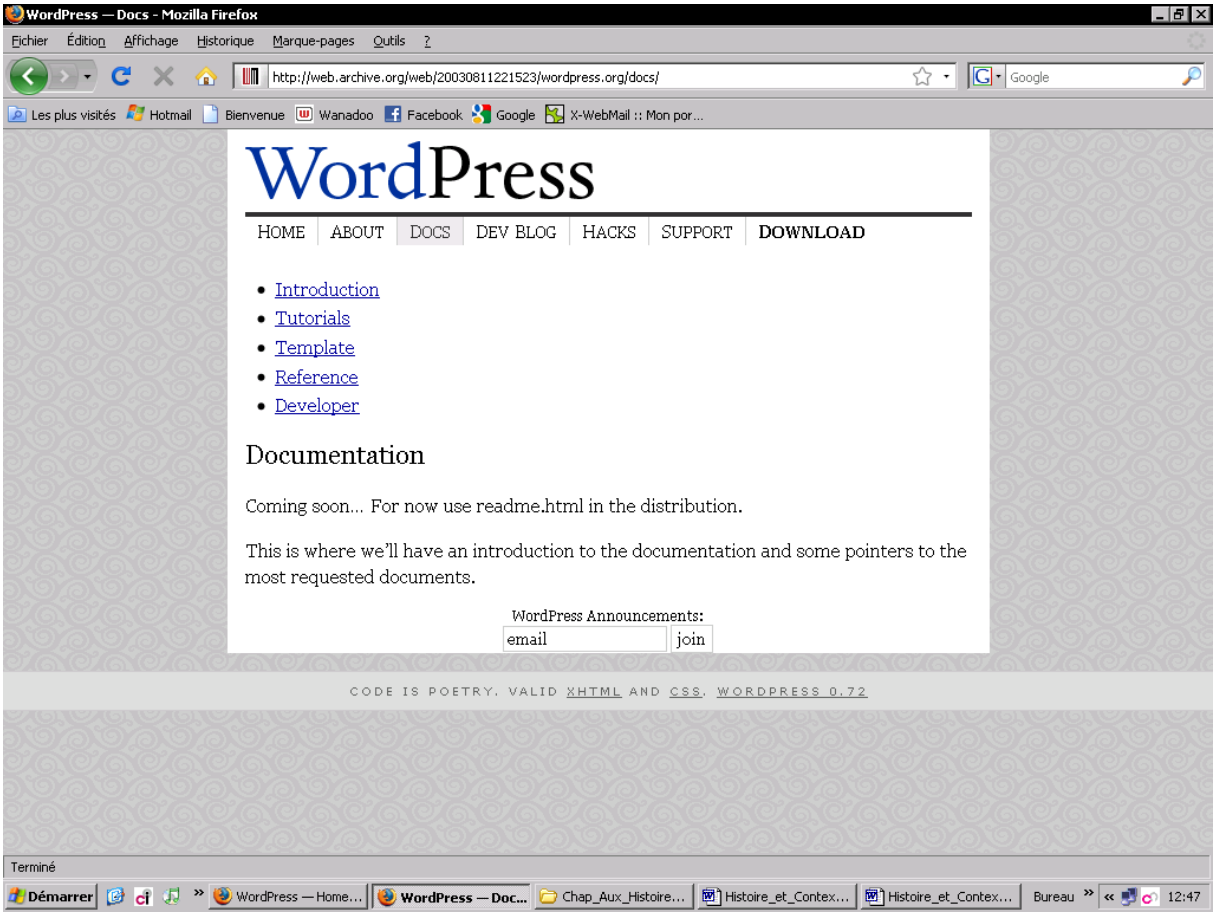


Figure 15 : les premières ébauches de documentation[189]

Le travail en question consiste d'abord à répertorier les thématiques pertinentes sur le contenu du forum, à sociabiliser les contributeurs (les qualifications de chacun), à préciser quelques éléments sur la vision du Codex (par Matt et tout particulièrement Cena[190]), et à accepter au moins le principe de centralisation de la documentation[191]. Un wiki n'a pas été mis en place tout de suite, et il fallait alors passer par des intermédiaires (comme Cena) pour effectuer des modifications sur les documents[192]. Une structure par FAQ a été la première source informationnelle envisagée à partir de l'information tirée du forum[193] et la petite équipe l'organise directement sur la liste de discussion, qui sert de « tableau noir » improvisé[194]. Les premières évocations du wiki datent de novembre 2003, le moteur

[189] Source : http://web.archive.org/web/20030811221523/wordpress.org/docs/ .
[190] Source : http://comox.textdrive.com/pipermail/docs/2003-November/000007.html .
[191] Source : http://comox.textdrive.com/pipermail/docs/2003-November/000008.html .
[192] Source : http://comox.textdrive.com/pipermail/docs/2003-November/000022.html .
[193] Source : http://comox.textdrive.com/pipermail/docs/2003-November/000024.html .
[194] Source : http://comox.textdrive.com/pipermail/docs/2003-November/000033.html .

« PmWiki » n'a pas été retenu. La licence choisie pour le logiciel est la GPL[195]. La première mise en ligne de la documentation date du 10 décembre[196] : elle ne s'est pas faite sur un wiki, ce fait est certain et confirmé par web.archives.org[197]. Le lancement de la première version de la documentation devait coïncider avec la version 1.0 du logiciel.

Un premier wiki sous moteur « PhPWiki » a été mis en place le 18 décembre 2003[198] ; le principe du wiki a été choisi par besoin d'une base de donnée centralisée, facile à éditer et dont les modifications pourraient être suivies[199]. Dès ce moment, on envisage une première documentation pour programmeurs[200], ce qui est le corrélat de la prise en compte de la pluralité des auditoires. On observe aussi les premiers débats sur l'organisation de l'activité de documentation et la mise en avant de l'idée de « méritocratie »[201], en ce sens que ceux qui auraient le pouvoir seraient ceux qui ont le plus contribué. Cependant, en cette période, le wiki n'est alors pas considéré comme une documentation officielle[202]. L'ajout de contenu est bien lancé et on observe de premiers actes de destruction de contenu[203]. La montée en puissance de la documentation et cette mise en place a entraîné des discussions sur sa licence[204], avec la prise de conscience des risques possibles d'opportunisme avec l'écriture d'un ouvrage à partir d'elle[205] (la licence Creative Commons est alors exclue d'emblée). La communauté connaît ses premiers problèmes d'élimination du spam depuis sa précoce montée en popularité. L'usage du wiki pour désengorger le forum est envisagé[206], forum dont l'organisation commence à être discutée sur la liste de documentation en parallèle[207]. Ainsi, le processus de documentation est lancé, avec plusieurs demandes de documentation sur des points spécifiques ou récurrents. En parallèle du wiki, sont composés d'autres textes, comme les premiers brouillons de documentation personnelle de Podz[208], un des gros contributeurs au forum et au wiki. L'organisation se poursuit, avec la socialisation (présentation et mention de compétences) de nouveaux membres (et futurs gros contributeurs), une proposition d'utiliser DocBook plutôt que le wiki rapidement écartée, et les premières traces de routines de

[195] Source : http://comox.textdrive.com/pipermail/docs/2003-November/000034.html .
[196] Source : http://comox.textdrive.com/pipermail/docs/2003-December/000043.html .
[197] Source : http://web.archive.org/web/20040629040432/wordpress.org/docs/ .
[198] Source : http://comox.textdrive.com/pipermail/docs/2003-December/000072.html .
[199] Source : http://comox.textdrive.com/pipermail/docs/2003-December/000066.html .
[200] Source : http://comox.textdrive.com/pipermail/docs/2003-December/000080.html .
[201] Source : http://comox.textdrive.com/pipermail/docs/2003-December/000082.html .
[202] Source : http://comox.textdrive.com/pipermail/docs/2003-December/000087.html .
[203] Source : http://comox.textdrive.com/pipermail/docs/2004-January/000109.html .
[204] Source : http://comox.textdrive.com/pipermail/docs/2004-January/000125.html .
[205] Source : http://comox.textdrive.com/pipermail/docs/2004-January/000137.html .
[206] Source : http://comox.textdrive.com/pipermail/docs/2004-February/000152.html .
[207] Source : http://comox.textdrive.com/pipermail/docs/2004-February/000160.html .
[208] Source : http://comox.textdrive.com/pipermail/docs/2004-April/000177.html .

rédaction[209] après des discussions sur la manière de faire. Il faut noter qu'au début de cette documentation, on ne voulait pas d'un manuel[210], une idée qui émergera plus tardivement. En matière de contenu, on observe en avril un sérieux travail sur le guide des options du logiciel[211], une demande de création d'un glossaire et de FAQ, en même temps que l'arrivée de nouveaux volontaires[212] et de l'organisation de la communauté avec la publication des « guidelines ».

Figure 16 : état du wiki le 24 mars 2004[213]

A partir de mai 2004, on observe une *accélération* en matière de contributions par rapport aux premiers tâtonnements précédents, en relation avec le succès de plus en plus

[209] Source : http://comox.textdrive.com/pipermail/docs/2004-April/000207.html .
[210] Source : http://comox.textdrive.com/pipermail/docs/2004-April/000193.html .
[211] Source : http://comox.textdrive.com/pipermail/docs/2004-April/000244.html .
[212] Source : http://comox.textdrive.com/pipermail/docs/2004-April/000241.html .
[213] Source : http://web.archive.org/web/20040323105321/wiki.wordpress.org/ .

important du projet[214]. La première apparition de la documentation des fonctions[215] et la mise en route du listage des plugins[216] datent de mai, ainsi que le lancement d'un manuel utilisateur sur le wiki[217]. Les demandes de collaborations par vérifications du contenu se multiplient[218], ainsi que la détection des informations les plus demandées sur le forum[219]. En matière de structuration du contenu grandissant, on observe en particulier le lancement des sous-pages sur le wiki[220], une idée qui se révéla malencontreuse.

En effet, une phase de désorganisation et de remise en question est observée en mai et juin 2004, probablement en contrepoint avec la phase précédente. Certes, l'apport d'information se poursuit, avec un travail sur la documentation de l'installation du logiciel, mais on observe des signes notables de dispersion de la localisation de plusieurs sources d'information (FAQ, plugins)[221]. En juillet, la documentation est dispersée sur cinq supports[222]. On observe plusieurs mentions de problèmes de désorganisation de la population traitant la documentation, qui se devait de travailler en équipe, et de la documentation elle-même, ce qui n'a cependant pas empêché que le wiki soit considéré comme la documentation officielle. La prise de conscience partagée de la nécessité de la conception d'un guide pour utilisateurs novices a certes été reconnue[223], mais le projet a été mal lancé faute de coordination[224]. Par ailleurs, sont soulignés un manque de pédagogie de l'information pour les utilisateurs novices[225], un manque d'implication des hackers pour la documentation[226] et un manque de leadership[227]. Les références sur la liste de discussions des hackers montrent que le wiki n'est guère accessible aux utilisateurs (« user-friendly »)[228], qu'un autre moteur wiki serait préférable[229], et font aussi mention de problèmes de structuration[230]. Par ailleurs, il s'agit aussi certainement d'une période de désorganisation pour les programmateurs et le forum[231], ce qui peut se comprendre, la communauté étant jeune. Un regard en arrière d'un

[214] Source : http://comox.textdrive.com/pipermail/docs/2004-May/000270.html .
[215] Source : http://comox.textdrive.com/pipermail/docs/2004-May/000257.html .
[216] Source : http://comox.textdrive.com/pipermail/docs/2004-May/000291.html .
[217] Source : http://comox.textdrive.com/pipermail/docs/2004-May/000259.html .
[218] Source : http://comox.textdrive.com/pipermail/docs/2004-May/000276.html .
[219] Source : http://comox.textdrive.com/pipermail/docs/2004-May/000287.html .
[220] Source : http://comox.textdrive.com/pipermail/docs/2004-May/000313.html .
[221] Source : http://comox.textdrive.com/pipermail/docs/2004-May/000292.html .
[222] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000366.html .
[223] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000371.html .
[224] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000375.html .
[225] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000369.html .
[226] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000376.html .
[227] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000377.html .
[228] Source : http://comox.textdrive.com/pipermail/hackers/2004-July/000967.html .
[229] Source : http://comox.textdrive.com/pipermail/hackers/2004-July/000990.html .
[230] Source : http://comox.textdrive.com/pipermail/hackers/2004-June/000365.html .
[231] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000406.html (par exemple).

des intervenants ultérieurs sur la liste mentionne un « chaos »[232]. Ainsi, il ne faut pas s'étonner que la documentation ait subi un premier échec. Bien entendu, des nuances sont observées : on remarque une activité en matière de documentation de plugins[233], une reconnaissance du caractère « amical » de la communauté[234], des performances meilleures que ce qui se fait sur Sourceforge[235], mais ce n'est pas suffisant pour faire contrepoids aux inconvénients précédents.

Par ailleurs, une autre référence émerge au-delà de celles des autres documentations logicielles, avec les premières mentions de Wikipedia comme modèle pour la conception du wiki[236], ce qui sera en fait l'une des raisons du lancement d'un autre système d'information avec un autre outil.

4.3.3 Le lancement du deuxième wiki ouvre une période de croissance forte et d'organisation adhocratique.

En conséquence, un second wiki sur moteur MediaWiki, celui de Wikipedia, a été lancé le 30 juillet 2004[237]. Son nom[238], qui comporte des enjeux en matière de communication, a donné lieu à un intense « brainstorming »[239]. La communauté désirait absolument éviter le désordre observé sur le précédent wiki et a cherché à structurer l'information dès le début[240] ainsi qu'à mettre en place un processus éditorial avec des régulateurs[241]. Cela explique la prise en compte de la nécessité de « manager » la documentation[242] suite aux problèmes de structuration pouvant émerger sur le wiki[243]. Cette régulation assume une double inspiration, provenant à la fois de l'open-source et de Wikipedia[244]. *La centralisation totale et la spécialisation en matière de documentation était*

---

[232] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000512.html .
[233] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000368.html .
[234] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000379.html .
[235] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000386.html .
[236] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000349.html et http://comox.textdrive.com/pipermail/docs/2004-July/000365.html.
[237] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000404.html .
[238] Nous signalons sur ce point que changer « wiki » en « Codex » sur l'Url du site s'avéra une erreur coûteuse d'un point de vue informatique, elle a retardé une mise à jour qui sera évoquée par la suite. On peut nommer cela « lock-in », verrouillage.
[239] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000401.html .
[240] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000411.html .
[241] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000412.html .
[242] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000521.html .
[243] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000534.html .
[244] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000535.html .

imparfaite à cette date : on observe des formes de documentations concurrentes dans d'autres langues que l'anglais[245], et l'ancien wiki est toujours édité, c'est là une politique officielle[246]. On observe aussi de fortes interactions avec le forum[247] et une discussion de ses problèmes, voire une convergence entre activité du wiki et du forum[248] . Il n'y a pas encore à ce moment-là de canal de discussion spécifique pour lui[249] : le forum était à l'époque un élément de la documentation[250], concept manifestement conçu au sens « large ». Cependant, à la marge, on observe un désir de centralisation des tutoriaux sur le wiki[251].

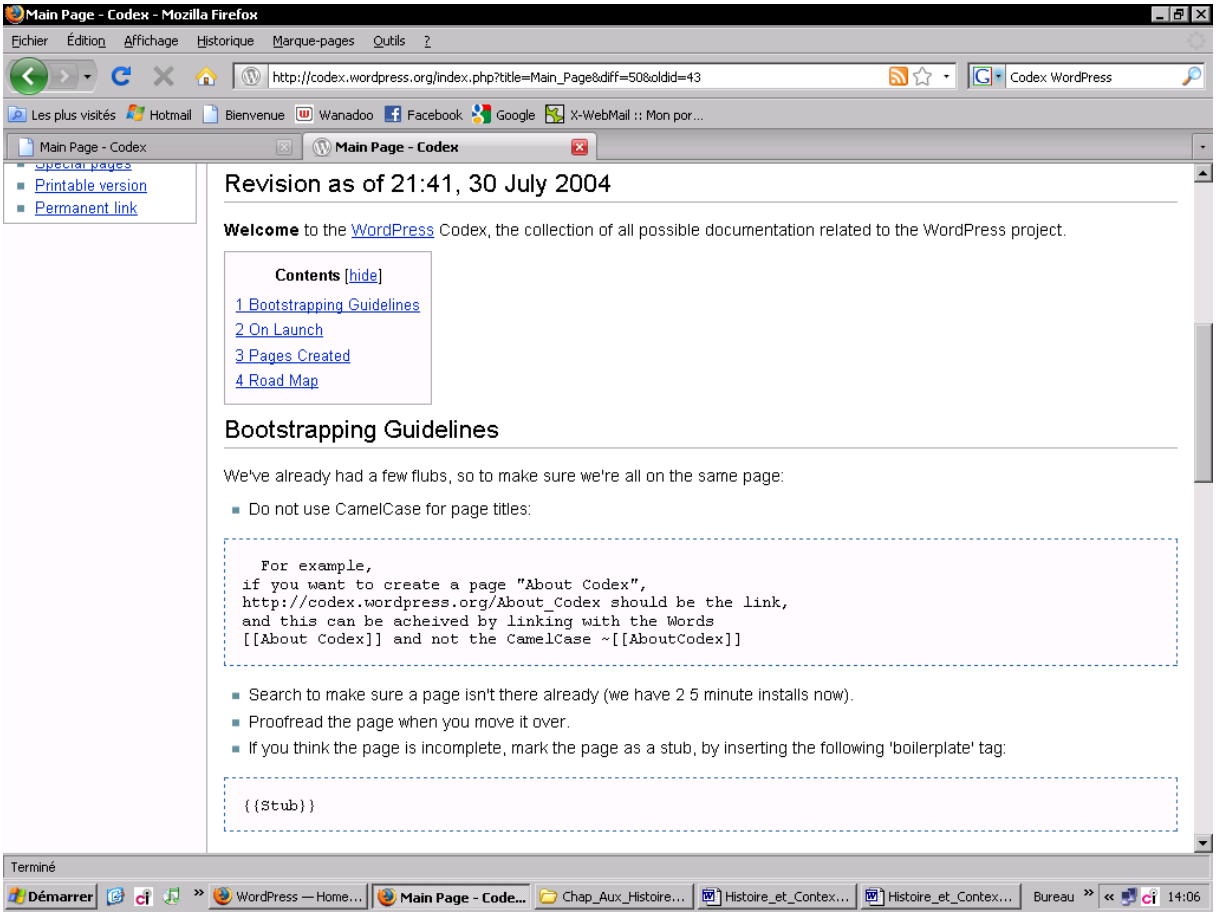


Figure 17: les premières éditions sur le Codex le 30 juillet 2004[252]

[245] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000427.html .
[246] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000457.html .
[247] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000464.html .
[248] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000533.html .
[249] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000482.html .
[250] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000507.html .
[251] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000481.html .
[252] Source : http://codex.wordpress.org/index.php?title=Main_Page&diff=50&oldid=43 .

Le « lancement » du Codex s'est fait un peu plus tard en concomitance avec (et pour) la version 1.3. du logiciel[253]. En matière de contenu, on observe à ce moment là la première évocation de la possibilité d'insertion du code[254] et le lancement de l'idée d'un manuel plus élaboré que le premier sur le premier wiki[255]. L'importance de la documentation pour le « marketing » du logiciel est soulignée par plusieurs intervenants[256], ce qui explique aussi l'insistance sur le management de la documentation[257]. Le premier conflit est observé dans le feu de l'action, entre activité de contribution et de régulation : Craig Hartel (« NuclearMoose »), administrateur et contributeur notable, se vexe fort des demandes d'organisation et des critiques de l'un des principaux développeurs[258], un incident qui finit par des excuses publiques[259]. La politique d'encouragements à contribuer à volonté (« at will ») en effet fonctionne à plein malgré le travail en cours d'organisation[260]. On note en la matière une recherche explicite et plus précise de « bonnes pratiques » dans les autres projets open-source, au-delà de la référence officielle à Wikipedia[261], en particulier lorsqu'elle est faite sur des wikis ; cette légère dérive sur les spécificités de l'outil occasionne un retour sur l'aspect en premier lieu *sémantique* d'une documentation, au-delà du support wiki[262]. On observe des réflexions sur la structuration de contenu, et les premiers problèmes de duplication des tâches[263], d'où des fusions d'articles[264]. Cette phase d'organisation importante serait en lien avec des formes d'apprentissage collectif observées pour faire de la documentation[265].

Lors de l'automne 2004, la constitution de la documentation se poursuit d'une façon énergique. Podz et Carthik reviennent après une période d'absence pour contribuer au Codex[266]. Plusieurs opérations sont observées, la rédaction de la documentation des templates[267], un travail sur les FAQ[268], et la rédaction de guidelines[269]. La mise en relation entre Codex et forum se poursuit[270], mais l'ancien wiki fonctionne toujours à cet effet[271].

---

[253] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000561.html .
[254] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000567.html .
[255] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000645.html .
[256] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000651.html .
[257] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000655.html .
[258] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000662.html .
[259] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000671.html .
[260] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000667.html .
[261] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000661.html .
[262] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000709.html .
[263] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000672.html .
[264] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000685.html .
[265] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000704.html .
[266] Source : http://comox.textdrive.com/pipermail/docs/2004-October/000753.html .
[267] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000717.html .
[268] Source : http://comox.textdrive.com/pipermail/docs/2004-October/000761.html .
[269] Source : http://comox.textdrive.com/pipermail/docs/2004-October/000763.html .
[270] Source : http://comox.textdrive.com/pipermail/docs/2004-October/000736.html .

Novembre est une période de forte implication de Podz[272], qui continue cependant à travailler ses documents propres[273]. La priorité a semblé être donnée à la documentation pour l'utilisateur novice[274], d'où l'extraction de beaucoup d'informations et d'évaluation des besoins de documentation visibles sur le forum[275], mais aussi la documentation des Templates, et les thèmes[276]. On observe aussi que la documentation développeurs de plugins est travaillée[277], ainsi que celle du développement des thèmes[278]

En décembre 2004, on observe un retour des préoccupations d'organisation du wiki. On observe le lancement de l'idée d'un « wiki day »[279], des problèmes d'organisation sur le wiki[280], auxquels répond en particulier l'idée de structurer le contenu sous la forme « task-based »[281]. On note aussi la découverte et la correction de formes d'anarchie dans la mise en forme des titres[282]. On observe beaucoup de discussions des principaux contributeurs sur la structuration du Codex[283], contributeurs qui ont délibérément accepté de prendre beaucoup de temps sur ces questions pour ne pas répéter les erreurs du précédent wiki[284], et ce, aux dépends de la documentation effective. Une tentative d'organiser davantage la « démocratie » de la conception du wiki n'a apparemment pas abouti[285] malgré les risques *supposés* de « dictature »[286]. A ce moment, le wiki semble à présent beaucoup plus abouti[287], d'où sans doute une décision de centralisation des FAQ sur le wiki[288], au point que la communauté va loin en matière de qualité : est évoqué le modèle de présentation d' « Edward Tufte », le grand maître des représentations graphiques, comme norme pour la documentation[289] et on observe

---

[271] Source : http://comox.textdrive.com/pipermail/docs/2004-October/000734.html .
[272] Source : http://comox.textdrive.com/pipermail/docs/2004-November/000821.html .
[273] Source : http://comox.textdrive.com/pipermail/docs/2004-November/000795.html .
[274] Source : http://comox.textdrive.com/pipermail/docs/2004-November/000809.html .
[275] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000923.html .
[276] Source : http://comox.textdrive.com/pipermail/docs/2004-November/000820.html et http://comox.textdrive.com/pipermail/docs/2004-November/000841.html.
[277] Source : http://comox.textdrive.com/pipermail/docs/2004-November/000824.html .
[278] Source : http://comox.textdrive.com/pipermail/docs/2004-November/000838.html .
[279] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000858.html et http://comox.textdrive.com/pipermail/docs/2004-December/000999.html.
[280] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000883.html.
[281] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000926.html .
[282] Source : http://comox.textdrive.com/pipermail/docs/2004-December/001027.html .
[283] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000938.html .
[284] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000947.html .
[285] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000966.html .
[286] Source : http://comox.textdrive.com/pipermail/docs/2004-December/001036.html . C'est hélas un point que les sources ne permettent pas ici de documenter plus avant.
[287] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000976.html .
[288] Source : http://comox.textdrive.com/pipermail/docs/2004-December/001025.html .
[289] Source : http://comox.textdrive.com/pipermail/docs/2004-December/001044.html .

beaucoup de discussions sur la mise en forme de la page principale, ce qui montre l'importance désormais sérieuse du wiki pour le « marketing » de la communauté[290].

On observe la création d'un nouvelle liste le 21 février 2005[291], suite à des problèmes techniques peu clairs. On ne sait quasiment rien de ce qui se passe en janvier 2005 et pendant l'essentiel de février 2005. Les archives de janvier manquent, ce qui est une limite pour comprendre l'histoire de la communauté. Il est difficile donc de dire si on aurait pu observer une rupture pendant cette période.

Les débuts de la nouvelle liste sont surtout marqués par une poursuite du mouvement d'organisation, avec l'insistance[292] et l'acceptation de la règle de composition des brouillons sur les pages des utilisateurs (une routine explicitée dans le chapitre suivant). Malgré des contestations sur leur manque de clarté[293] et d'utilité sur le long terme[294], la routine devient une pratique admise dans les « guidelines »[295]. D'autres points d'organisation sont observés plus à la marge, i.e. des discussions sur la façon de rédiger les articles[296] et le rejet de l'idée de documenter uniquement la version 1.5. sur le wiki[297]. On observe aussi une grosse panne[298] du wiki et une extension excessive de la liste des plugins[299], ce qui aura quelques répercussions immédiates.

4.3.4. Une phase importante de documentation managée par Lorelle Van Fossen, future administratrice, contribue largement à aboutir à une documentation plus exhaustive.

Les premières interventions de Lorelle Van Fossen, une photographe nomade et blogueuse, sont d'abord celles d'une novice qui ne sait pas bien comment contribuer[300], malgré des réécritures antérieures d'articles significatifs[301].

---

[290] Source : http://comox.textdrive.com/pipermail/docs/2004-December/001050.html .
[291] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-February/000000.html.
[292] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-February/000007.html .
[293] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000099.html .
[294] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000118.html.
[295] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000146.html.
[296] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-February/000009.html et http://comox.textdrive.com/pipermail/wp-docs/2005-February/000011.html .
[297] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-February/000041.html .
[298] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-February/000050.html .
[299] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000058.html et http://comox.textdrive.com/pipermail/wp-docs/2005-March/000070.html .
[300] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000133.html .

A ce moment-là, la communauté observait surtout un manque crucial de documentation pour les développeurs[302] : le codex étant alors orienté « end-user »[303]. Ce constat a occasionné un mouvement d'écriture[304] : l'interaction entre questions techniques d'informatique et documentation s'approfondissent[305], menée notamment par un hacker[306] et on organise aussi la documentation des plugins[307]. C'est *en contraste* avec ce mouvement que l'on observe l'émergence des premiers textes de Lorelle sur les Leçons[308], la principale documentation orientée « end-user », qui coïncide avec la décision de séparer la documentation « end-user » et « développeurs »[309]. Lorelle écrit de façon collaborative ses premiers textes mais les contrôle. Le projet aurait été lancé en partie à cause d'une panne du Codex sur une période importante (sans doute celle que nous avions mentionnée précédemment) et les versions « bêta » ont alors été placées sur le forum[310].

Le mois de mars 2005 fut par la suite justement *une période de forte activité* de documentation, menée par Lorelle. La communauté qui demande toujours des contributeurs décide de conserver un modèle ouvert. Au-delà des discussions sur la durée des exclusions des indésirables[311], une longue discussion posait le problème de la conservation des comptes utilisateurs inactifs et a conduit au maintien de ceux-ci, l'ouverture maximale de la communauté étant jugée prioritaire, et cette question démographique de moindre importance par rapport à celle du management de contenu[312]. Les questions des contributeurs et utilisateurs les plus novices sont en effet utiles pour constituer la documentation de base[313]. En matière de contenu, on remarque principalement la préparation de la documentation pour la version 1.5. du logiciel[314], l'organisation du « wiki week-end » de fin mars[315] et le lancement de la partie du Codex en espagnol[316]. Sur le plan technique, suite à des pannes, les

---

[301] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000093.html .
[302] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000167.html et http://comox.textdrive.com/pipermail/wp-docs/2005-March/000161.html .
[303] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000194.html .
[304] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000164.html .
[305] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000201.html et http://comox.textdrive.com/pipermail/wp-docs/2005-March/000183.html .
[306] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000203.html .
[307] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000244.html .
[308] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000098.html et http://comox.textdrive.com/pipermail/wp-docs/2005-March/000198.html .
[309] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000262.html .
[310] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000377.html .
[311] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000236.html .
[312] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000308.html .
[313] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000343.html .
[314] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000351.html .
[315] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000368.html .
[316] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000389.html .

administrateurs prennent conscience de la nécessité de trouver des « sites miroirs » en cas de panne du Codex, une solution permettant la duplication du contenu adoptée ultérieurement[317]. Enfin, la décision de créer une section « Leçons », pour les débutants, dans le Codex est définitivement prise, les articles de Lorelle étant pressentis pour en former une part significative[318]. On observera un succès immédiat des Leçons[319].

Le mois d'avril 2005 est caractérisé à nouveau par une phase d'organisation. La régulation du forum est officiellement centralisée provisoirement sur la liste de documentation[320]. Des débats sur les problèmes de référencement des pages de brouillons[321] amènent à la création d'un outil supplémentaire utile de catégorisation, le template « draft »[322], qui marque par une balise les articles à l'état de brouillon. Le codex commence à être dans un état suffisamment satisfaisant pour être relié à d'autres parties de la communauté[323], après continuation du travail des « Leçons »[324] (une sorte de « wordpress for dummies »[325]), de la documentation de la programmation[326] et des guidelines de la documentation[327]. La documentation apparaît alors cependant comme trop abondante[328], et nécessitant organisation. Dans cette optique, Carthik a créé des pages d'aide à la communauté de contributeurs[329], Lorelle demande à tous de s'inscrire sur une liste sur le Codex, afin de tenter de clarifier les tâches respectives de chacun[330] et on recherche les « trous » dans la documentation[331]. Sur le plan technique, les administrateurs rencontrent un premier échec de mise à jour du wiki[332] et de nombreux problèmes avec le logiciel Mediawiki, par exemple pour déplacer les pages[333].

Le mois de juin 2005 est clairement caractérisé par une période intense de maintenance (dont l'idée avait fait son chemin le mois précédent[334]) à l'exception d'un effort

---

[317] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000398.html .
[318] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000406.html .
[319] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000496.html .
[320] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000428.html .
[321] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000446.html .
[322] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000460.html .
[323] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000464.html .
[324] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000476.html .
[325] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000482.html .
[326] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000463.html .
[327] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000427.html .
[328] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000479.html .
[329] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-May/000516.html .
[330] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-May/000517.html .
[331] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-May/000536.html .
[332] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000490.html .
[333] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-May/000547.html .
[334] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-May/000546.html .

de classification des terminologies[335], du développement des FAQ[336], de modifications significatives de la page d'accueil[337], de l'internationalisation du logiciel qui demande des développements spécifiques de la documentation[338] et d'un projet de « featured articles » pour le forum[339]. Nous observons un lancement d'une intense communication par Lorelle pour effectuer une semaine entière d'écriture d'articles et de « housekeeping »[340]. La « Codex week » fut lancée et se déroula correctement, à savoir que le mouvement d'augmentation ponctuelle brutale des contributions au Codex, aussi bien la correction de la forme que la création de nouvelles pages, fut effectif[341].

A l'inverse, juillet 2005 est un mois de transition aux caractéristiques hétérogènes. Des discussions sur la réputation et l'expertise dans la communauté, une ignorance de Matt qui traduirait un relâchement de son implication[342] et le refus du modèle de structure « narrative » de la communauté Drupal[343] pour le Codex pourraient suggérer un début *d'indépendance* de l'équipe de documentation. Lorelle se sent d'ailleurs désormais responsable au point de prévenir qu'elle sera indisponible à cause de l'arrivée du fameux cyclone « Dennis » (elle vit de façon « nomade » dans une caravane). Des *insuffisances* sont toujours relevées, des problèmes de structure observés, liés à l'aspect touffu (« bush-like ») du wiki, non chronologique[344]. On n'observe pas de projet de « livre » à partir du Codex encore trop incomplet à ce moment[345], mais plutôt le lancement d'une opération pour combler les « trous » dans la documentation[346], qui a du mal à être lancée[347], et beaucoup d'incertitudes mentionnées par l'équipe de documentation, qui envisage que ce projet reste sérieusement possible[348]. Une mise à jour et changement d'hébergement du Codex[349] sont tentés, et des difficultés techniques s'ensuivent, car l'hébergement reste problématique, ce qui est de moins en moins tolérable[350]. On remarque aussi le lancement de la documentation française[351], la

[335] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000567.html .
[336] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000571.html .
[337] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000574.html .
[338] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000580.html .
[339] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000592.html .
[340] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000581.html .
[341] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000626.html et http://comox.textdrive.com/pipermail/wp-docs/2005-June/000647.html .
[342] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000656.html .
[343] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000668.html .
[344] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000670.html .
[345] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000691.html .
[346] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000693.html .
[347] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000709.html .
[348] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000718.html .
[349] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000720.html .
[350] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000730.html .
[351] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000688.html .

décision de rendre obsolète le précédent wiki et la promotion de la wordpress back-up week[352] (« semaine de la sauvegarde »). La création d'un plugin permettant de lister les réponses du Codex en premier (avant le forum) dans les résultats du moteur de recherche, ce qui deviendra la politique officielle de la communauté, montre le caractère central du Codex.

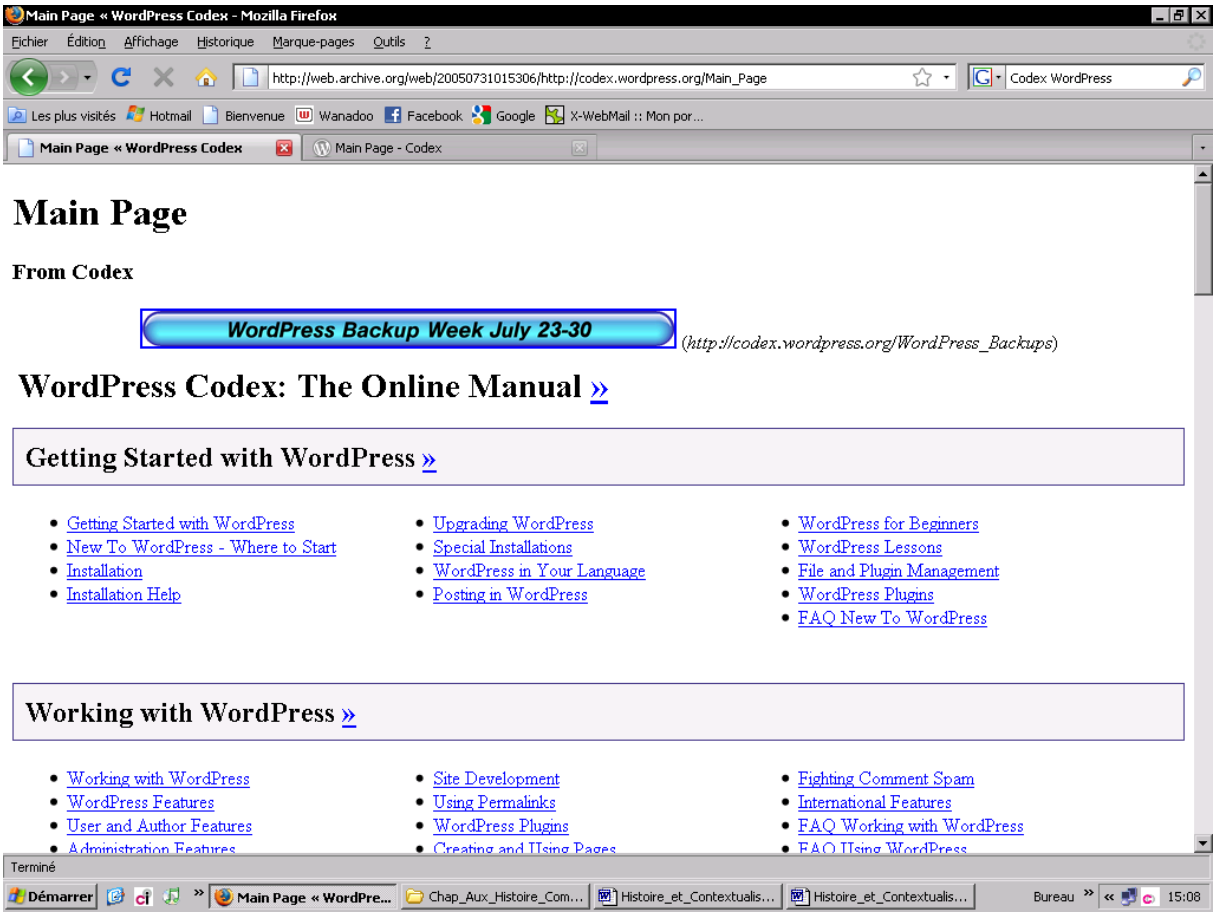


Figure 18: état du Codex après la « WordPress Backup Week », le 30/07/05[353]

Août et septembre 2005 sont des mois caractérisés par beaucoup de discussions juridiques. Le contexte est celui d'une *prise d'importance de cette documentation* : en effet l'officialisation du Codex est planifiée[354]. Un projet de nouveau design pour le Codex est pour cela en cours[355]. L'articulation entre le forum d'aide aux utilisateurs et le Codex est prise en compte et travaillée, et permet d'améliorer de façon significative la situation du premier de ce

[352] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000763.html .
[353] Source : http://web.archive.org/web/20050731015306/http://codex.wordpress.org/Main_Page .
[354] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000784.html .
[355] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/001056.html .

fait[356]. L'intégration forum-Codex se poursuit en septembre au niveau des liens entre les deux SI et la décision d'intégrer le Codex à la communauté se fait plus claire (notamment au niveau de l'esthétique). En ce qui concerne le référencement, qui permet l'accès des utilisateurs, Matt a éliminé les pages autres que celles du « main space » des moteurs de recherche[357]. Cependant, malgré cette prise d'importance, on observe *des éléments de remise en question*. En parallèle, on observe en effet une tentative de Podz de proposer le Codex sous format PDF[358], un projet apprécié y compris par une partie de la communauté, mais qui détourne l'attention du wiki, bien que cette transposition soit entre autres causée par une période significative de fermeture du Codex. Un début de discussion et d'utilisation de l'outil de visualisation des thèmes[359] est observé, alors que c'était le wiki qui répertoriait ces fonctionnalités.

C'est cette initiative de Podz qui fait ressurgir la question du copyright, apparemment non réglée. La question de la protection des droits des contributeurs commence à s'ouvrir. A ce stade la communauté n'envisageait cependant pas explicitement de projet d'ouvrage sur support papier, car entre autres raisons le Codex reste incomplet[360]. Lorelle doit partir à cause de l'arrivé du deuxième cyclone (« Katrina ») et s'absente sur l'essentiel des débats[361], alors qu'elle est la première concernée. Podz décide brusquement de ne plus s'intéresser au Codex à cause de la prise de contrôle apparente du Codex par Lorelle (un « we » d'appropriation selon lui)[362], ce qu'il continuera pourtant à faire, mais qui est révélateur de tensions. Podz demande ensuite des volontaires pour sa documentation sous format PDF et Lorelle organise pourtant le comblement de trous dans le Codex afin de permettre la création des fichiers en question[363]. La problématique de l'appropriation de la documentation (problème économique de « hold up ») est discutée, plusieurs solutions sont proposées, y compris l'appropriation par WordPress Inc.. La communauté prit conscience de la nécessité des avertissements (« disclaimers ») à propos des droits des contributeurs, ce qui montre le sérieux de ce type de préoccupations[364]. Cependant, Podz met brusquement fin à son projet de documents PDF[365], manifestement non viable. De plus, il apparaît qu'il est quasi-impossible de changer de

[356] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000795.html .
[357] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000934.html .
[358] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000814.html .
[359] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000832.html .
[360] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000848.html .
[361] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000877.html .
[362] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000912.html .
[363] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000943.html .
[364] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000891.html .
[365] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000986.html .

licence[366]. De nombreuses discussions ont lieu jusqu'à ce que Matt parle à un juriste, ce qui semble rassurer l'équipe… jusqu'à ce que Michael Hancock découvre qu'une inconnue à la communauté est en train d'écrire un ouvrage à partir du Codex, ce qui provoque encore une décision non suivie d'effet de quitter le projet de la part de Podz qui accusa publiquement à tort Lorelle[367]. Bref, cette période se caractérise sans doute par « beaucoup de bruit pour rien », mais est révélatrice de la prise de conscience par la communauté de la *valeur du bien produit*, de fortes susceptibilités en la matière, et de quelques *tensions* interpersonnelles qui ne sont pas toujours totalement claires. En matière de contenu, une restructuration des FAQ en cours[368], ainsi qu'une intégration des moteurs de recherche interne, qui deviennent communs au forum et au Codex[369].

D'octobre à décembre, on constate au contraire une période qui apparaît moins tendue et plus productive : la communauté coordonne son travail avec la sortie de la version 1.6.[370], ce qui pose des problèmes pour documenter la transition entre les différentes versions du logiciel[371], voire des incidents ailleurs dans la communauté[372]. Le Codex sert à la coordination de la chasse aux bugs (« bug hunt »)[373]. Cependant, on note quelques bémols : l'arrivée de spammeurs[374] entraîne la protection de la page principale, stratégique[375] ; la tentative de Podz de proposer la documentation en format PDF[376] (toujours disponibles alors) est toujours appréciée[377]. En décembre 2005, les attaques de spam s'intensifient davantage, probablement en lien avec le succès de la communauté. La préparation de la documentation pour la version 2.0 du logiciel, importante avec de nombreuses modifications, a été différée en particulier parce que la préparation de la documentation était nécessaire[378]. L'avancement cependant conséquent de la documentation se concrétise par un changement d'apparence du Codex (« blanchissage ») qui en a dérouté plus d'un[379] et qui oblige les contributeurs réguliers à configurer l'apparence en question du wiki (« skin ») de l'outil pour contribuer et à traiter

[366] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/001006.html .
[367] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/001043.html .
[368] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000886.html .
[369] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000895.html .
[370] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-October/001067.html.
[371] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-October/001068.html .
[372] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-October/001078.html .
[373] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-November/001104.html .
[374] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-November/001103.html .
[375] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-October/001088.html .
[376] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-November/001134.html .
[377] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-November/001138.html .
[378] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-December/001162.html .
[379] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-December/001173.html .

quelques problèmes techniques[380]. *Cependant, c'est là une étape importante et un point de rupture, la consécration visible du Codex comme documentation officielle*[381].

Figure 19 : changement d'apparence du Codex avec son statut de documentation officielle[382]

4.3.5. On observe à partir de janvier 2006 remises en question et mise en quasi-sommeil de la régulation de la documentation pendant presque une année

Michael Hancock, qui a besoin de disposer de fonctionnalités d'administrateurs pour faire son travail[383], est promu au vu de ses mérites[384], mais pas un autre[385].

---

[380] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-December/001180.html .
[381] La phase de stabilisation aurait duré un certain temps d'après d'autres sources comme les minutes de chat IRC, où il y a un écart entre les datations : "Yes, the codex is stable" (source : http://codex.wordpress.org/IRC_Meetups/2005/June/June22RawLog) et: "it has stopped growing in the way it used to" (source : http://codex.wordpress.org/IRC_Meetups/2006/January/January04RawLog).
[382] Source : http://web.archive.org/web/20051228203100/http://codex.wordpress.org/Main_Page .
[383] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-January/001194.html .

A partir de janvier 2006, on observe des *fluctuations et remise en question*. Il y aurait eu une période d'inactivité sur le *Codex*[386], et des pannes récurrentes. Des *conflits* apparaissent : les propos malveillants d'un contributeur à l'encontre de Matt Mullenweg a permis de révéler que l'intervention dans la communauté suppose *désormais* des comportements plus professionnels[387] et une régulation sur ce point ; un départ d'un contributeur est observé à cause de la *supposée* prise de contrôle du Codex par Lorelle qui aurait causé une séparation entre forum et Codex[388]. Sur le plan technique, Matt Mullenweg lance l'idée de faire usage du moteur « Blicki » à la place de MediaWiki[389], à cause entre autres des problèmes de mise à jour[390], que la communauté subit de façon significative[391]. « MichaelH » ne réussit pas à résoudre techniquement le problème[392]. Les autres listes, ici utiles, montrent une forte désorganisation portant préjudice aux utilisateurs novices[393], les conséquences de l'instabilité technique du moteur MediaWiki[394], des problèmes de leadership, d'« effet d'éviction » à cause de la montée en puissance des activités marchandes[395], et de management[396]. On le constate donc, les causes d'une « crise » au premier trimestre 2006 sont nombreuses et convergentes. Cela ne signifie pas que le travail s'arrête ; par exemple la documentation de WP-MU et des FAQ s'est poursuivie pendant l'été, mais on ne remarque pas beaucoup d'animation de l'équipe de documentation (ce qui se voit d'ailleurs sur une série chronologique), certainement à cause de l'absence assez mystérieuse de Lorelle van Fossen[397], sur laquelle nous ne pouvons établir sans témérité de motifs définitifs.

---

[384] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-January/001198.html .
[385] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-January/001200.html .
[386] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-January/001205.html .
[387] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-January/001244.html .
[388] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-February/001291.html .
[389] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-February/001297.html .
[390] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-February/001301.html .
[391] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-February/001303.html .
[392] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-March/001310.html .
[393] Source : http://comox.textdrive.com/pipermail/wp-hackers/2006-February/004793.html .
[394] Source : http://comox.textdrive.com/pipermail/wp-hackers/2006-February/005082.html .
[395] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-March/001532.html .
[396] Source : http://comox.textdrive.com/pipermail/wp-hackers/2006-March/005216.html .
[397] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-September/001342.html .

4.3.6. On observe une des reprises périodiques de l'activité de coordination à partir de décembre 2006, caractérisée des évolutions notables mais aussi des formes de permanences

Matt Mullenweg change le Codex de serveur en décembre 2006[398], et le transfère sur celui de l'entreprise Automattic. Ce n'est sans doute pas neutre, mais c'est surtout pour lui l'occasion de tenter de relancer l'activité de la documentation qu'il avait négligée[399]. La communauté prend finalement la décision de ne pas utiliser le moteur Blicki à la place de Mediawiki, après un délai qui est un indicateur significatif au minimum d'une relégation au second plan et au pire d'une mise en sommeil prolongée. Aussi tente-t-elle encore une mise à jour du Codex[400]. Ensuite, certains administrateurs font le constat de l'échec d'une organisation « plate » et demandent davantage d'organisation avec l'idée d'arriver à un livre[401], ce qui est discuté[402], sans qu'une décision définitive soit prise sur ce point controversé. On observe surtout une relance de l'activité sur la documentation pour développeurs par une nouvelle venue, Jennifer Hodgdon[403] en janvier, en particulier en vue de la version 2.1.[404] : cette documentation avait été négligée depuis longtemps[405]. Son travail et son implication en la matière sont reconnus[406]. Les contributeurs à cette documentation ont résolu des problèmes de structure[407], amélioré la documentation des plugins, d'importance stratégique pour la communauté, et décidé de maintenir la documentation des fonctions dans le Codex[408]. On observe aussi une restructuration importante de la page principale du Codex[409]. La prise de conscience collective de la publication prochaine de livres de documentation sur WordPress ne décourage en effet manifestement aucunement la rédaction[410]. Ce mouvement n'est pas sans nuances négatives. On observe deux cas de retours négatifs motivés sur la documentation[411], des causes de désordre, un ajout plus rapide de

[398] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-December/001366.html .
[399] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-December/001370.html et http://comox.textdrive.com/pipermail/wp-docs/2006-December/001391.html .
[400] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-December/001392.html .
[401] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-December/001389.html .
[402] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-December/001402.html .
[403] C'est une chercheuse en physique avec des compétences en programmation qui a finalement créé sa propre entreprise de service en matière de logiciels de création de contenu, http://www.poplarware.com/ . Elle propose en particulier WordPress.
[404] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-January/001409.html .
[405] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-February/001455.html .
[406] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-February/001433.html .
[407] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-February/001436.html .
[408] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-February/001448.html .
[409] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-February/001455.html .
[410] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001485.html .
[411] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001464.html et http://comox.textdrive.com/pipermail/wp-docs/2007-March/001484.html .

contenu que de suivi du management[412], un problème observé plusieurs fois dans la communauté. Jennifer Hodgdon en est venue à assumer l'organisation de la réorganisation de la documentation développeurs[413], réorganisation aidée par Lorelle qui se poursuit en mars 2007[414]. Cependant, les développeurs ne parviennent toujours pas à mettre à jour le moteur MediaWiki[415]. On observe enfin en juillet 2007 des débats sur la suppression des listes de thèmes du Codex, à cause de la volonté de Matt Mullenweg de refuser les « thèmes sponsorisés » ; la volonté de Matt va jusqu'à refuser toute duplication de liste de thèmes, alors que l'équipe de documentation envisageait cette possibilité[416].

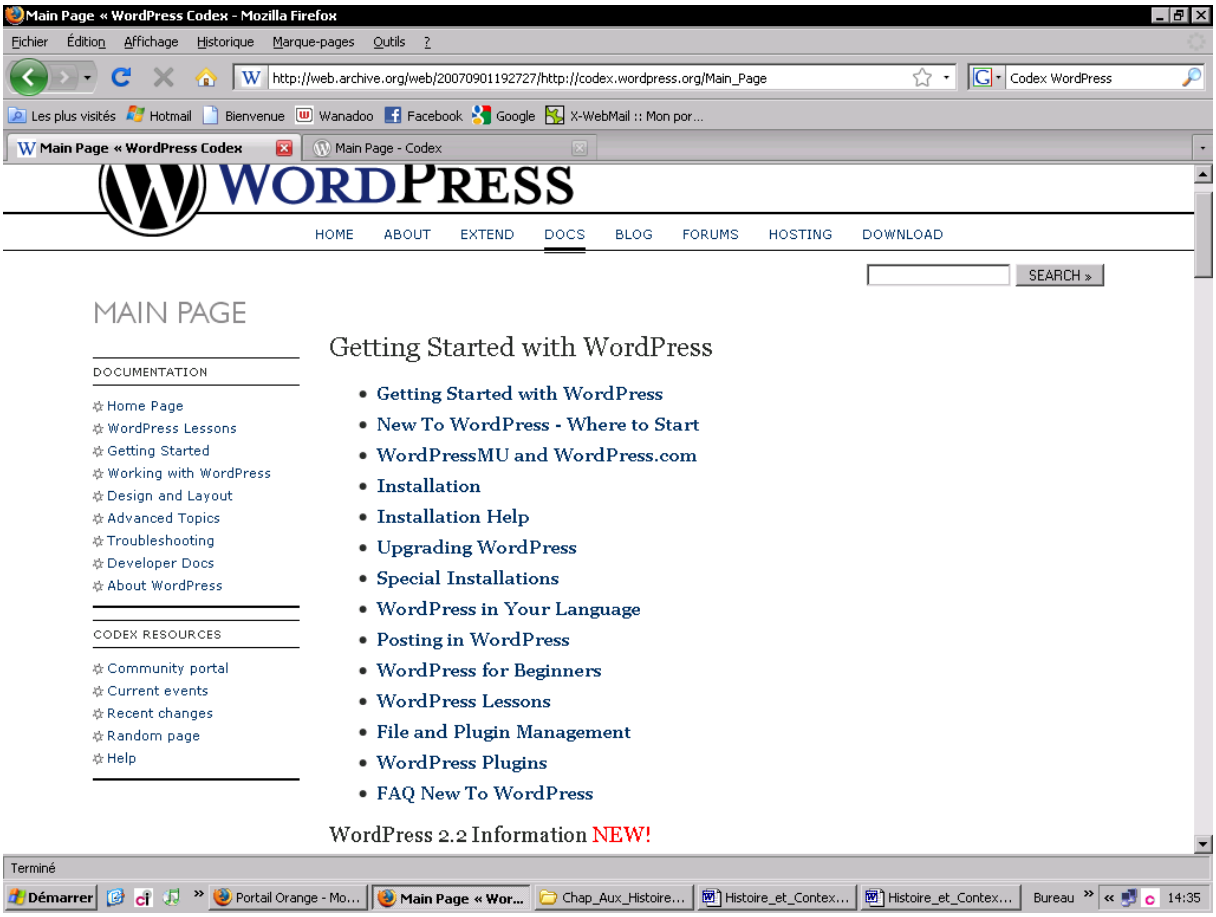


Figure 20 : état du Codex le 01/09/07[417]

[412] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001486.html .
[413] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001488.html .
[414] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001499.html .
[415] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-June/001539.html .
[416] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-July/001551.html..
[417] Source : http://web.archive.org/web/20070901192727/http://codex.wordpress.org/Main_Page .

On observe en septembre, une relance contrastée de l'activité de coordination avec la rédaction de la documentation de la version 2.3. du logiciel[418], après un ralentissement des discussions pendant l'été 2007[419]. *Le moteur MediaWiki a alors été finalement mis à jour*[420]. Si on rapporte tout de même des opinions positives sur la documentation[421], *des insuffisances et des désaccords sont toujours relevés*, avec par la première mention des liens entre parties du Codex anglaises et étrangères qui ne marchent pas[422], un mouvement de regret et de protestation (sans doute latente) à propos de la suppression des plugins du Codex[423], une reprise des discussions sur les inconvénients de la licence GPL pour la documentation[424] sans qu'une décision de modification claire soit prise.

En mars 2008, une phase significative de tâtonnements aboutissant parfois à des réorganisations en parallèle du renforcement de ce qui se faisait avant est observée, avec la modification des « guidelines »[425] et une tentative de réduction de la taille des pages[426]. A partir de mars avril 2008, on remarque à nouveau un sérieux travail supplémentaire en matière de documentation, beaucoup d'échanges sur le contenu, une réorganisation des fonctions PhP de référence[427] et d'autres parties de la communauté, mais aussi une évolution ratée des routines de catégorisation des pages[428]. L'ajout d'un guide supplémentaire[429] a relancé entre autres l'expansion de la documentation pour les utilisateurs novices. Les frontières de l'organisation de la documentation des fonctions PhP a fait aussi l'objet de débats qui n'ont mené à aucune restructuration, mais souligne *la permanence* de l'usage du Codex en cette matière. En décembre 2008, un nouveau cycle de documentation commence avec la sortie de la version 2.7. du logiciel[430], l'insertion d'images[431], et la continuation de l'organisation des

---

[418] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-September/001555.html .
[419] L'explication pourrait se trouver ici : « [17:14] <photomatt_lap> there was one I think, but docs discussion seems to have shifted to wp-forums, strangely
[17:14] <photomatt_lap> also lorelle isn't contributing to codex any more, and michaelH is much quieter
[17:15] <photomatt_lap> anyway, we're tangenting"
(source:http://codex.wordpress.org/IRC_Meetups/2006/June/June07RawLog).
[420] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-December/001566.html .
[421] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-September/001559.html .
[422] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-January/001576.html .
[423] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-January/001594.html .
[424] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-February/001611.html .
[425] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-March/001634.html.
[426] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-March/001637.html .
[427] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-April/001723.html .
[428] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-April/001715.html .
[429] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-April/001733.html .
[430] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-December/001788.html .
[431] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-December/001791.html .

fonctions PhP[432]. *En cette matière, comme pour la programmation, des régularités sont désormais observées.*

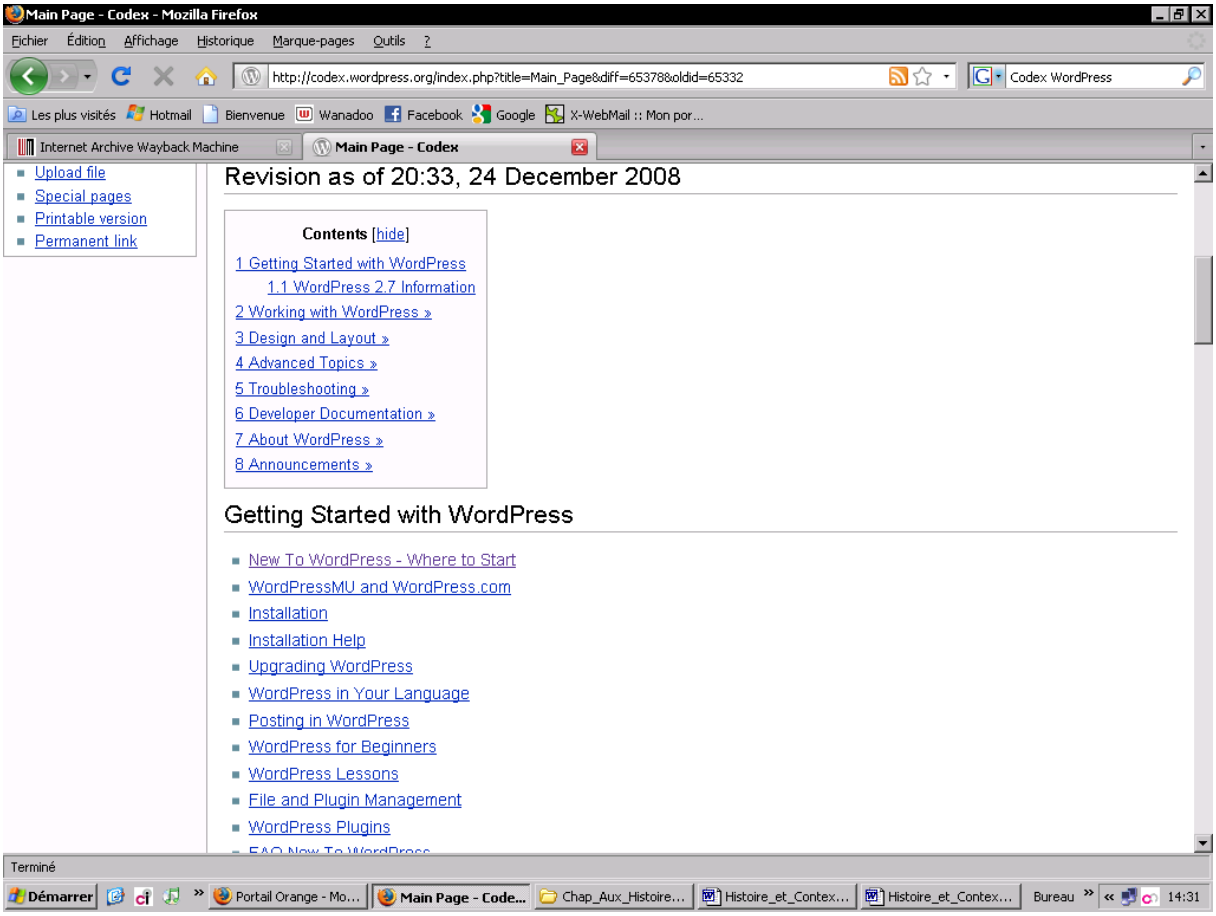


Figure 21 : état final du Codex le 24/12/2008[433]

La période qui commence en janvier 2009, que nous n'évoquerons pas, étant donné que nous devions borner notre recherche de thèse, se caractérise entre autres par l'idée de sortir un manuel extérieur au wiki, qui resterait en ligne avec un niveau « expert », alors que le manuel sur support « subversion » serait destiné aux plus novices[434]. Un tel projet serait *une énorme rupture*, mais cela n'a toujours pas abouti à notre connaissance en août 2009. Cette phase plus incertaine dans cette partie du projet pourrait être une rupture chronologique, qui

[432] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-December/001804.html .
[433] Source : http://codex.wordpress.org/index.php?title=Main_Page&diff=65378&oldid=65332 .
[434] Source : http://comox.textdrive.com/pipermail/wp-docs/2009-January/001862.html .

justifie aussi notre idée de nous arrêter au 1er janvier 2009. De plus, en 2009, trois nouveaux administrateurs sont apparus[435], le dernier le 21 juillet 2009.

4.3.7. Propositions de synthèse par des repères chronologiques

- Avril 2003 : premières traces numériques de la communauté.
- Août 2003 : premières traces numériques de documentations.
- 13 novembre 2003 : lancement des échanges sur la documentation par liste de discussions.
- 18 décembre 2003 : lancement du premier wiki.
- Janvier-avril 2004 : phase de contribution sur le wiki.
- Mai-juin 2004 : phase de croissance forte et désorganisation.
- 30 juillet 2004 : lancement du Codex sur moteur MediaWiki.
- Août 2004 - décembre 2004 : phase de croissance et d'organisation intense.
- Janvier-février 2005 : histoire mal connue et panne de la liste de discussions.
- 21 février 2005 : lancement d'une seconde liste de discussion, « wp-docs ».
- Mars - juillet 2005 : phase intense de documentation et forte implication de Lorelle Van Fossen.
- Août – septembre 2005 : phase de discussions juridiques.
- 27 décembre 2005 : officialisation du Codex comme documentation officielle et intégration esthétique au site.
- Janvier-mars 2006 : diminution de l'activité et constat de désorganisation.
- Avril-novembre 2006 : ralentissement notable de la régulation de la documentation.
- Décembre 2006-mars 2007 : transfert du Codex sur le serveur de l'entreprise Automattic et relance de la documentation à destination des développeurs.
- Avril – août 2007 : ralentissement notable de l'activité de documentation.
- Septembre 2007 : mise à jour de MediaWiki et reprise de l'activité de documentation avec la sortie de la version 2.3. .
- Mai – novembre 2008 : ralentissement notable de l'activité de documentation.
- Décembre 2008 : reprise de l'activité de documentation avec la sortie de la version 2.7. .

[435] Nous avons des doutes sur le sérieux de deux d'entre eux.

4.4. Evolution des interventions des administrateurs :

Nous venons de présenter une histoire linéaire du projet, qui mentionnait l'intervention à plusieurs reprises des administrateurs, qui peuvent avoir des rôles plus ou moins conséquents. Nous approfondissons ici ce point.

En effet, le statut d'administrateur donne au minimum des possibilités d'interventions techniques supplémentaires à celui qui en dispose : il s'agit pour l'essentiel de bloquer sur plusieurs points les interventions des utilisateurs au besoin, protéger des pages, intervenir sur les historiques d'éditions, contrôler les éditions des utilisateurs (« patrouiller »), importer du contenu, changer l'emplacement des pages[436]. Ces possibilités ne disent rien en revanche du comportement effectif des administrateurs, en particulier en matière de management du contenu et des utilisateurs.

Une fonction spéciale apparue sur Mediawiki[437] permet de dater la promotion comme « administrateur » de *certains* administrateurs. Certains ne sont pas répertoriés par cette fonction, d'autres preuves de la datation proposée sont alors insérées *en plus* en note. On obtient alors ce qui suit :

- Pour les premiers « admins », Podz et Matt, il est impossible de dater la promotion[438]. Matt a cependant dû recevoir le statut d'administrateur dès le 30 juillet 2004, étant donné que c'est lui qui a effectué l'installation.
- 27 Décembre 2004, NuclearMoose, et 31 December 2004, Carthik.
- 7 mars 2005, Kafkaesqui[439] et MDAWaffe[440]
- Juin 2005 (*estimation*), Lorelle Van Fossen[441].
- 28 août 2005, Skippy et Westi.
- 4 janvier 2006 : MichaelH.

---

[436] Le détail des fonctionnalités se trouve ici : http://codex.wordpress.org/Special:ListGroupRights (consulté le 02/08/09).

[437] Source : http://codex.wordpress.org/index.php?title=Special%3ALog&type=rights&user=&page=&year=&month=-1

[438] Pour ce qui est de Podz, on sait juste qu'il l'était avant le 12 mai 2005 (source : http://codex.wordpress.org/index.php?title=Documentation_team&diff=prev&oldid=11991) et après août 2004, ce qui n'est guère précis.

[439] Source : http://codex.wordpress.org/User_talk:Kafkaesqui#Welcome_SysOp .

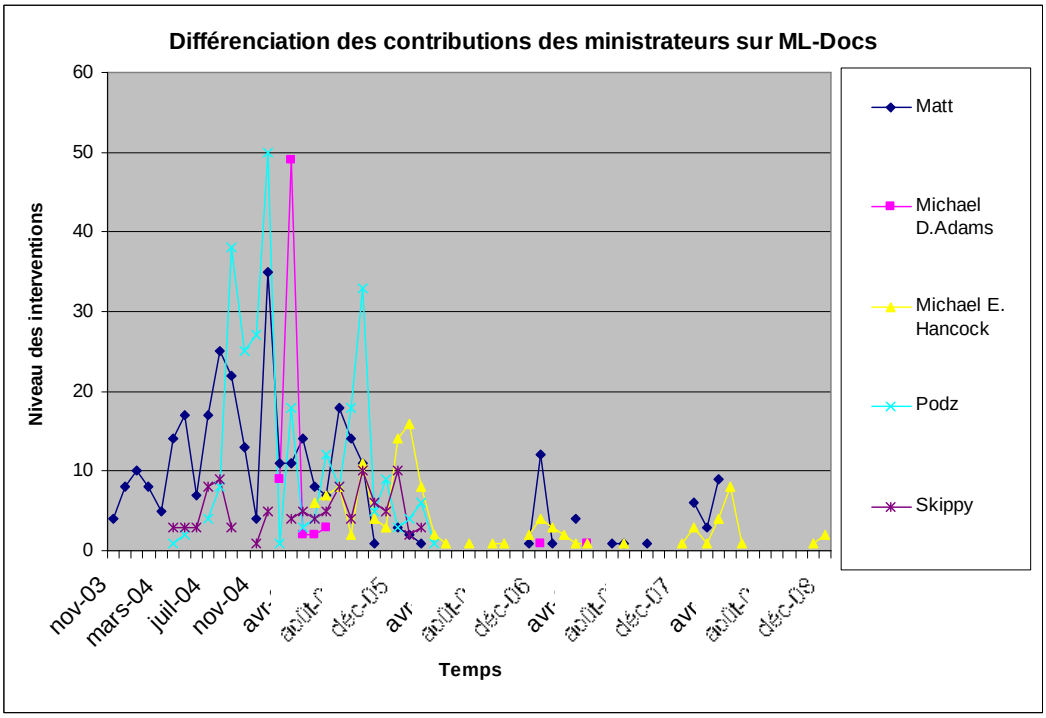
[440] Source : http://codex.wordpress.org/User_talk:MDAWaffe#Welcome_SysOp .

[441] En revanche, curieusement, Lorelle n'est pas répertoriée, et rien dans les archives ne permet de dater son arrivée avec grande précision. C'est là un mystère. Nous avons pu effectuer une datation possible à partir de l'apparition de Lorelle sur la liste des administrateurs, le 18 juin 2005 (source : http://codex.wordpress.org/index.php?title=Documentation_team&diff=next&oldid=16599) et de sa première intervention sur la liste de discussion, 4 mars 2005 (source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000092.html).

- 5 mars 2007 : Foolswisdom.

On observe donc plusieurs vagues de promotion, avec trois regroupements possibles.
Les critères de sélection ne sont pas clairs d'après les archives étudiées, à part le nombre de contributions, pour certains (MichaelH). La méritocratie fonctionne, mais pas au même degré pour tous et en fonction des circonstances. Nous verrons en effet au chapitre suivant que le nombre de leurs interventions sur le wiki lui-même est hétérogène, de même que leurs interventions sur la liste de discussion de la documentation.

Nous présentons à présent leurs interventions sur la liste de discussion, qui permet de comprendre la diversité de leurs interventions. Un seul graphique était difficilement lisible, aussi nous avons divisé *arbitrairement* la population en deux populations, pour meilleure présentation *seulement* :



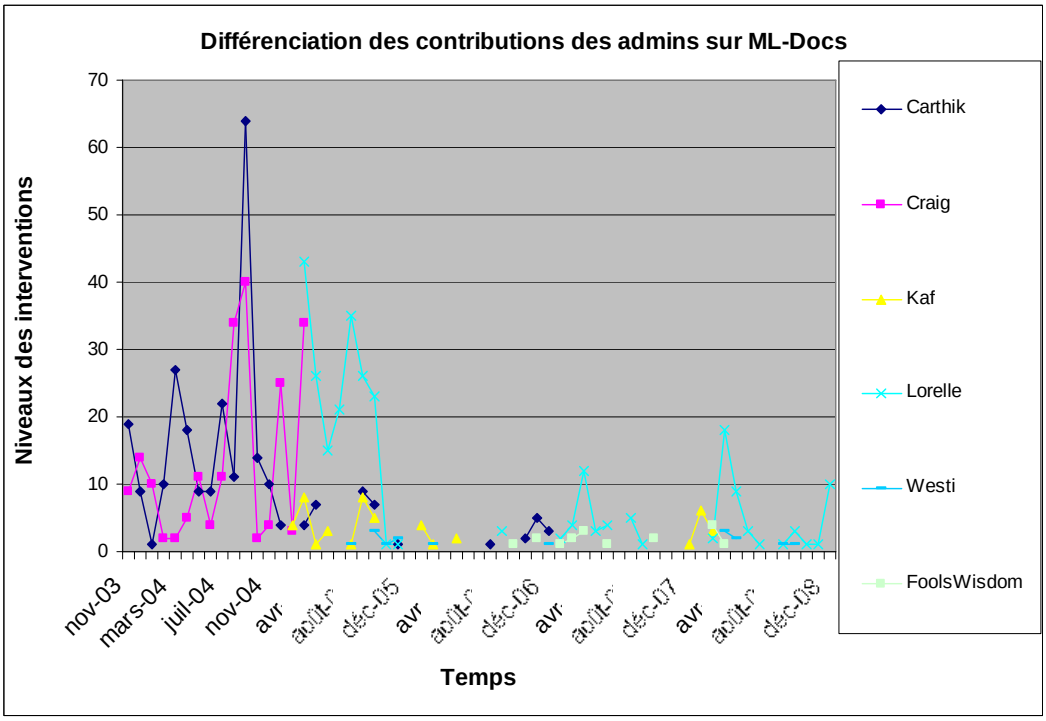


Figure 22 : évolution des interventions des administrateurs sur la liste «wp-docs »

Ce qu'il faut retenir de la structuration des interventions des administrateurs est l'hétérogénéité : certains se caractérisent par des pics de discussions importants (Carthik par exemple), d'autres ne contribuent que de façon mineure et épisodique après une intervention massive (MDAWaffe). Un *renouvellement* (« turnover ») s'est aussi opéré, certains gros contributeurs du début disparaissant (Podz, qui se concentre sur le forum et la start-up et qui doutait du Codex, Skippy, qui a quitté le projet suite à des différents. D'autres sont arrivés plus tard (MichaelH, Foolswidom).

Il semble que les contributeurs mineurs (Kaf, Westi) soient plus préoccupés par d'autres parties du projet (Westi est un hacker), bien que présents sur la totalité du projet de documentation. Le rôle de leadership de certains est clair (MichaelH, Lorelle).

Matt, le chef de l'ensemble du projet WordPress, a un statut particulier. Il est là depuis le début, s'est éclipsé début 2006, mais a repris la documentation en main fin 2006 début 2007.

On observe donc des périodes différenciées de présence des administrateurs de la liste de discussions ; mais ce n'est pas parce qu'ils ne parlent pas qu'ils ne reçoivent pas les mails, et ne contribuent pas ailleurs.

4.5. Rythmes des activités et mise en évidence de dynamiques à partir des séries chronologiques

Nous avons centré nos propos précédemment sur l'histoire de la création de la documentation et sur une partie significative de ceux qui y ont aidé, les administrateurs. Cependant, la documentation n'est pas isolée du reste de la communauté WordPress, elle s'y insère. Etudier le reste de la communauté dans une certaine mesure permet de comprendre le contexte de la création de cette documentation. Nous partirons de l'activité d'ingénierie et nous étudierons la relation entre utilisateurs et la communauté de volontaires, pour arriver aux dynamiques propres à la documentation. Ces recherches permettront des comparaisons et rapprochements entre les dynamiques observées.

Techniquement, il s'agit pour cela d'aborder l'histoire de la communauté à partir des données quantitatives disponibles, des séries chronologiques. Leur présentation suppose surtout des formes de dépouillements conséquentes des archives présentées.

4.5.1. L'activité de programmation

Les dynamiques de l'activité de programmation sont essentielles pour comprendre l'activité de la communauté, même si notre sujet ne porte pas directement là dessus: en effet, il s'agit de celles qui sont propres à l'élaboration du produit WordPress, dans une communauté formée essentiellement d'informaticiens[442]. L'histoire de sa relation avec le wiki sera elle abordée en son temps dans l'avant-dernier chapitre.

4.5.1.1. Le rythme de publication des nouvelles versions du logiciel

Le dépouillement des données des archives des listes de discussions et la quantification permet de dégager de grandes tendances de l'histoire de la communauté WordPress. Cependant, le dépouillement d'autres sources pertinentes, en particulier le blog de suivi de l'activité des développeurs[443] permet d'établir une chronologie des nouvelles versions

[442] L'histoire de l'activité d'ingénierie informatique mériterait dans une autre optique de plus amples développements si on l'étudiait pour elle-même, mais ce n'est pas notre sujet.
[443] Source : http://wordpress.org/development/ ou http://codex.wordpress.org/Changelog .

du logiciel[444], ce qui est un rythme important que le lecteur doit rapidement connaître au préalable. Ce travail permet de distinguer trois phases : la première en 2003, jusqu'à juillet 2003, avec les premières versions, une deuxième plus calme jusqu'à novembre 2005[445], et enfin une phase où les nouvelles versions sortent de façon plus régulière, ce qui correspondrait à la politique officielle de la communauté[446].

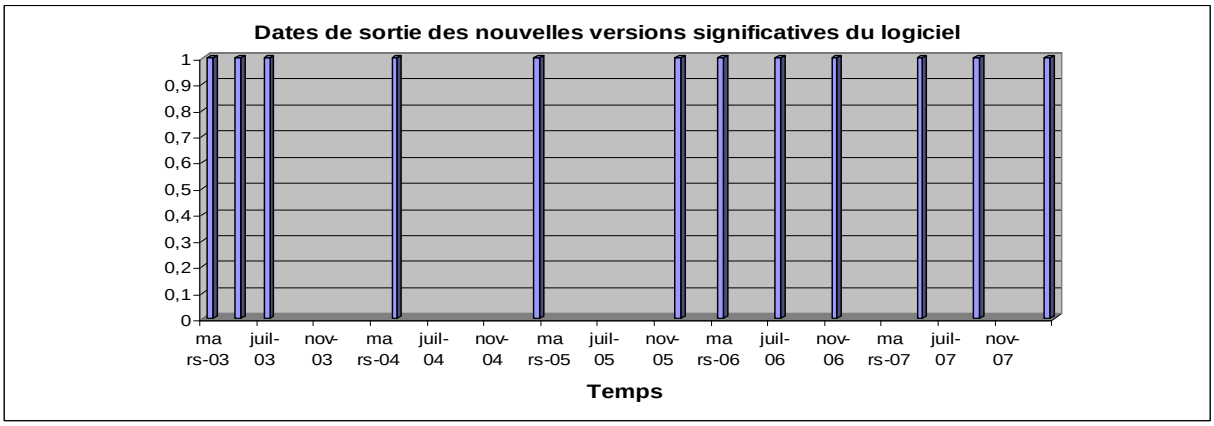


Figure 23 : datation des sorties des nouvelles versions significatives du logiciel

4.5.1.2. Les activités de programmation

Le dépouillement intégral des archives des listes « hackers » (« wp-hackers » est fusionnée avec celle-ci), « trac », « CVS » (« SVN » est fusionnée avec celle-ci) et « testers »[447] sur cinq ans permet de construire les séries chronologiques qui suivent. Nous présentons d'abord chacune d'elle avant de les agréger, ce qui servira enfin à commenter l'ensemble :

[444] Chaque nouvelle version significative du logiciel porte le nom d'un grand musicien de Jazz (Matt Mullenweg est grand amateur), ce qui souligne aussi encore une fois l'importance de *l'esthétique* dans cette communauté. Nous verrons davantage cela dans le chapitre ultérieur sur les routines.

[445] On observe une prise de conscience de la volonté de ralentir le développement logiciel, par exemple : http://comox.textdrive.com/pipermail/wp-docs/2005-February/000043.html .

[446] Nous avons choisi de ne pas être précis au point de mentionner les publications des nouvelles versions « mineures ».

[447] Nous renvoyons au chapitre de méthodologie pour une présentation détaillée de ces sources.

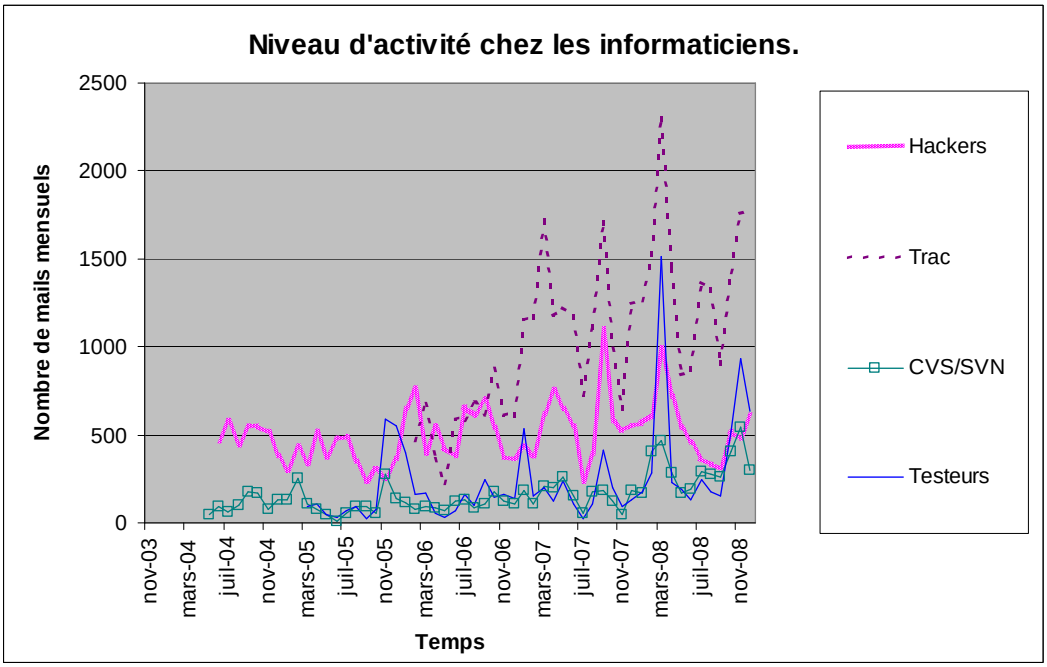


Figure 24 : séries chronologiques de l'activité chez les informaticiens.

Hackers	27703
Trac	37159
CVS/SVN	9078
Testeurs	10990

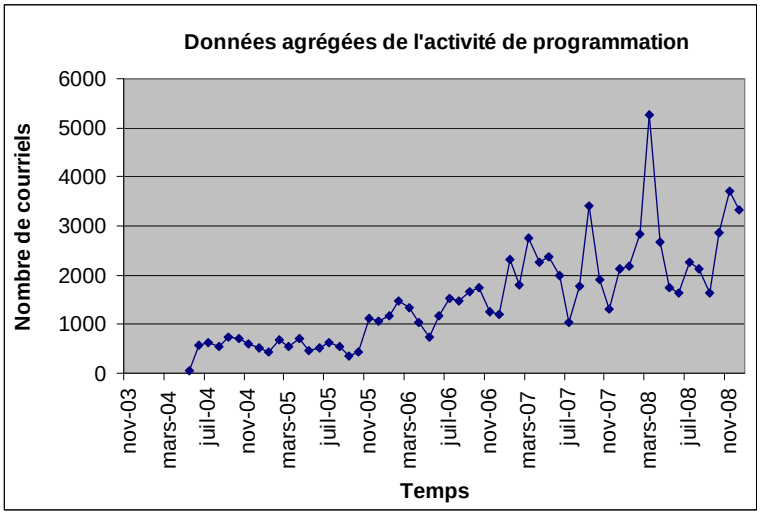


Figure 25 : agrégation des listes principales d'informaticiens et addition des séries chronologiques des informaticiens

L'activité de programmation se caractérise donc, sur la série agrégée précédente, par une période de stabilité (autour de 600 courriels par mois) jusqu'à novembre 2005, et par l'émergence d'une activité croissante mais cyclique à partir de décembre 2006. La deuxième période montre un accroissement notable de l'activité de programmation, avec un « trend » *croissant* net, mais aussi des *variations* d'une amplitude de plus en plus forte (augmentation de la variance). Sur le premier graphique avec les quatre séries chronologiques, on constate

que les années 2004, 2005 et 2006 ont consisté en une activité non coordonnée des différentes activités informatiques. Les années 2007 et 2008 se caractérisent au contraire par l'émergence de cycles: les pics se *resserrent* et sont plus marqués sur les données agrégées (on voit mieux en « reculant » comme dans le film *Blow-up* d'Antonioni) où les quatre listes de discussions deviennent de plus en plus fortement corrélées. Le détail des listes montre que ces variations fortes sont dues à des corrélations partielles entre l'activité des listes (les « crêtes de montagne » se trouvent alignées).

Il est en revanche plus difficile d'évaluer la logique des autres projets informatiques secondaires liés au projet, qui sont plus décentralisés. Mais nous constaterons que les examiner permet de confirmer les évolutions observées précédemment. Nous travaillons avec ces données de la même façon que pour les séries chronologiques précédentes. Voici ce que donne le dépouillement des archives de coordination des autres projets d'ingénierie informatique :

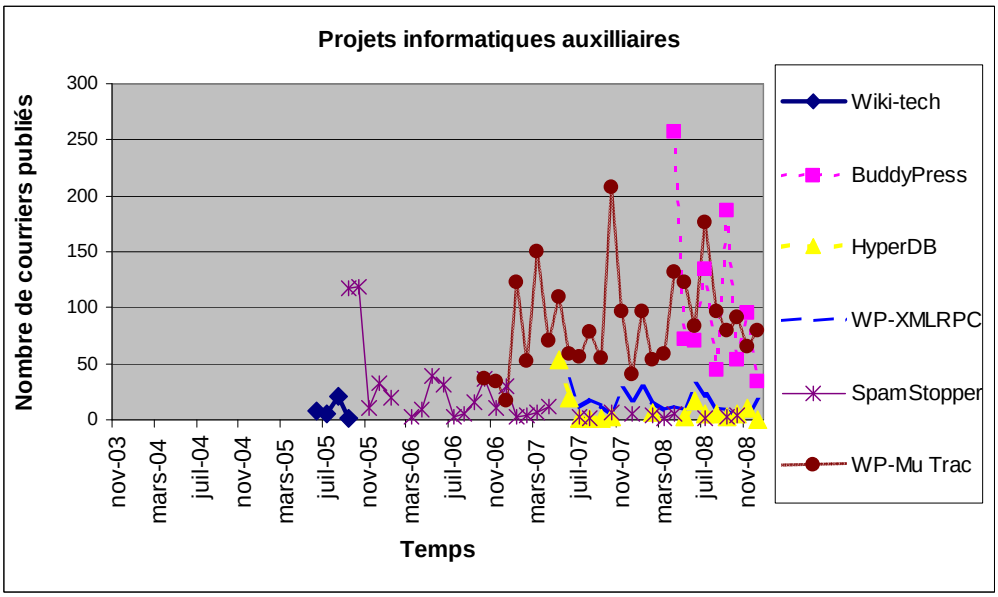


Figure 26 : séries chronologiques des listes de discussions des projets informatiques secondaires

Les projets sont d'importance et de durée de vie diverses, et leur logique de synchronisation *ne ressemble pas du tout* à ce qui a été observé précédemment (pas de cycles). En revanche, la sommation des activités (graphique ci-dessous) est utile de la même façon que précédemment : il permet de décrire *à nouveau deux phases* qui apparaissent mal sur le premier graphique : peu ou pas d'activité avant fin 2006, et une deuxième phase de en 2007 et

2008. Cette logique ressemble sur ce point à celle observée précédemment, et qui confirme que les années 2007 et 2008 ont vu une *importante activité d'ingénierie informatique*. C'est ce que nous retiendrons.

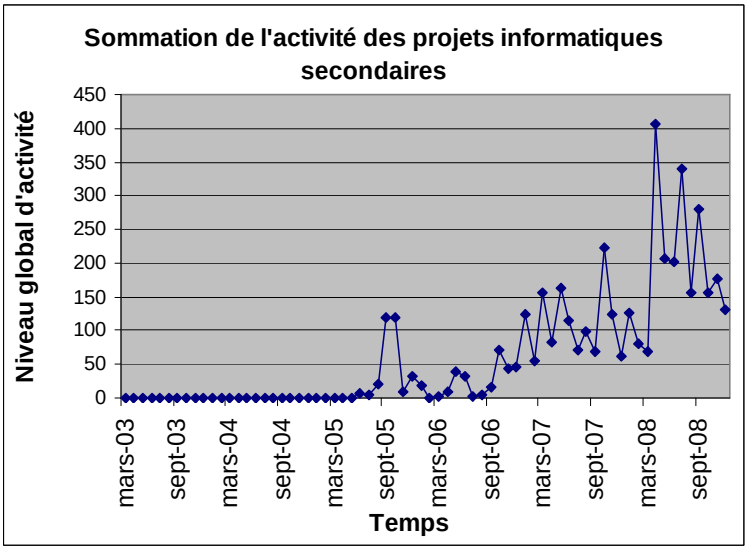


Figure 27 : série chronologique de la sommation de l'activité des projets informatiques secondaires

4.5.2. La fréquentation de la communauté et le développement en conséquence des activités de support

Nous venons d'examiner les grandes tendances de l'activité d'ingénierie informatique ; nous allons aborder un autre aspect de cette communauté. Voici à présent une tentative de compréhension de l'étendue du succès de WordPress auprès de ses utilisateurs et de ce qui s'ensuit, la relation entre utilisateurs et la communauté de volontaires qui cherchent à les aider, communauté dont la composition va au-delà de la population d'informaticiens qui contribue à la programmation.

4.5.2.1. La fréquentation de la communauté :

L'outil Alexa[448] permet d'estimer le succès de la communauté, au-delà de sa renommée. Le premier graphique est le plus important :

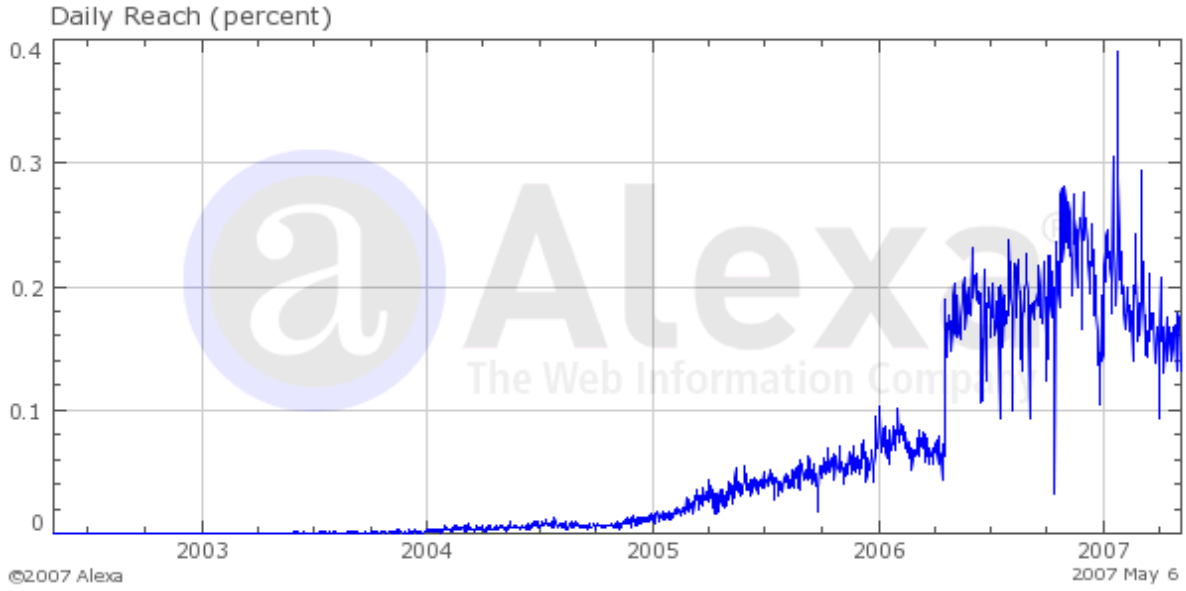


Figure 28 : graphique de fréquentation de la communauté, par Alexa de 2003 à 2007

Lorsqu'il a été obtenu (en mai 2007), WordPress était l'un des sites les plus visités au monde, parmi les 500 premiers. Par ailleurs, il montre d'une part le caractère décisif de *l'année 2005*, qui apparaissait dans la liste de discussion et qui trouve confirmation ici, d'autre part la présence d'une *rupture* au premier semestre 2006. Cette information est importante car c'est peut être une hypothèse explicative supplémentaire pertinente à la crise observée début 2006 dans la liste de la documentation (et évoquée dans l'histoire événementielle), i.e. une fréquentation supplémentaire que la communauté n'aurait peut-être pas su gérer.

Un deuxième graphique montre la permanence de cette importante fréquentation, malgré une baisse relative:

[448] http://www.alexa.com/ permet d'obtenir gratuitement ou contre rémunération des informations de qualité diverse sur la fréquentation des sites.

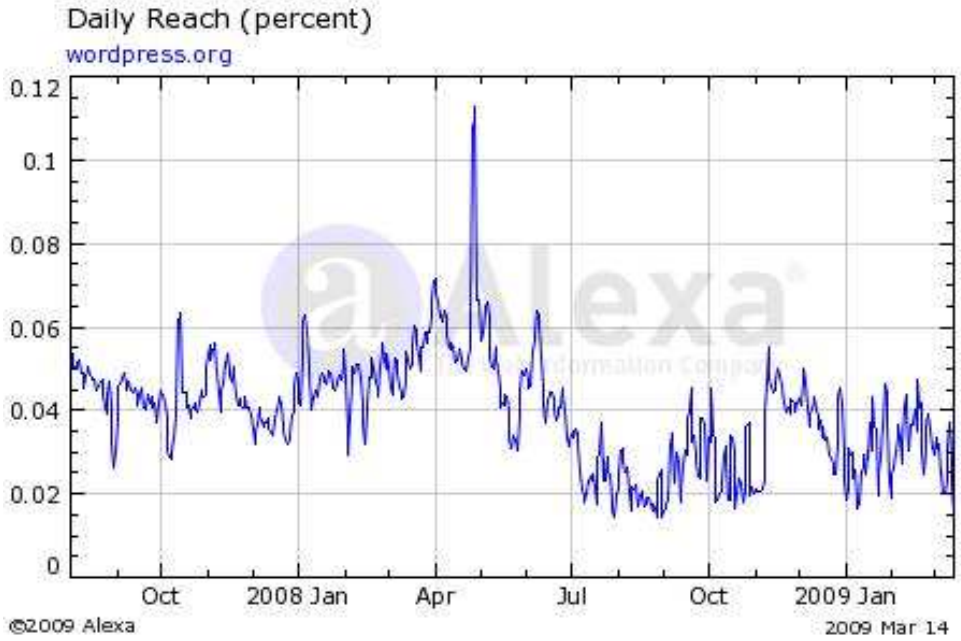


Figure 29 : graphique de fréquentation de wordpress.org, par Alexa, période de août 2007 à février 2009

4.5.2.2. Le forum de discussions et l'activité marchande

L'intérêt des statistiques du forum de discussions est de montrer *l'importance* de l'activité de support aux utilisateurs, ce qui est *aussi* le rôle du SI sur support wiki. Il s'agit également d'un indicateur du succès de la communauté, au-delà de sa réputation « médiatique », et enfin aussi un indicateur de *la charge de travail* qu'elle doit supporter, ce qui n'est pas anodin pour une communauté de volontaires.

Elles ont pu être déduites de l'archivage automatique de captures d'écran de cette partie de la communauté par archives.org[449]. L'échelle logarithmique[450] permet de visualiser ci-dessous une rupture (relative) dans le rythme de croissance exponentielle que ne permettait pas l'échelle normale. Le point de rupture au niveau de l'activité du forum, entre mars et septembre 2004, n'est pas le même que dans les autres parties de la communauté.

[449] Source : http://web.archive.org/web/*/http://wordpress.org/support/ . Nous n'avons pas tout dépouillé, notamment à partir de 2006, car cela ne présentait pas grand intérêt, le « trend » apparaissant clairement, excepté dans le cas d'une analyse fine qui ici n'est pas notre objet d'étude mais qui est bien sûr envisageable. Nous avons ici l'un des apports essentiels de web.archives, la datation de l'activité du forum « sur le forum » étant insuffisamment précise (post écrit « deux ans auparavant ») par comparaison; nous ne parlons même pas des difficultés du traitement direct des données.

[450] Nous n'avons pas exposé les données brutes, qui présentaient des courbes fortement croissantes.

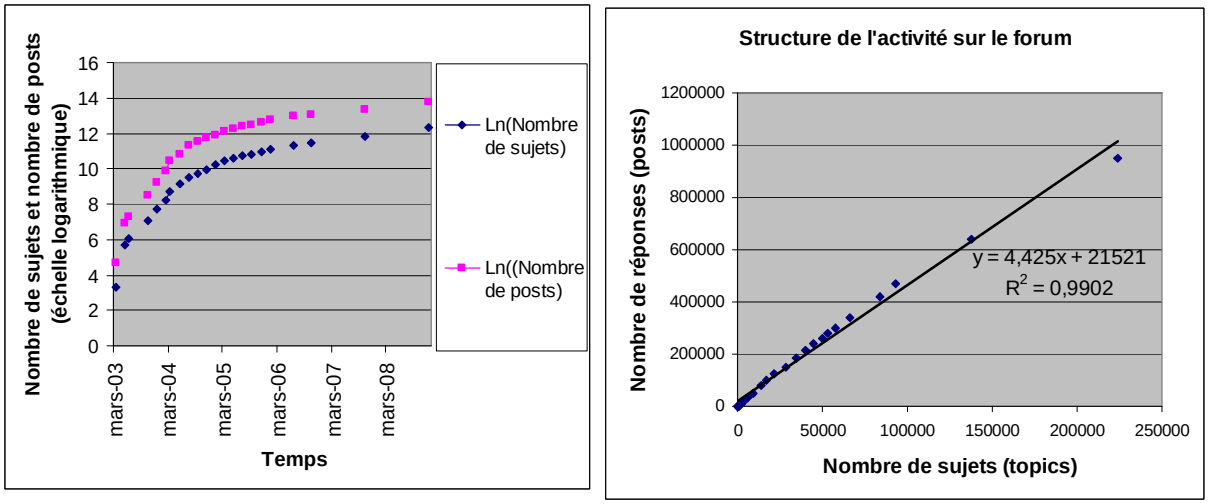


Figure 30 : série chronologique du nombre de sujets postés et du nombre de réponses « posts » (échelle logarithmique)

Le graphique de droite (avec des données brutes) ci-dessus montre que, malgré tout, la structuration des fils du forum obéit à une régularité nette au niveau agrégé. On la devinait sur le graphique de gauche mais elle méritait une explicitation plus poussée : pour un sujet posté, 4,5 réponses sont données en moyenne. Ce chiffre moyen et agrégé est difficile à interpréter ; cependant, il montre au moins la constance de la *présence* d'une communauté de répondants volontaires, ce qui est un résultat significatif. *L'efficience* réelle du support utilisateur est en revanche une toute autre question, qui est voisine de notre sujet, mais à laquelle le graphique de gauche ci-dessous, obtenu à partir des données disponibles à l'état de traces sur web archives, ne permet pas de donner de réponse vraiment satisfaisante :

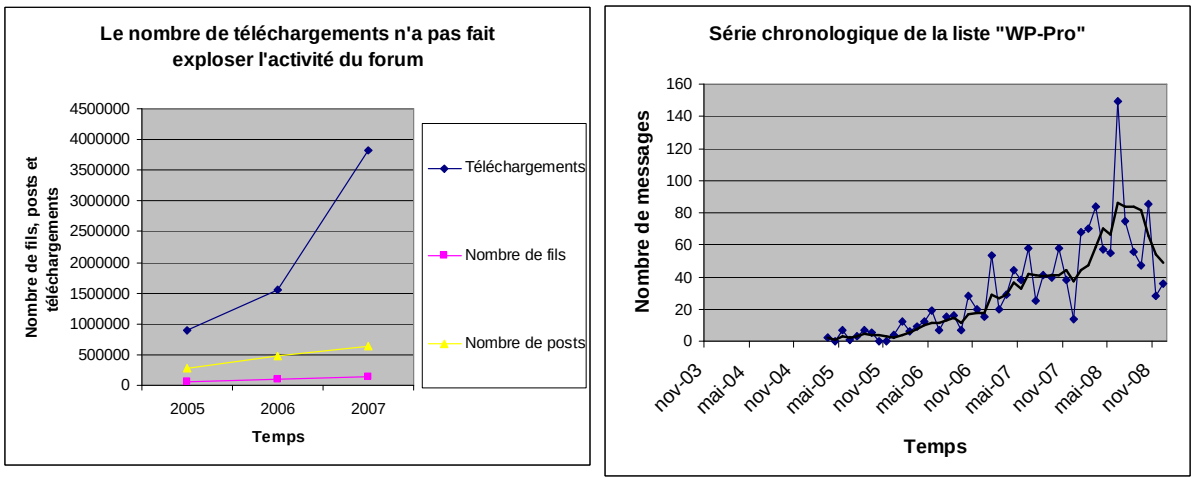


Figure 31 : évolution des téléchargements et de la liste de discussion « WP-Pro »

On constate sur ce premier graphique que le nombre de téléchargements augmente beaucoup plus vite que l'activité du forum. Plusieurs hypothèses explicatives peuvent être retenues pour comprendre cela dans un contexte de fréquentation forte de la communauté : la documentation est suffisamment efficace pour remplacer le forum et combler l'écart entre les courbes, le logiciel est plus pratique et soulève moins de questions, l'information s'est suffisamment accumulée dans le forum et les utilisateurs posent moins de questions, l'activité marchande remplace le travail communautaire, ou bien le forum n'est pas efficace de réputation, et serait négligé… mais on ne peut retenir sérieusement celle selon laquelle le forum n'arriverait pas à « suivre », le ratio post/réponses précédent restant en effet *constant et donc l'activité reste soutenue.* Même si l'efficience de la communauté en matière d'aide aux utilisateurs (forum, documentation) est vraisemblable, *nous ne pouvons conclure définitivement sur cette question,* à partir des données de web.archives, c'est là l'une de leurs limites : si la fréquentation de la communauté augmente, si le nombre de téléchargement augmente, la nature autres données disponibles mise en relation avec ces indicateurs de succès du logiciel ne permettent pas des conclusions définitives, étant donné qu'il n'est pas possible de départager les hypothèses explicatives que nous venons de mentionner. En ce qui concerne l'efficience de la documentation, des travaux supplémentaires sont donc nécessaires, et nous aborderons la question au chapitre 8.

Le graphique à droite ci-dessus vient compléter le premier. Il montre l'activité de la liste « wp-pro », qui est le lieu de discussion et de proposition d'emploi salarié en lien avec le logiciel wordpress. C'est également un indicateur de succès pertinent : il rappelle la croissance de cet aspect de la communauté, mais il en donne une vision incomplète, étant donné que l'activité de la start-up wordpress.com n'est pas rapportée dans cette liste; nous n'avons pas accès à ces données et cette question ne se rapporte pas à notre sujet.

Les indicateurs ci-dessus sont des indices forts de succès de la communauté, mais ils ne permettent pas de conclure définitivement à propos de son efficience en matière de recherche d'information, i.e. de satisfaction des utilisateurs.

4.5.3. L'activité sur le wiki

Après avoir examiné le reste de la communauté, nous étudions les dynamiques du wiki. Les données disponibles permettent de dégager au moins trois rythmes, correspondant à

trois activités aux rythmes distincts, d'un point de vue purement quantitatifs : la création d'articles, la gouvernance (la liste de discussions), et les éditions.

4.5.3.1. La liste de discussions de coordination de l'activité de documentation

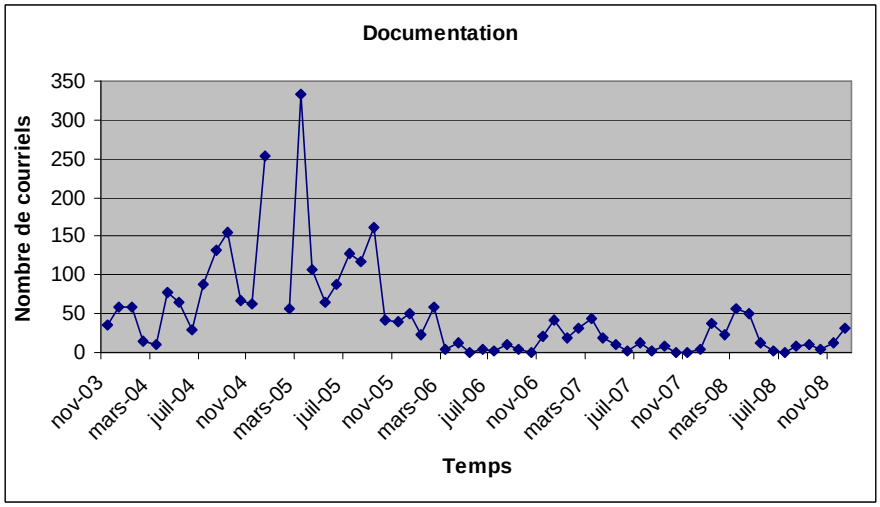


Figure 32 : série chronologique de la liste de discussion de la documentation

L'analyse poussée de la structuration de la liste fera l'objet d'un chapitre entier par la suite. Il suffit pour le moment dans ce chapitre de constater deux périodes marquées, l'une, de novembre 2003 à février 2006, comporte une phase de croissance marquée et de décroissance, l'autre, de mars 2006 à décembre 2008, est marquée par un régime plus faible, avec des oscillations de faible ampleur.

On remarque à partir de l'année 2006 une forme de saisonnalité stricte : la période été-automne se caractérise par une baisse de l'activité. Cette régularité tend à se vérifier en 2009, ce qui n'apparaît pas sur le graphique. Son interprétation serait qu'on délaisserait davantage l'ordinateur et donc les contributions aux sites pendant les périodes estivales[451]. Cependant, les oscillations peuvent avoir d'autres causes, que nous verrons ensuite.

[451] Nous remercions ici Carla Taramasco du CREA de nous avoir signalé qu'on observe ce type de saisonnalité chez les blogueurs, justement (Jean-Philippe Cointet du CREA a travaillé sur les blogs).

4.5.3.2. L'évolution du nombre d'articles

L'analyse des dynamiques comparées à l'intérieur de la communauté suppose de sélectionner les matériaux pour extraire de la façon la plus productive les données pertinentes. La quantité conséquente de données à dépouiller dans le cadre d'une analyse fine oblige à une sélection lorsque l'extraction est *coûteuse*, comme ici. Le dépouillement des articles comportant du contenu[452] a été fait d'une façon *manuelle*, pendant la semaine du 22 au 26 septembre 2008, ce qui neutralise les variations à la marge de la population pendant cette courte période (création d'articles supplémentaires pendant cette période) [453]. Cela correspond à la période de cette thèse où l'opération a été effectuée; compléter les données sur les mois ultérieurs est bien entendu possible ultérieurement.

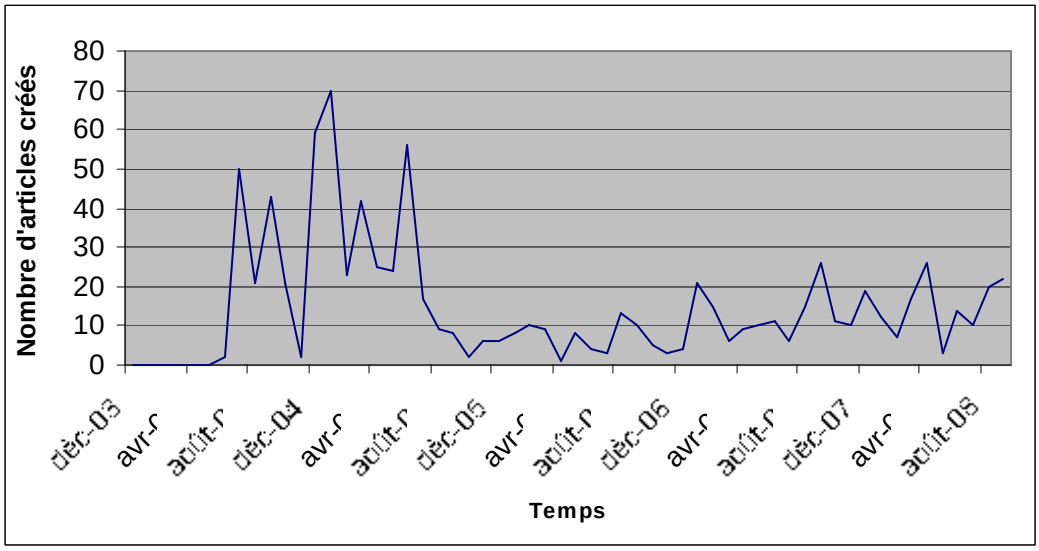


Figure 33 : graphique montrant l'évolution de la création d'articles

L'activité de création d'articles sur le wiki a été continue sur la totalité de la période. Ce graphique, montre qu'il existe quatre phases dans l'histoire de la communauté de la documentation:

- Une première phase correspondant à la vie du premier wiki, qui n'apparaît pas ici[454].

[452] Nous avons exclu les articles n'appartenant pas au « main space » (comme les pages de discussions) et les articles en langue étrangère.

[453] Obtenir cette courbe suppose un travail conséquent de dépouillement des 862 articles du wiki. Nous insistons sur le fait qu'étudier des archives de la Toile de façon fine n'est pas du tout ergonomique (nombre considérable de clics de souris par exemple). D'autres méthodes auraient été préférables et intéresseraient des chercheurs ultérieurs.

[454] Le service Web archives ne permet pas hélas de retrouver des statistiques sur l'ancien SI.

- Une deuxième phase, de juillet 2004 à *août* 2005, correspondant à la phase de croissance de l'activité de documentation.
- Une troisième phase d'activité plus réduite, de septembre 2005 à décembre 2006, avec des plages de quasi inactivité et des phases ponctuelles de reprise.
- Une phase de reprise lente d'activité à partir de janvier 2007, avec des oscillations de faible ampleur.

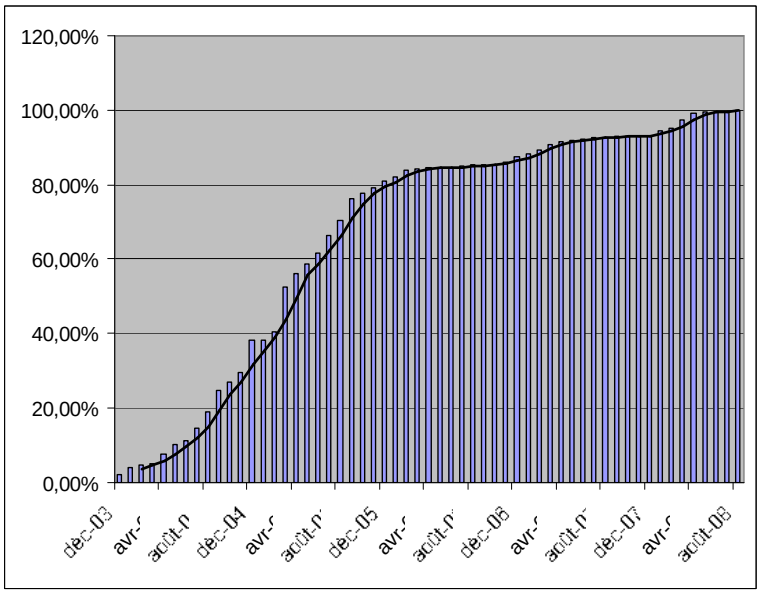


Figure 34 : croissance du nombre d'articles créés (en pourcentage cumulé)

Le coefficient de corrélation entre la série chronologique des nouveaux articles et celle de la discussion, de 0,63, bien qu'honorable, montre que la corrélation est partielle, malgré des ressemblances générales: on peut en effet remarquer facilement que les deux séries sont semblables. Il existe cependant des décalages ; en particulier, la baisse du nombre d'articles créés de septembre 2005 a lieu avant la baisse de l'activité de coordination, qui est vraiment marquée début 2006.

Il faut insister sur les limites de cet indicateur, pourtant légitime et employé par d'autres[455] : il ne s'agit que des dates de créations d'articles ; il suffit de parcourir quelques historiques d'articles pour se rendre compte que l'activité d'un nombre significatif d'entre eux dure plus d'un an. En revanche, beaucoup de petits articles sont édités à leurs débuts, ce qui signifie qu'il désigne tout de même un mouvement d'éditions *significatif*, et cet indicateur est utile pour comprendre directement l'évolution de la *structure* du wiki, qui dépend de façon

[455] BESTEN et DALLE (2008).

cruciale de la croissance des articles : un excès d'articles fin 2005, comme le suggère la courbe, peut expliquer des difficultés évoquées en matière de structuration anarchique du contenu qui apparaissaient dans notre récit d'histoire immédiate.

Ce graphique montre aussi que 80 % des articles existaient déjà en décembre 2005, ce qui *confirme* les résultats de l'histoire narrative précédente sur la date de stabilisation du Codex. Il confirme aussi plus à la marge la présence des périodes d'organisation de décembre 2004 et avril 2005 (« marches d'escalier » sur le graphique) également dégagées lors la précédente analyse, mais ce point pourrait davantage prêter à discussion. L'activité de régulation de la documentation comporte une *structure semblable* dans ses grandes lignes, l'essentiel de l'activité de régulation est effectué à 80 % entre novembre 2003 et janvier 2006. Nous ne présentons pas le graphique qui ressemble à s'y méprendre au précédent.

Cette similitude et corrélation entre création d'articles et gouvernance sur la liste de discussions souligne l'importance de la structuration du contenu par création de nouveaux articles. On peut faire l'hypothèse que la corrélation forte entre régulation/gouvernance de la documentation et nombre de nouveaux articles créés montre que ce serait surtout eux qui seraient régulés, et beaucoup moins l'activité décentralisée d'édition ultérieure des articles, plus continue voire routinière. Ceci est cohérent avec l'idée de régulation de la *structure* du wiki, qui posait l'essentiel des problèmes. Cette corrélation est significative, mais l'analyse *qualitative* précédente de la liste de discussion montre que la gouvernance ne se résume pas à cela, la corrélation n'étant pas parfaite, rappelons-le.

Cette chronologie ne va pas de soi, et ne peut pas clairement être déduite des données qualitatives. On peut cependant comprendre que la croissance du nombre d'articles pour un manuel ne soit pas infinie, et que la communauté désire en limiter le nombre, ce qui est certainement différent du modèle encyclopédique de Wikipedia.

4.5.3.3. La reconstitution du nombre d'éditions sur le wiki

Il a été nécessaire et possible de reconstituer l'activité du wiki, à partir des données issues du dépouillement de la datation des articles, présentée précédemment. A partir de web.archives.org, nous disposions de quelques traces à intervalles réguliers sur le nombre d'éditions sur le wiki, mais aussi du ratio nombre d'éditions par article, qui augmente légèrement mais régulièrement sur la durée de vie du wiki (entre 4 et 6 éditions en moyenne par article). Cette quasi constance du ratio est *un moyen commode* qui permet de déduire par

interpolations et tâtonnements un ratio mensuel sur la durée de vie du wiki. A partir du nombre d'articles et de l'estimation mensuelle du ratio en évolution légère, on comprendra que nous ayons pu obtenir par multiplication une estimation du nombre *total* d'éditions mensuelles sur le wiki pour les articles comportant du contenu.

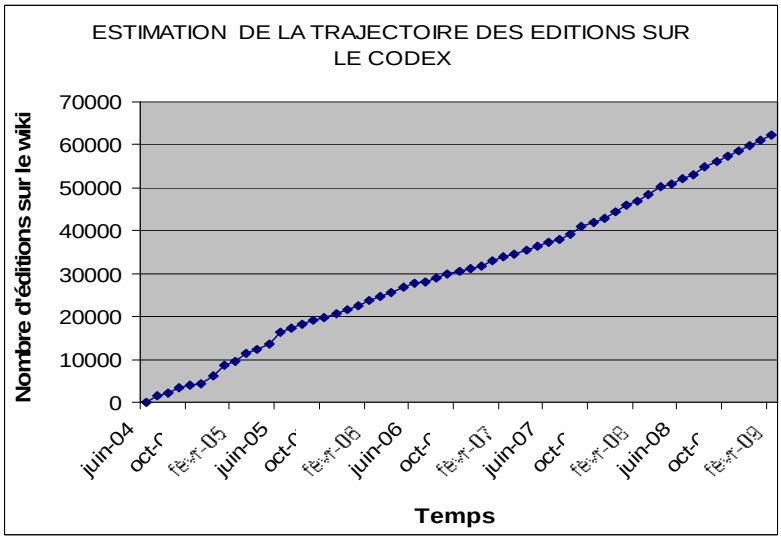


Figure 35 : reconstitution de l'évolution du nombre d'éditions sur le wiki

Cette estimation permet au minimum de comprendre que la trajectoire des éditions sur le wiki *ne ressemble pas du tout* à celle de l'activité de coordination et à celle de la création d'articles. L'activité d'éditions et d'ajout de contenu est continue et régulière. Cela n'allait pas de soi de prime abord.

Pour parler de façon imagée, il semblerait que l'activité sur le wiki ressemble à un décollage au moyen d'un tremplin : le tremplin est l'activité de coordination initiale décrite précédemment (structure des pages, liste de discussion) dont la pente ralentit, l'objet « volant » dont la trajectoire est croissante et rectiligne étant l'activité d'édition[456].

[456] On pourrait trouver assez facilement une corrélation entre l'activité sur le forum et la croissance du nombre d'éditions sur le wiki, mais un tel rapprochement serait critiquable (on peut toujours trouver une corrélation entre deux relations linéaires alors qu'elles n'ont aucun lien entre elles). On peut s'attendre au contraire à une absence de corrélation, les deux activités étant plutôt substituables (une hypothèse difficile à tester aussi dans un contexte de forte demande d'aide aux utilisateurs).

4.5.4. Comparaison des rythmes des différentes activités de la communauté et mise en évidence de la singularité de la documentation.

Nous tentons à présent une synthèse à partir des séries chronologiques précédentes.

Pourquoi a-t-on observé les cycles précédents dans l'activité de programmation? Un passage parmi les sources permet de comprendre:

> "In 2007 we adopted a regular release schedule, putting out major feature releases roughly every 3-4 months, or three times a year."[457]

C'est effectivement ce qui s'est passé : on observe une mise en relation de l'activité informatique de la communauté et de la publication des nouvelles versions. Cependant, l'apport du dépouillement et de la quantification permet de montrer que cette politique a été *effectivement* suivie ; cela aurait bien pu ne pas être le cas, faute d'un nombre de programmeurs ou d'un leadership suffisant, par exemple.

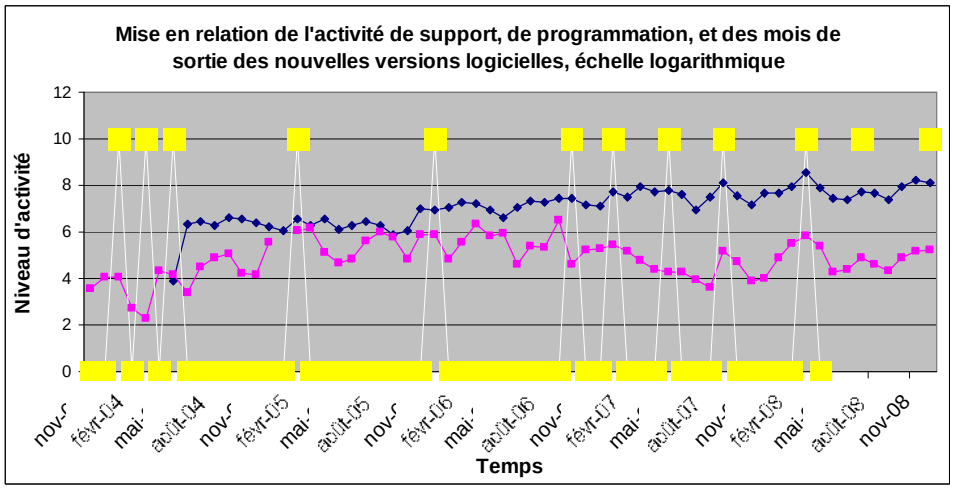


Figure 36 : séries chronologiques de synthèse des activités de support et de programmation, échelle logarithmique

Ici, la *courbe du bas* correspond à l'addition de l'activité mensuelle de régulation des activités de support (10899 courriels), qui inclut la documentation. La *courbe du haut* correspond à l'addition de l'activité mensuelle des activités de programmation (84930 courriels). On constate que la coïncidence entre les sorties du logiciel (les carrés en haut des lignes blanches) et l'activité des listes de programmation ne fonctionne bien qu'en 2007-2008. On remarque aussi, et c'est là l'utilité entre autres de l'échelle logarithmique, certaines corrélations *plus*

[457] Source: http://wordpress.org/about/ .

locales entre l'activité de support et de programmation, dans la mesure où les nouvelles versions du logiciel doivent être documentées : ce sont les sorties des nouvelles versions qui causeraient dans l'une et l'autre des activités de programmation et de support une augmentation de l'intensité des discussions.

Lorsqu'on examine seulement les deux séries chronologiques agrégées précédentes, on peut remarquer plusieurs éléments. Les niveaux d'activité *comparables* des activités agrégées de support aux utilisateurs et de programmation de 2003 à 2006 montrent que la première a été prise en compte avec un sérieux semblable. Après janvier 2007, la programmation a pris le dessus. Ces données permettent de dire clairement que la documentation n'a pas « suivi » l'activité de programmation, et qu'elle l'a à coup sûr significativement précédée, au moins sur les listes de discussions, ce qui peut être important pour expliquer le succès significatif de celle-ci. Ces données peuvent suggérer aussi deux résultats, i.e. que les problèmes de bien collectif du premier ordre (contributions simples) comme du second ordre (contribution à l'organisation) au niveau des activités auxiliaires de documentation ont été résolus malgré la dévalorisation associée à l'activité de documentation, ou que la croissance de l'activité plus prestigieuse de programmation s'est faite *suffisamment tard* pour ne plus provoquer un « effet d'éviction » gênant. Nous entendons par là que la montée en puissance d'une activité prestigieuse, la programmation, incite à la contribution en matière de code, et n'incite pas à donner autrement, en matière d'aide à l'utilisateur.

Par ailleurs, sur le graphique précédent de l'historique de sortie des versions du logiciel, on a constaté en particulier en 2005 un ralentissement de cette activité. Elle pourrait s'expliquer par une insistance par *substitution* sur l'activité de support au détriment du code. Ce n'est pas certain, mais cela est cohérent avec le précédent constat. Les chiffres sont aussi relatifs, étant donné qu'il existait tout de même une activité notable de programmation dès le début de l'existence de la communauté. Cependant, il nous était difficile de ne pas souligner une probable singularité de cet ordre.

Enfin, l'activité de programmation et l'activité de support dans son ensemble n'obéissent pas à la même logique et aux mêmes rythmes. L'activité de support aux utilisateurs comporte trois phases de régulation *successives* de ses composantes, documentation, forum utilisateur, traduction de la documentation :

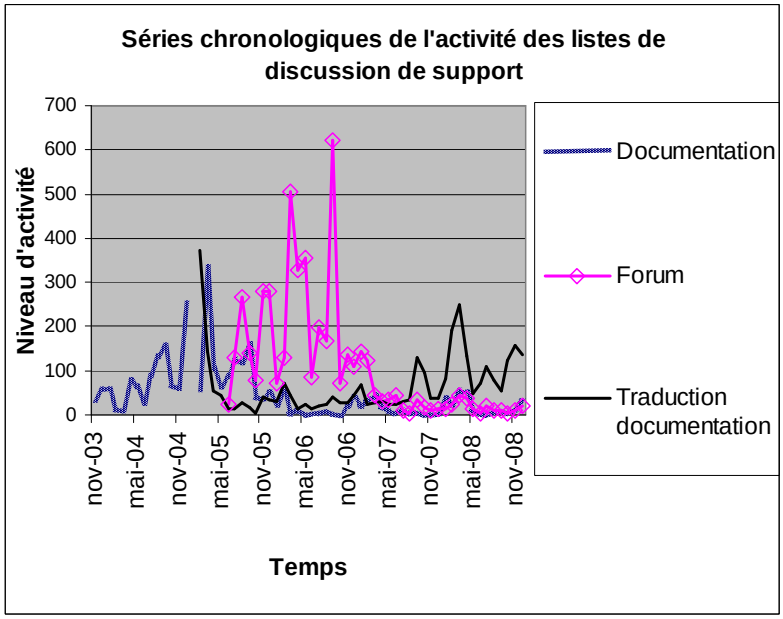


Figure 37 : séries chronologiques de l'activité des listes de discussions de support aux utilisateurs

On observe donc ci-dessus des rythmes différents de ceux de la programmation : les activités de programmation ont fait *converger* leurs activités, afin de les synchroniser. Ici, au contraire, on observe trois vagues *successives*, comme si les tâches accomplies (mise en place de régulations) s'étaient effectuées en un temps limité et défini[458]. Les rythmes et les complémentarités à l'intérieur des deux parties de la communauté obéissent donc à des régularités *complètement différentes*.

4.6. Synthèse, conclusions, limites de notre approche et ouverture vers les autres chapitres d'approfondissements

Nous arrivons à quelques conclusions, qui ont une portée historique pour l'histoire de la communauté, et « managériale ».

En effet, nous avons mis en évidence des *spécificités* en matière de rythmes dans la communauté, des convergences ou des contrastes, qu'il s'agisse de l'activité interne à l'activité de programmation (convergences), de l'activité de documentation (convergence et

[458] On pourrait certes approfondir l'étude de leurs différences d'un point de vue quantitatif par exemple, mais cela dépasserait notre sujet.

divergence), des activités de support (divergences). Les régularités observées suggèrent qu'il existerait plusieurs phases de régulations et d'activités, différenciées et successives. Tout cela aide à comprendre dans une certaine mesure la collaboration distribuée dans un contexte communautaire différent de Wikipedia : cette forme organisationnelle n'est pas un objet isolé, elle comporte une dimension *temporelle*. Elle comporte des rythmes spécifiques mais aussi des formes d'articulations, dont on peut observer l'évolution en se fondant sur les traces laissées, sous forme d'archives, sur le web. Le système d'information de « knowledge management » n'est pas indépendant du reste de la communauté, mais il comporte des spécificités.

L'histoire « événementielle » est également riche d'enseignements. La documentation a subi un processus d'évolution : d'une structure de FAQ à un wiki ouvert, pour finir sur un modèle intermédiaire entre Scholarpedia et Wikipedia, avec un processus de validation, comme pour le code. Elle montre un travail permanent, approfondi, d'organisation de la documentation, avec des formes de cycles, au-delà des phases d'activité importante et de calme prolongé. On observe en particulier une alternance de phases de contributions et d'organisation, lorsqu'on « tamise » les principaux événements sur des périodes d'un ou deux mois. Ce n'est pas seulement un travail d'ajout de contenu, de mise en forme, mais un *véritable travail de conception collective d'un système d'information*.

Sur le plan méthodologique, notre présent travail apporte des enseignements, car les études longitudinales, ici historiques, d'une communauté, ne sont pas nombreuses. Il montre que des formes simples mais parfois laborieuses de dépouillements d'archives communautaires, des reconstitutions historiques par interpolations à partir du service web.archives.org permettent de mettre en évidence des régularités et des spécificités nettes. Il propose aussi à d'autres chercheurs des conseils de prudence, certains points n'étant pas faciles à interpréter (l'activité d'un forum). Cependant, *l'approche quantitative seule est insuffisante* pour interpréter et examiner la portée des données. La tentative qui doit être soulignée est celle de l'histoire « événementielle », les listes de discussions n'ayant pas été à notre connaissance traitées dans cette optique dans la littérature.

Pour un « gestionnaire » de communauté open-source, ce type de dépouillement et de prise de recul historique seul peut être riche en enseignements : il donne une idée de *ce à quoi s'attendre*, un repère pour établir un point de *comparaison*, et, dans une certaine mesure, un *modèle à suivre*. Le cas particulier de WordPress enseigne que la prise en compte de la documentation *avant* (plutôt qu'après) que ne se mette en place « le rythme de croisière » de la programmation peut être un élément tout à fait significatif. Deux mécanismes sont à

l'œuvre. *D'une part*, l'accroissement remarquable de l'activité de la programmation et de la communauté dépend de l'existence d'une masse critique d'utilisateurs… qui dépend de l'efficacité des activités de support, en particulier de la documentation, qui doit être donc constituée *avant*. Ce processus est d'importance, même s'il n'est pas précisément documenté ici. *D'autre part*, si la documentation d'un projet connu peut être motivante, elle peut aussi l'être moins cependant lors d'une phase ultérieure de succès grandissant, où les programmeurs retirent davantage de bénéfices à programmer. Pour ceux qui veulent tirer des bénéfices de l'activité communautaire (services), mieux vaut « en profiter » ainsi qu'en aidant les utilisateurs en écrivant de la documentation. C'est là un « effet d'éviction » (« crowding-out effect ») qui aurait été ici évité.

Ce chapitre historique sur les dynamiques de la communauté a permis de prendre un recul macroscopique. Il a donc fait seulement allusion à des thèmes essentiels pour notre problématique. Il justifie donc des analyses ultérieures plus approfondies que nous mènerons plus loin dans cette thèse. Il s'agira en particulier :

- De l'analyse de la population de contributeurs, des contributions, et des motivations qui permettent le passage de l'un à l'autre. En effet, c'est de cela que dépend par exemple la dynamique des éditions sur le wiki que nous avons dégagée dans ce présent chapitre.
- De l'analyse des routines mise en place par la communauté. Notre récit a fait allusion à certaines d'entre elles, comme celle des sous-pages utilisateurs. Pourquoi ont-elles émergé ?
- De l'analyse de la dynamique de la liste de discussions. Nous n'en avons donné qu'une vue ponctuelle ; il manque une étude interne de sa structuration, essentielle pour mieux comprendre les formes de collaborations des contributeurs au Codex.
- De l'analyse de l'interface entre le wiki et le reste de la communauté. Rappelons-le, nous n'avons pas pu obtenir de résultats définitifs sur ce point dans ce chapitre. Il n'est pas facile de caractériser le lien « rythmique » et « physiologique » entre le SI du wiki et les autres parties de la communauté, mentionnées ici. L'analyse dans le dernier chapitre de l'efficience du SI à partir de traces laisse entendre qu'une telle voie de recherche est particulièrement ardue.

Le présent travail comporte des limites et suggère d'autres perspectives de recherches, au-delà des précédentes que nous aborderons dans cette thèse. La présente approche

historique comporte des limites par comparaison avec une thèse d'histoire « pure ». Nous aurions pu étudier beaucoup d'autres thématiques comportant des enjeux gestionnaires et toujours selon une approche historique, mais différentes de celle abordée ici[459]. Il faudrait peut-être trouver des moyens d'estimer de façon plus précise encore les éditions sur le wiki à partir d'outils d'extraction informatique. Rappelons également que les tendances observées comportent des limites, car nous avons pris parti d'estimer à partir des listes de discussions, ce qui donne des tendances dans une certaine mesure agrégées.

Il reste aussi à savoir *en priorité* si de telles régularités temporelles se retrouvent dans d'autres communautés ou si elles sont significativement originales. Les études longitudinales des communautés ne sont pas fréquentes, de sorte qu'il est impossible actuellement de répondre à cette question d'après ce que nous savons de la littérature empirique ayant porté sur ces sujets *précis*.

[459] Pour ce qui concerne les informaticiens, le noyau dur des communautés, on pourra penser en particulier à l'analyse de donnée de type « Concurrent version system » (CVS) brutes (voir sur ce sujet par exemple ROBLES et al., 2005), ou à l'utilisation sur des données d'informaticiens du même type d'une analyse de réseaux avec Pajek, (perspective ouverte par LOPEZ-FERNANDEZ et al., 2004).

Chapitre 5 : « just my 2 cents »[460] ? Analyse des contributions des participants au Codex et de leurs motivations

5.1. Introduction

Nous nous intéressons dans ce chapitre à l'un des points décrits selon une perspective dynamique dans le chapitre précédent, la constitution selon une trajectoire *linéaire*, à partir de *contributions bénévoles* multiples, du bien informationnel qu'est le Codex. L'économie cognitive problématise ce type de situation comme la constitution d'un bien public ou collectif (chapitre 1). Une telle émergence est une problématique de gouvernance, même si le processus peut être spontané, ou qu'il ne pose problème que de façon marginale.

Nous cherchons donc ici à *comprendre* de façon plus approfondie à partir d'analyse quantitatives et qualitatives les *caractéristiques* de la constitution du bien collectif (i.e. un processus dynamique).

Ainsi, si nous nous intéressons à une problématique d'économie cognitive, nous n'envisageons pas de tester des théories existantes (à part l'infirmation désormais classique du paradoxe du bien collectif), mais ce n'est pas à exclure par la suite. Notre priorité était de *tester l'intrigue* précédente, de spécifier un processus, plutôt qu'une théorie. En un seul chapitre, c'était un objectif raisonnable, car traiter une telle question sans aide informatique est ardu.

Aborder cette question directement n'est pas possible avec les moyens dont nous disposons. Nous en étudierons donc les caractéristiques à partir de plusieurs points de vue (correspondant à plusieurs sources), les trajectoires d'intervention des contributeurs, et des données qualitatives permettant de comprendre leurs motivations, la difficulté de la situation de gestion et la mesure de la résolution du problème.

Après avoir brièvement présenté notre méthodologie, nous commencerons par présenter les difficultés du travail de documentation, à l'origine de la problématique de bien collectif. Nous présentons, à partir d'une base de données des contributeurs au Codex, des éléments sur les contributions individuelles de chacun (quantitatifs, temporels). Ensuite, à

[460] Cette expression est très employée dans la communauté : elle signifie qu'une personne apporte une contribution qu'elle juge minime (comme à la quête par exemple), qu'il s'agisse d'idées sur la liste de discussion, ou de contributions directes au Codex. Mais poser ici la question signifie que cela peut aller bien au-delà dans un nombre significatif de cas.

partir d'un échantillon d'articles du Codex, nous présentons des éléments quantitatifs sur la structuration des contributions au contenu. Puis, à partir de la précédente base de données « contributeurs », nous dégageons des données qualitatives disponibles les motivations qui y apparaissent. Enfin, à partir des données qualitatives issues de la liste de discussions de la documentation, nous renforçons l'idée que la communauté a résolu la problématique de formation du bien collectif en matière de contributions.

5.2. Eléments de méthodologie

Nous spécifions la méthodologie en complément du chapitre général précédent sur la question. La démarche adoptée est essentiellement de type archéologique : nous avons en effet travaillé les strates des historiques des wikis afin d'en extraire les éléments pertinents, comme un archéologue cherche des traces de la vie quotidienne d'un village sur un seul site, sous plusieurs mètres de terre.

Tout d'abord, nous avons constitué une *base de données* à partir des informations disponibles sur les utilisateurs[461] :

- pour les données *qualitatives*, il s'agit des commentaires divers sur leurs pages utilisateurs, (« user:page », une fonctionnalité proposée sur MediaWiki), ou d'autres informations obtenues à partir de ce qu'ils disent d'eux sur leurs blogs, en particulier les pages dites « about », devenues standards, qui permettent d'obtenir des éléments biographiques ou des indices sur les motivations.
- pour les données *quantitatives*, il s'agit du nombre d'éditions sur le wiki, de différents types, ou du nombre de jours d'intervention (uniquement pour ceux qui ont contribué plus de 10 fois).

D'autres formes de dépouillements ont été effectuées, en particulier un sondage sur un peu plus d'une centaine d'articles du wiki. Nous avons également extrait de la liste de discussion de la documentation un certain nombre d'éléments pertinents correspondant aux motivations ou aux problématiques de constitution du bien collectif, ainsi que quelques fils de discussions du forum utilisateurs extraits à partir des mots clefs associés à la documentation ou au Codex. On peut associer ce type d'extraction à des formes de codage simples, avec un

[461] Nous n'avons étudié que les contributeurs anglophones et francophones. En effet, un nombre significatif de contributeurs viennent de pays très divers, ils peuvent contribuer en langue anglaise ou dans leur langue maternelle, en particulier en matière de traduction.

seul code. Mais l'utilité marginale qui consiste à nommer ainsi le traitement des fragments pertinents n'est pas significative.

La méthodologie historique qui consistait à croiser diverses sources au prix de dépouillements conséquents était ici légitime : un wiki est un objet laborieux à étudier comportant un nombre considérable de modifications correspondant à chaque édition.

5.3. La documentation est une activité laborieuse et guère reconnue

Nous verrons ici que la documentation est une activité laborieuse et guère reconnue et que pourtant elle doit se caractériser par l'efficacité professionnelle. Il s'agit donc un objet de gestion, au sens de J. Girin, tel que nous l'avions défini dans notre chapitre introductif. Nous devions ici approfondir ce point.

Les données disponibles, des traces, montrent que la constitution du bien collectif qui nous intéresse butte sur deux obstacles, au-delà de l'absence de bénéfice clair associé : la difficulté de la tâche, et l'absence de reconnaissance, ce qui n'est pas incitatif. Que la tâche soit ardue est bien documenté, en particulier par des mots particulièrement forts : nous citons à différentes dates quatre passages de cet ordre où nous soulignons nous-mêmes les termes en question qui ont l'avantage de donner au lecteur un aperçu concret de cette difficulté:

> "Yes, that's *tedious as HELL* to write, but you know, it's worth doing in the long run."[462]
>
> "Having written a number of Docs & Procedure manuals for labs I know that Doc organisation is a *nightmare*. I will add my 2bobs worth if I can. My impression is the codex does a fair job of a *giant task*."[463]
>
> "However, putting questions of technology aside, the larger problem with WordPress' documentation is that, except for the recent screencast guy who publishes release videos, the documentation effort seems to be mainly put in the hands of the community. Isn't this a lot to ask of the community? I think WordPress should hire a professional technical writer to manage the entire project. I don't mean any disrepect to the documentation team with this recommendation, only that this is *too monumental of a task* to leave entirely up to the community to create and maintain. Tom."[464]
>
> « There are about 1000 pages on the Codex, and though that includes pages which are not a part of the documentation, managing this huge list of pages and resources is *challenging* »[465]

[462] Source : http://comox.textdrive.com/pipermail/docs/2003-November/000026.html .
[463] Source : http://wordpress.org/support/topic/37107. Le post date d'il y a « quatre ans », la précision de la datation du forum ne permet pas de dater précisément.
[464] Source : http://comox.textdrive.com/pipermail/wp-docs/2009-January/001876.html .
[465] Source : http://codex.wordpress.org/Codex:Maintenance (consulté le 01/09/09).

La difficulté de la documentation, ainsi bien établie, comporte plusieurs aspects, que nous ne faisons ci-dessous que présenter rapidement.

Au fur et à mesure de sa croissance, un tel travail en vient à correspondre à un emploi à plein temps, ce qui n'a pas empêché la question d'être négligée (sous-traitée à la communauté ?) par la start-up :

> "One thing I can't understand about Automattic is their lack of attention about documentation. […] Don't take this as an insult to the wonderful content you've created on the Codex. I'm just saying the Codex is a full-time job to maintain, and all the useful info on the forums, the hacks and the unknown tips, should be captured and somehow put into the Codex."[466].

Cette idée était déjà visible dans le troisième fragment cité dans cette sous-partie (« should hire a professional technical writer to manage the entire project »). Lors de cette évolution, la nécessité du professionnalisme apparaît, ce qui ne veut pas dire que le remplacement de la création collaborative par un emploi à plein temps soit pour autant envisagé. Au contraire, le modèle de collaboration et contribution sous-jacent à Wikipedia comporte une légitimité une année après le début de la documentation:

> "Wikis are awesome tools, but there is a level of professionalism that must be maintained in order for users to feel like it's authoritative. The main reason I find Wikipedia so outstanding is that so many pages have been so thoroughly and professionally reviewed that its users have a high level of confidence in its content. Our documentation must look equally professional."[467]

La difficulté de la documentation se caractérise aussi par la nécessité d'une expertise locale[468] et une prise en compte de la pluralité des auditoires, une idée récurrente pour laquelle nous ne présentons qu'un fragment : "We need to cater to all needs from totally web ignorant to master developer. That makes the task of the Codex structure a challenge"[469]. Enfin, il peut exister des sujets sensibles, par exemple ceux liés à la sécurité du logiciel, qui rendent l'écriture de la documentation délicate, ce qui est aussi dissuasif :

> « Hardening WordPress - This is really, really controversial. I know it's not on your list, but security topics are and this is one of them. We need to clearly address the issue of security flaws and potential risks in WordPress, especially in WordPress Plugins, while not exposing flaws or potential "back doors" into WordPress. This is a big one and we really need to get a consensus and policy on how to write about this because it is SO important, while not opening a door we don't want open."[470]

[466] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-May/001737.html .
[467] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000926.html .
[468] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-May/000505.html .
[469] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001478.html .
[470] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001491.html .

Nous avons souligné le fait que la documentation soit difficile et exigeante. Nous faisons de même à présent avec les problèmes de légitimation. L'idée de formes de récompenses monétaires a été proposée, mais n'a pas été concrétisée[471]. Par comparaison avec les avantages connus de la constitution du code par les développeurs, dont la communauté a conscience, on observe à plusieurs reprises des plaintes concernant le manque de reconnaissance de la documentation, que ce soit à l'extérieur de la communauté, comme le montre le premier fragment suivant, ou à l'intérieur, comme le montre le second ci-dessous:

> "The folks contributing documentation don't get even that intangible reward. Docs are every bit as thankless as writing code, and _far_ less likely to generate wishlist purchases or Paypal donations. I don't think there's anything wrong at all with some visible attribution in the Codex so that the contributors can generate for themselves the same kind of documented record as the developers enjoy. This may help the documentors in job interviews, other Free Software projects, etc. "[472]
>
> "Podz is defending the position that we _DO_ care. Writing docs is hard. A lot of effort went into making the Codex the resource that it is. The contributors deserve every bit as much credit as the person who submits an elegant piece of code to solve some problem inside WordPress."[473]

La question de savoir si cette légitimité a émergé par la suite est une autre question, sur laquelle nous n'avons pas de certitude d'après notre connaissance des archives. Nous ne faisons que montrer que le problème est apparu et explicité en septembre 2005, la date des deux fragments ci-dessus, et donc qu'il a existé au minimum pendant les deux premières années de la documentation, c'est-à-dire pendant sa phase de croissance, comme on l'a vu au chapitre précédent.

5.4. Le problème d'incitation du premier ordre a été résolu par la communauté, mais d'autres problèmes de création du bien collectif se posent

Nous avons examiné précédemment en quoi la constitution du bien collectif était ardue dans un tel contexte communautaire. Cependant, le dépouillement des archives montre que le problème de motivation à la contribution[474] n'est pas du tout central pendant la totalité de l'histoire de la communauté, ce qui justifie de porter notre attention dans le reste de cette thèse sur d'autres thématiques mentionnées précédemment. Aucun fil de discussion ne porte

[471] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000884.html . Pour plus de détails sur ce point, nous renvoyons au chapitre suivant.
[472] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000988.html .
[473] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000956.html .
[474] Signalons qu'il est nécessaire de distinguer problème d'incitation et problème de reconnaissance pour les contributions offertes.

spécifiquement sur le sujet par exemple, nous n'avons trouvé que quelques fragments dispersés dans les listes de discussions qui ne permettent pas de conclure à des problèmes récurrents sur des périodes longues ; si cela avait été le cas, nous en aurions observé au moins un. On observe en effet quelques passages où des insuffisances en matière de contributions pourraient apparaître[475], mais elles sont marginales. On observe aussi de rares passages localisés lors d'événements demandant plus de main d'oeuvre, comme celui-ci qui porte sur l'événement « wiki-day » :

> "Why aren't there more people wanting to contribute to Wiki-Day given that I announced it in here as well as on the forums ? We've still got people asking about having more docs, better docs, docs that describe something which probably only they will ever need and people who say inane things like "why don't WE make a wiki page" when what they mean is someone else should do it for them.
>
> I know the docs is a task that will never attract lots of help, but just looking back through the forums to see what we need, and seeing the amount of discussion that has taken place on here, I just wonder why all these silent members are here and being included."[476]

L'histoire de la communauté a montré qu'il y avait eu une période de crise début 2006. A ce moment, effectivement, un passage montre que les contributions ont baissé de régime :

> "Hello All,
>
> We need to increase interest in the codex. Plenty of people use it, but not many have added / edited it during the past few months. Lorelle was great at getting people to work on it, and now that her presence on IRC has diminished, I think we need to push ourselves. How about a Codex drive (a week? two weeks?) sometime in the near future. Stuff we could look at could include new documentation (roles etc) for WP 2.0. Thoughts, anyone? »[477]

Si on cherche des problèmes d'incitations, on pourra toujours finir par en trouver[478], mais cela ne signifie pas qu'ils soient décisifs. Si cela avait réellement posé problème, on aurait trouvé plus de citations. Cela ne veut pas dire que les wikis ou la collaboration massivement distribuée ne soient pas confrontés à ce problème, mais il semble résolu ici… sans dispositif incitatif *institutionnalisé*, tel que le concevrait l'économie.

En effet, d'autres fragments montrent au contraire qu'il n'y a jamais eu de crise sérieuse des « vocations », et la masse des contributeurs a été satisfaisante pendant toute la phase de croissance (jusqu'à décembre 2005) :

---

[475] Ils sont trois, en comptant large pourtant : http://comox.textdrive.com/pipermail/docs/2004-May/000294.html et http://comox.textdrive.com/pipermail/wp-docs/2005-April/000422.html ou ici http://comox.textdrive.com/pipermail/wp-docs/2005-April/000431.html .

[476] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000881.html .

[477] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-January/001205.html .

[478] Hervé Dumez dans un séminaire sur l'étude de cas du DEA GDO en 2005 mettait en garde contre ce type de lecture qui permet d'appliquer un cadre théorique non-pertinent sur n'importe quel cas.

> “Fortunately, we have always had more than just a pair of good hands at work at the documentation, throughout, and with a few more volunteers, I think that part of the show can keep running.”[479]

Malgré des difficultés marquées, la communauté souligne à la fois les difficultés de constitution et le caractère satisfaisant du produit fini, comme le montrent par exemple les deux passages ci-dessous :

> “I feel your frustration. WordPress is just another in a long line of developments which suffers from issues with documentation. It's the nature of the beast... documentation is the least fun, and the least relevant to those with the knowledge to document it. It does need to be done, I agree completely, but it's hard enough to get it done properly when you're paying people to do it... let alone when time-poor developers have to drop bug fixes or new features and put their technical writing hats on. It sucks, but it's certainly understandable. Having said all that, wordpress is one of the better examples out there, of volunteer-based support and documentation. Far from perfect, I agree, but nowhere near 'bad'.”[480]
>
> “I'm aware of the fact that Wordpress is an open source project and that documentation is traditionally a difficult part of open source projects. However, I think that Wordpress is doing very good on this point and that the documentation is very complete. But I think that if the structure of the documentation is improved, it can be a lot more usable and make plugin and theme development easier.”[481]

Il faut ici se rappeler des résultats quantitatifs du chapitre précédent, en particulier de la trajectoire des éditions : on constatait en cohérence avec ce qui précède que le mouvement de contributions au Codex croît de façon linéaire. Ainsi, les données quantitatives et qualitatives montrent que soutenir l’existence d’un problème d’incitation significatif n’est pas tenable.

En revanche, un *mécanisme plus significatif serait celui de la désincitation* : l’organisation problématique du wiki ne rend pas les contributions toujours aisées, ce qui peut avoir un effet dissuasif. Il y aurait ici un lien avec la problématique d’incitation du second ordre (OSTERLOH et ROTA, 2004): si cette dernière n’est pas résolue, celle qui précède, celle du premier ordre, risque d’en pâtir. Les propos de ce contributeur sont révélateurs en la matière :

> “I've been helping a number of people on irc and sometimes on the forums, and I have a bit of experience writing documentation, but with the Wiki, I just don't know where to start because I don't know how what I'm going to write fits in with an overall structure. This, I believe, is an inherent difficulty in using wikis for documentation: not that we shouldn't do it (it has the huge advantages that it makes contributing technically easy and creates a uniform appearance), but we need to cater for this.”[482]

La gouvernance, d’après la définition d’Hervé Dumez du chapitre 1, consiste à orienter « l’équipage », ici pour aider l’équipe à contribuer, les membres de la communauté ne sachant pas toujours s’y retrouver. Le problème de désincitation peut toucher même les

[479] Source : http://comox.textdrive.com/pipermail/wp-hackers/2005-May/000870.html .
[480] Source : http://wordpress.org/support/topic/138583 .
[481] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-February/001436.html .
[482] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000931.html .

administrateurs, comme ici, début 2006 : « Forgive me for not leaping to participate when the current wiki situation is abominable. »[483].

L'histoire de la communauté montre que les problèmes qu'elle a rencontrés consistent davantage en problèmes d'organisation qu'en problème d'incitations du premier ordre ou d'outil wiki, comme le montre la crise de début 2006 évoquée dans l'histoire de la communauté. Le phénomène est *encore* marqué début 2007, mais dans une moindre mesure, malgré tous les efforts effectués.

> "I have to say that the documentation on Wordpress is very complete and covers almost everything, which (I think) is a very good thing. However, I think there is a rather big problem with the Wordpress documentation, and that is its structure. It is very hard to find the thing you're looking for."[484]

La raison apparaît après, les contributions en soi posent moins de problèmes que leur organisation. Toute organisation a un coût, ce qu'on appelle une *problématique d'incitation du second ordre*, qui reste davantage pertinente:

> "As documents and additions have been coming faster than structures management, things have gotten a bit out of control. We're working on this and will have a new "version" very soon. Stay tuned and you'll see"[485]

Le passage suivant, plus précis et concret, montre que même avec la meilleure volonté du monde, cela ne suffit pas toujours et donne envie de jeter son tablier (… de « gouvernante ») :

> "Lorelle, you are a maintenance nightmare - a few more prolific users like you to churn out articles at this rate, and I throw in my towel! ;) "[486]

Ainsi, toute analyse d'une communauté ou d'une organisation qui ferait usage d'un wiki qui se focaliserait sur les seuls symptômes en matière d'outil technique ou d'incitations du premier ordre ferait un diagnostic insuffisant. *Ces considérations seront de ce fait amplement développées dans le prochain chapitre* sur les routines, les « editorial patterns » étant plus significatifs que les problèmes de motivation et d'incitation.

5.5. Eléments quantitatifs sur la structuration des contributions au Codex

Nous avons vu précédemment que la tâche de documentation n'était pas aisée, mais que le problème économique de premier ordre a été significativement résolu. Nous proposons

[483] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-March/001532.html .
[484] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-February/001436.html .
[485] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001486.html .
[486] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000199.html .

à présent des éléments quantitatifs sur la structuration du système d'information qui a quand même émergé : ces éléments structurels permettent de mieux le comprendre.

5.5.1. La structuration des articles

A partir des « pages spéciales » du wiki[487], il est possible d'obtenir une liste des articles le constituant, des plus longs aux plus courts. Plus précisément, cette liste donne le nombre d'articles comportant un nombre d'éditions donné. Par exemple, il y a 57 articles avec deux éditions, et 63 avec trois éditions. Le dépouillement pour ce qui concerne cette variable a eu lieu fin janvier 2008, et nous n'avons pas jugé utile de le mettre à jour, au vu du résultat classique obtenu.

Ce classement est pertinent, car la distribution observée est significativement différente des manuels classiques où les sections ont une taille uniforme (un constat utile pour le chapitre suivant). Il est par ailleurs utile de la connaître pour travailler la population d'articles au besoin.

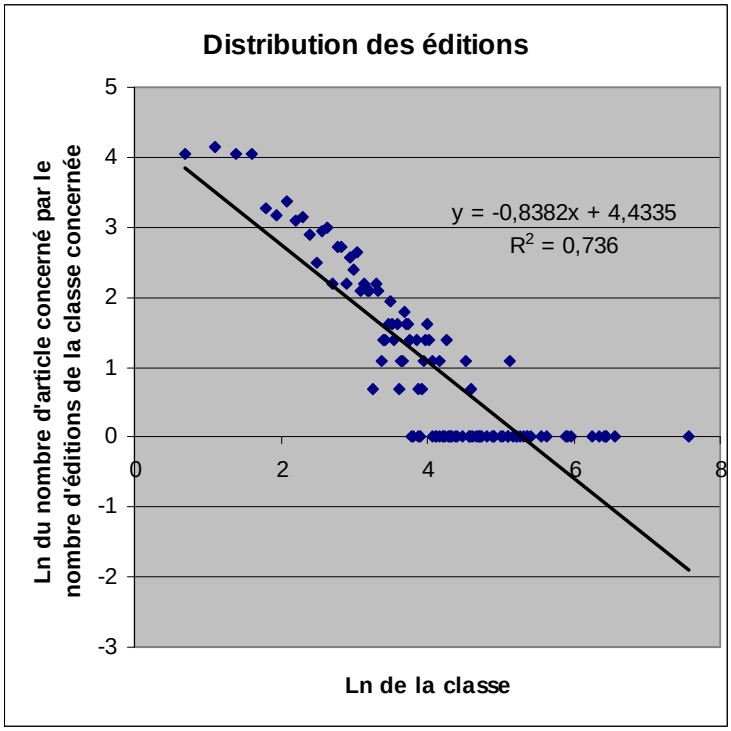


Figure 38 : distribution de la taille des articles sur le wiki en forme de loi de puissance

[487] Il s'agit de l'outil permettant de créer une liste des pages comportant le plus de révisions, des plus travaillées aux moins travaillées (Source : http://codex.wordpress.org/Special:MostRevisions).

Ainsi, après ordonnancement des données sur Excel, *on observe un phénomène de loi de puissance* en ce qui concerne la structuration du nombre des articles en fonction de leur nombre d'éditions. Bien qu'elle soit tronquée vers la gauche du fait que les articles comportent rarement une seule édition, sa forme est correcte bien que le nombre d'éléments dans la population d'articles étudiée soit moyen, ce qui ne favorise pas l'émergence de ce type de distribution. C'est ce que montre une régression simple, ajoutée ci-dessous ; il n'est guère nécessaire d'aller plus loin. On peut interpréter cela de différentes façons, l'origine des lois de puissance étant une question délicate, et il est impossible ici d'en déterminer l'origine. Cependant, cette structuration des articles au moment t peut avoir un rôle en matière de contributions au moment t+1 (« contribuer là où il y avait avant le plus d'éditions » ?) et créer un effet d'entraînement de type « boule de neige » ou « urne de Polya ». Il était donc nécessaire de la connaître.

5.5.2. La structuration des contributions

Voici à présent la modélisation (en deux temps) de la distribution des contributions de la population *entière* des contributeurs. La courbe obtenue ci-dessous à gauche a été tirée du dépouillement *manuel* pendant l'été 2006 des historiques des contributeurs « bleus » au wiki. Il s'agit de ceux qui ont « activé » leur page utilisateur, contrairement aux « rouges », qui ne l'ont pas fait[488]. Les contributeurs identifiables, et donc « bleus », sont ceux qui fournissent l'essentiel des contributions. Les données ont été travaillées sur Excel. Ils étaient 495 contributeurs « bleus » alors, correspondant à 21213 éditions[489]. Au vu de la difficulté de l'extraction de données, du temps imparti pour la thèse, et au vu de la nature du résultat, nous n'avons pas pu effectuer finalement de décompte supplémentaire par la suite, alors que le nombre des éditions a au moins triplé (nous renvoyons au chapitre précédent). Cependant, un tel décompte donne des informations précieuses sur la *première phase* cruciale de l'histoire de la communauté, que nous avons documentée dans le chapitre précédent. Voici ce que nous obtenons :

[488] Un « clic » pour édition sur un lien rouge correspondant à une page utilisateur n'existant pas encore suffit à le transformer en bleu et à créer une page vide attendant d'être remplie. C'est la raison pour laquelle *il ne faut pas y toucher*.

[489] Nous précisons que le dépouillement manuel des historiques suppose plusieurs semaines de travail, avec cependant l'aide d'onglets de comptage qui permettent de faciliter le travail pour les plus gros contributeurs. Ils permettent de séparer les historiques à volonté par « paquets » de 20, 50, 100, 500 éditions, plus aisées à décompter.

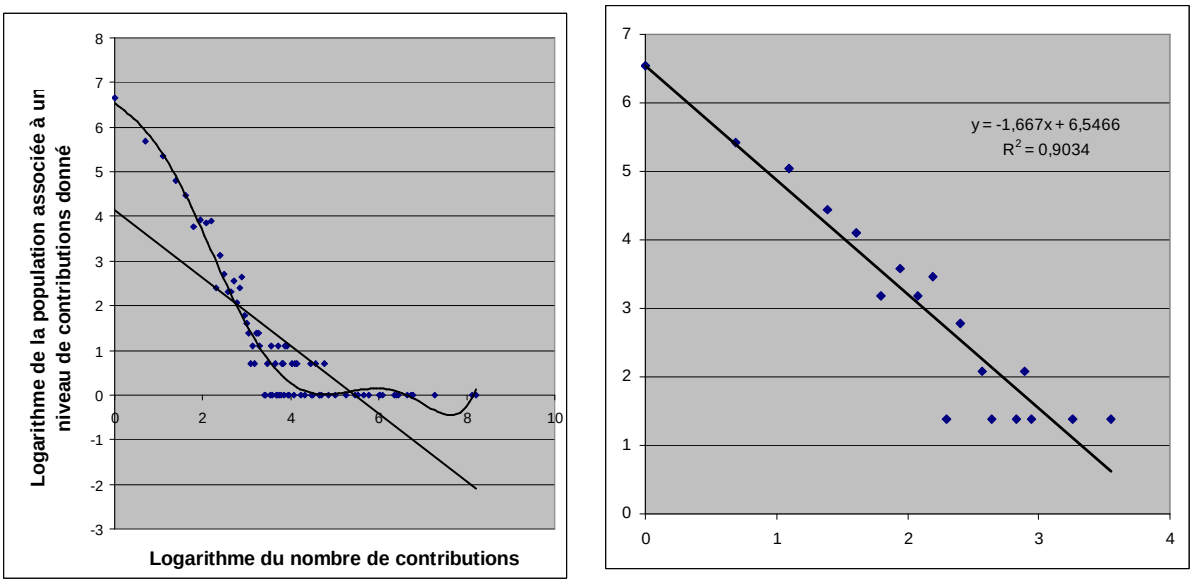


Figure 39 : distribution des contributions des volontaires « bleus » et « rouges »

Nous commentons le graphique de gauche ci-dessus. Il ne s'agit que des contributeurs « bleus », i.e. ceux qui ont activé leur page personnelle de présentation. Nous constatons donc qu'un nombre réduit de contributeurs se caractérise par un nombre considérable d'interventions. Cependant, à ce stade de l'histoire de la communauté, cette distribution est *intéressante*, car on constate un écart significatif entre une loi de puissance normale (rencontrée plus haut par exemple) qui irait d'une même valeur en abscisse et en ordonnée, précisément de 6 à 6 ici[490], et la distribution observée. Celle-ci passe de fait au dessus de cette droite, puis en-dessous, puis au-dessus[491].

Nous commentons à présent le graphique de droite ci-dessus. Il s'agit de la distribution des contributions des contributeurs « rouges ». Ils ne se sont pas identifiés et guère identifiables par le chercheur. Mais ils formeraient une part significative des contributions. Nous avons effectué un sondage en comptant les contributions d'un individu sur dix dans cette population, afin d'avoir une idée de ce qui se passe. Là encore, la forme de la distribution est semblable à ce qui est observé ci-dessus sur le graphique de gauche, une distribution en loi de puissance *imparfaite*.

[490] C'est la droite qui se rapprocherait *à vue d'œil* le plus de la distribution observée.
[491] Les contributions « bleues » présentées ici incluent les contributions qui consistent en discussions. Nous les avons enlevées pour avoir les contributions nettes, mais cela ne change pas significativement la forme de la courbe, que nous n'avons donc pas présentée ici.

Les plus gros contributeurs sont des administrateurs, qui contribuent de 233 éditions à 3389. *Les administrateurs ont effectué 12030 éditions : ces 2% des contributeurs ont ainsi effectué 56,71% des éditions*. Cependant, 29 autres contributeurs ont contribué plus de 50 fois, entre 50 et 878 ; cette dernière valeur est le nombre des éditions de « MacManX », dont la candidature comme administrateur avait été suggérée en décembre 2005, mais non acceptée.

5.5.3. La structuration temporelle des contributions

L'analyse de la structuration temporelle permet d'obtenir des résultats en matière de rythme de formation du Codex, d'intensité des contributions journalières et la mise en évidence d'un noyau permanent de contributeurs, présents, mais ne contribuant pas forcément. C'est là l'un des aspects de la communauté structurant les contributions.

Nous avons d'abord décompté le nombre de jours où *il y a effectivement contribution(s)* des contributeurs ayant effectué 10 éditions ou plus.

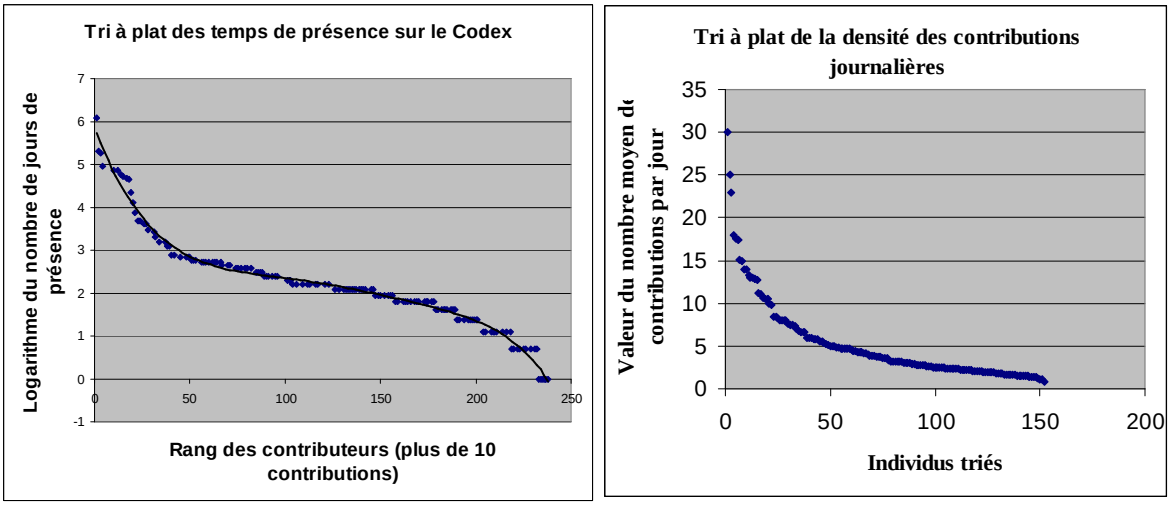


Figure 40 : tris à plat des temps de présence sur le Codex et de la densité des contributions

Le graphique de gauche ci-dessus montre que l'essentiel des contributeurs ayant effectué plus de 10 éditions se situe entre 5 et 20 jours de présence avec édition(s), points d'inflexions de la courbe (respectivement ln(5) et ln(20), en 1,6 et 3 sur le graphique avec échelle logarithmique)[492].

[492] Le point d'inflexion se caractérise par une droite tangente.

Le graphique de droite ci-dessus aligne les 150 plus gros contributeurs (plus de 10 éditions) en mettant en évidence cette fois le nombre moyen de contributions journalières (le ratio nombre d'éditions/nombre de jours d'éditions). On constate à partir de ce graphique que sur cette population, les deux tiers effectuent moins de 5 éditions en moyenne par jour, et qu'un autre tiers en fait entre 5 et 30. Sur ce point, nous avons encore des populations différentes.

Il est possible de raffiner encore ci-dessous, en prenant en compte cette fois la durée totale de présence de chaque individu, i.e. la durée entre sa première contribution et sa dernière, une variable *différente* de la précédente qui décrivait le nombre de jours où il y avait contribution *effective* au wiki. Pour donner un exemple concret, un individu peut contribuer deux jours, le premier janvier d'une année et le premier janvier de la suivante. Sa durée totale de présence (1 an) sera différente de celle d'un individu qui aura contribué autant de fois, le 1er et le 2e jour d'une seule année (2 jours de durée de présence). Cette variable est plus aisée à extraire manuellement, contrairement à la précédente, étant donné qu'il suffit de regarder les dates de débuts et de fin sur les historiques, et de calculer la différence.

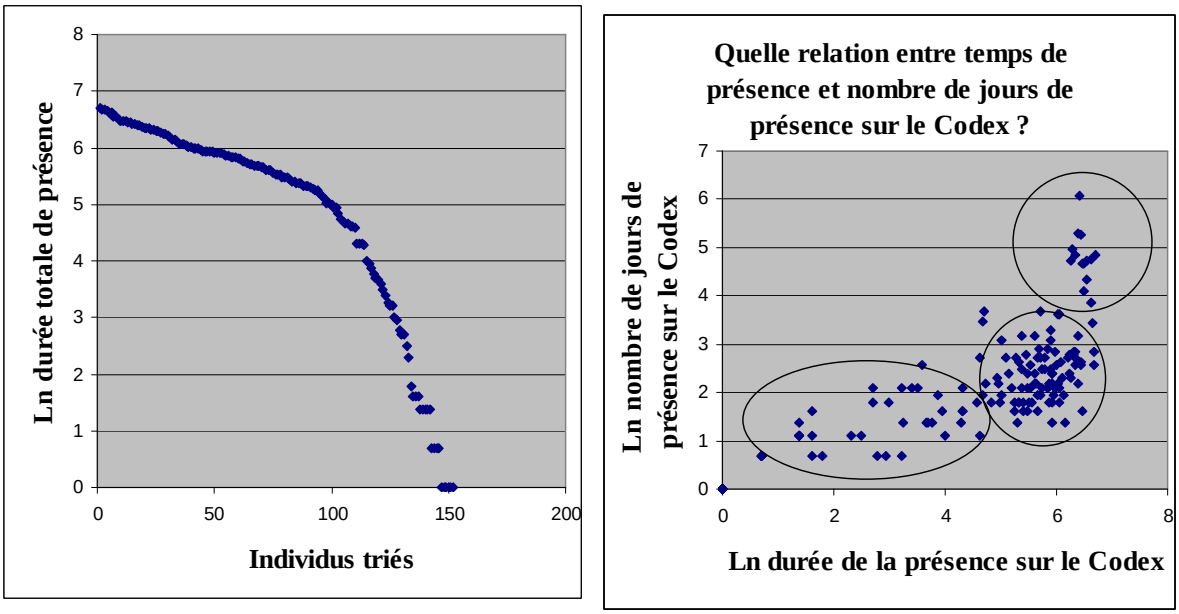


Figure 41 : tri à plat de la durée totale de présence et croisement avec la variable « temps de présence avec contributions ».

Le graphique de gauche permet de trier les individus par durée de présence décroissante. La distribution observée présentée avec une échelle logarithmique des valeurs

de durée est clairement différente de la première variable qui présentait le nombre de jours *avec* contributions. La distribution comporte une rupture en son centre, et non plus « aux deux ailes ».

Le graphique de droite permet de croiser les deux dernières variables, la durée totale de présence étant en abscisse, le nombre de jours où il y a contribution(s) en ordonnée. On observe ce faisant *trois populations différentes*, entourées par des cercles. La première à partir de la gauche correspond à des individus plus épars, qui contribuent sur des périodes pouvant aller jusqu'à 150 jours, et contribuent jusqu'à 20 jours en tout. La deuxième correspond à un noyau de contributeurs, visible par sa plus grande densité sur le graphique, qui contribuent sur des périodes allant de 150 jours à trois ans, et qui contribuent effectivement entre 5 et 40 jours. Enfin, un petit nombre de contributeurs, dont tous les administrateurs, a contribué sur des périodes de trois ans, toute la durée de vie du wiki lors du dépouillement, et contribue entre 40 et 400 jours.

Telles sont les caractéristiques temporelles des contributions au Codex.

L'une des originalités du dépouillement du nombre de jours où il y a contribution(s), ce qui n'a aucunement été effectué dans la littérature antérieure, est qu'il permet de dégager trois populations différentes en matière d'intensité temporelle des contributions au wiki, *et de montrer qu'il existe une population de contributeurs « permanents »*.

Ces caractéristiques de formation du bien collectif, plus temporelles, n'ont jamais été extraites à notre connaissance, le décomptage ne pouvant se faire avec un quelconque outil informatique : compter le nombre de jours avec contributions effectives est plus pénible manuellement et visuellement que celui du nombre d'éditions.

5.5.4. La structuration des discussions

A partir des mêmes sources que précédemment sur les contributeurs, on peut extraire des données caractérisant d'autres variables, comme le nombre d'éditions de discussions des contributeurs de la population. Les discussions sur le wiki sont significatives, car il s'agit aussi de contributions, qui ont un coût. Elles sont un élément de la formation d'un bien collectif du second ordre, le bien en question étant l'organisation de la formation du bien collectif du premier ordre (contributions simples). On observe des corrélations, mais uniquement à l'échelle logarithmique :

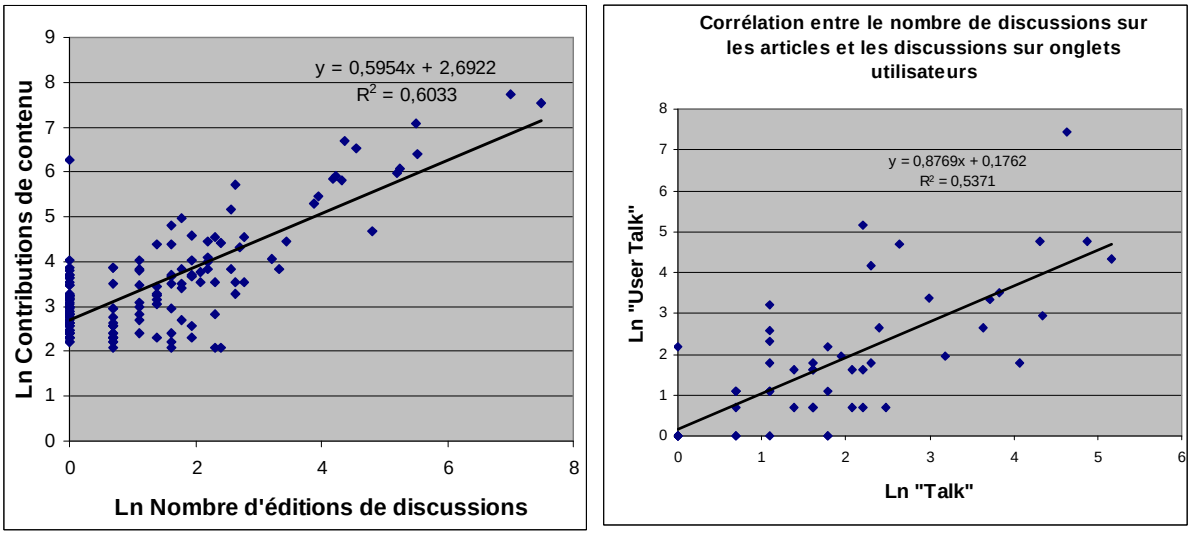


Figure 42 : corrélations entre les contributions et les discussions, et entre les types de discussions.

Le graphique de gauche montre que le nombre d'éditions de discussions, nul pour l'essentiel des contributeurs[493], tend à augmenter avec le nombre des contributions de ces mêmes contributeurs portant sur le contenu effectif du wiki. *Le graphique de droite* montre qu'il existe une relation structurelle entre les éditions de discussions via les articles (éditions « talk ») et le nombre d'éditions entre contributeurs non médiatées par des articles, directement sur les pages des contributeurs et entre contributeurs (« user : talk »). Les statistiques descriptives en pourcentage montrent que 16,36% des contributeurs effectuent des éditions « talk » et que 15,96% effectuent des éditions « user :talk ». Nous ne disposons pas du pourcentage de contributeurs qui effectuent des discussions, i.e. du pourcentage agrégé des deux pourcentages précédents. Les deux derniers chiffres donnent toutefois une idée du pourcentage total des éditions de discussions, qui ne doit pas concerner plus de 25% des contributeurs. Les administrateurs effectuent 72,21% des éditions « talk », et 83,95% des éditions « user :talk ». Nous ne pouvons dire si l'écart est significatif.

Les relations croissantes observées ne sont pas triviales, car les contributeurs bénévoles n'ont pas forcément envie de supporter un coût supplémentaire en matière de discussions. Cependant, elles montrent qu'il y a un lien entre contributions et régulations.

[493] Ils n'apparaissent pas ici, ln(0) n'étant pas défini.

5.5.5. Comparaison entre les éditions de contribution et les éditions de discussions

Nous effectuons à présent un tri à plat des deux variables de discussions précédentes, à côté desquelles nous plaçons aussi la variable « contributions sur les articles ». Il permet de comparer éditions de contribution et éditions de discussions, et de dégager la spécificité de ces dernières, qui sont une forme de régulation du bien public en question.

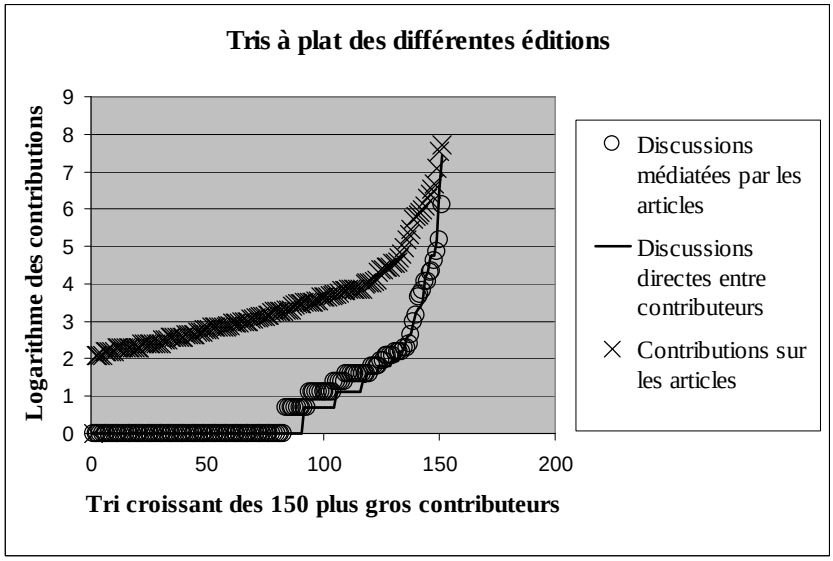


Figure 43: tris à plat des différents types d'éditions sur le wiki

Ce graphique synthétique permet de visualiser deux phénomènes *d'émergence*. D'une part, il montre que les contributions d'une petite partie des contributeurs de la documentation se différencient nettement à partir de 130 individus, ce qui pourrait être essentiel pour comprendre la viabilité du projet; il en est de même des discussions, où le seuil d'émergence est semblable, 140. D'autre part, on observe l'émergence des discussions chez des contributeurs de moindre importance plus tôt, à partir de 80 individus, sur le graphique. Enfin, la forme du tri à plat des éditions de discussions montre un phénomène de convergence, celles-ci rejoignent celles des éditions de contenu : il y a contact entre les courbes au point (150, 8).

Les variables ci-dessus sont corrélées, mais ce ne sont *pas forcément les mêmes individus* qui apparaissent au même point de l'abscisse en fonction de la variable étudiée, la corrélation n'étant pas parfaite, comme l'ont rappelé les graphiques ci-dessus.

5.5.6. Eléments sur la structuration du bien collectif à partir d'un échantillon d'articles

Nous abordons à présent la structuration des articles eux-mêmes. Nous avons pour cela effectué un sondage à partir d'un article sur quatre en avril 2006, en excluant les pages de management de la communauté, pour nous contenter d'étudier les pages comportant un contenu effectif, qui font partie d'un espace, dit « main space », rassemblant l'information « prête à usage » utile aux utilisateurs de WordPress. Cette différentiation lors des opérations de quantification sera une constante dans l'ensemble de ce travail. La coupe transversale pour les statistiques a été faite à un moment où le codex est arrivé à une certaine maturité, comme le montrent les données qualitatives, mais cela ne veut pas dire que le nombre d'article n'augmente plus. Nous en avons dépouillé 119, ce qui est une longue opération. Du fait de la pénibilité du dépouillement visuel, des erreurs de décompte dues au caractère faillible de l'observateur sont inévitables, mais nous supposons de façon tout à fait raisonnable qu'elles sont aléatoires. Nous dégageons tout d'abord quelques éléments structurels, en particulier le nombre d'intervenants de chaque catégorie par page. On observe des régularités, une courbe concave pour les administrateurs, une courbe convexe décroissante pour les autres :

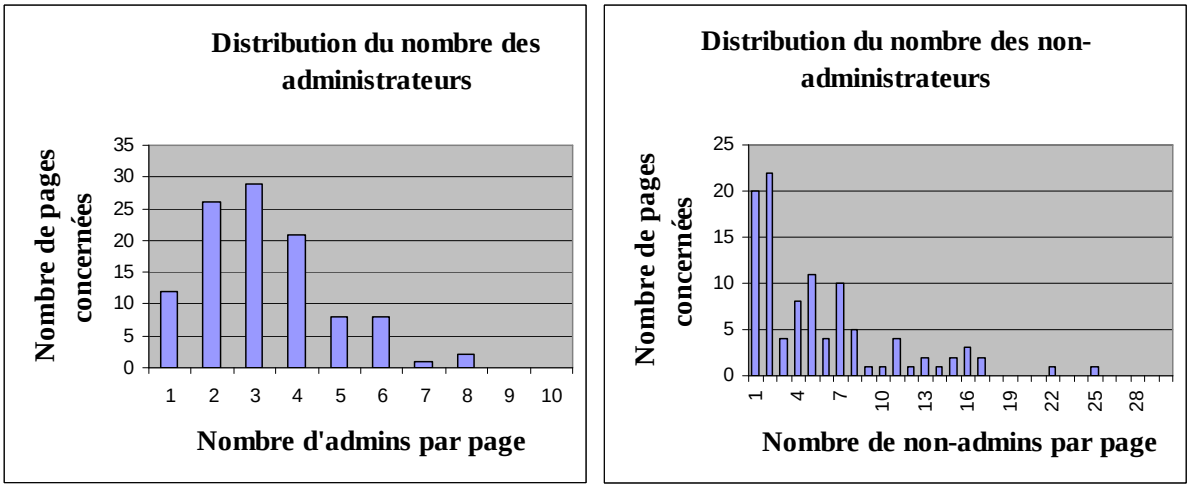


Figure 44 : effectifs des administrateurs et des non-administrateurs sur les pages du Codex

Ce résultat est intéressant, car il montre que l'essentiel des pages du wiki est constitué mi-2006 par au maximum 8 contributeurs non-administrateurs et 4 contributeurs administrateurs, ce qui montre *les limites* de la dimension « massivement distribuée » du processus de constitution du Codex. Par ailleurs, nous n'avons pu établir de corrélations entre les deux variables.

Ensuite, nous mettons en relation le nombre de contributeurs et le nombre d'éditions sur chaque article (graphique de gauche ci-dessous). On observe une nette corrélation sur ce point et donc une régularité structurelle soulignant sur ce point davantage que précédemment l'aspect distribué des éditions :

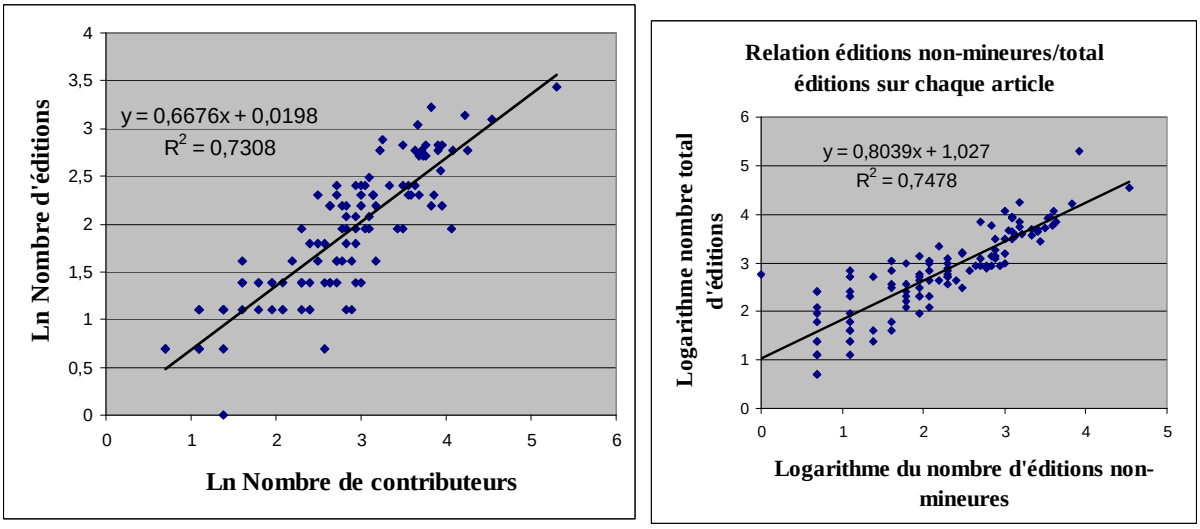


Figure 45 : corrélations entre le nombre de contributeurs et le nombre d'éditions, et entre les éditions non-mineures et le nombre total d'éditions

Le graphique de gauche montre que sur la population d'articles l'accroissement du nombre de contributeurs fait augmenter de façon régulière le nombre d'éditions. Le graphique de droite montre qu'il existe une relation structurelle entre le nombre d'éditions non-mineures et le nombre total d'éditions[494]. Ainsi, il s'agit de deux régularités dans la constitution du bien collectif. Le fait qu'il existe une relation linéaire entre le nombre de contributeurs et le nombre d'éditions montre qu'il existe une régularité dans le partage des tâches.

Sur les deux graphiques ci-dessous, nous avons différencié en éditions mineures et non-mineures. On voit que la relation précédente s'estompe, en particulier pour les éditions mineures.

[494] Les éditions « mineures », les autres, sont des éditions marginales de mise en forme de contenu existant, par exemple, corriger la forme ou l'orthographe. Le classement d'une édition comme mineure dépend cependant du jugement du contributeur en question et ne peut faire l'objet d'aucun contrôle.

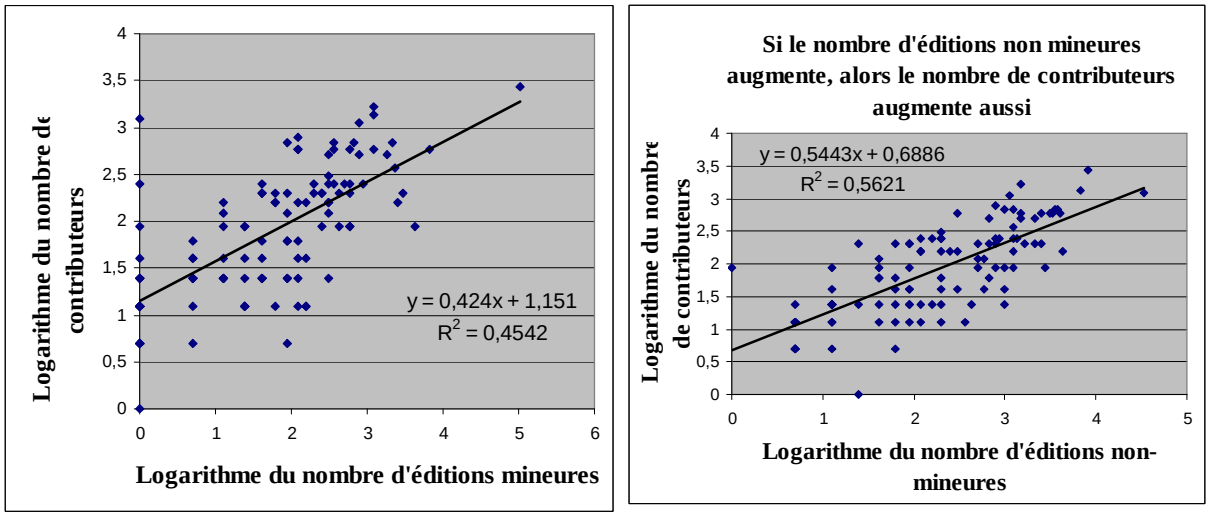


Figure 46 : corrélations entre le nombre de contributeurs et les éditions mineures et non-mineures

La structuration sémantique du Codex peut être estimée à partir de la liste des pages les plus catégorisées[495]. Ci-dessous, nous donnons le nombre de pages correspondant à chaque catégorie. Nous avons tiré ces données en juin 2009 ; nous avons présenté ci-dessus les catégories les plus fréquentes, c'est-à-dire celles qui regroupent plus de 50 pages du wiki. Nous avons donc ici les sections où on a le plus apporté au bien collectif. :

Functions	463
New page created	342
Advanced Topics	193
Getting Started	189
WordPress Development	169
Template Tags	160
Design and Layout	118
Stubs	116
WordPress Lessons	94
About WordPress	93
Troubleshooting	89
Meetings	86
Translating WordPress	81
Installation	70
WordPress Help	66
WPMU	66
Link	56
Plugins	55
Drafts	52

Tableau 2 : liste des catégories les plus fréquentes (plus de 50 pages, par ordre décroissant)

[495] Le principe est le même que précédemment : http://codex.wordpress.org/Special:MostLinkedCategories .

Il n'est pas nécessaire ici de détailler leurs significations[496]. On se rappellera seulement que les trois premières catégories qualifiant une page désignent des sections concernant les informaticiens : il s'agit de « functions », c'est-à-dire la description de fonctions du langage PhP, « advanced topics » (« sujets d'un niveau de difficulté avancé »), de même que « WordPress development ». Ensuite, l'essentiel des catégories concernent surtout les utilisateurs novices, la première étant « Getting started ».

5.6. Qui sont les contributeurs ?

Nous avons tiré et présenté, sur la base de données « utilisateurs » ayant permis la quantification précédente, précisément 100 fragments ayant un intérêt biographique sur les quelques 150 contributeurs ayant édité plus de 10 fois. Les contributeurs sont des blogueurs, des utilisateurs de WordPress, certains passionnés de blogs en possédant plusieurs.

L'écrasante majorité des contributeurs de cette population sont des informaticiens, en particuliers des réalisateurs de « plugins ». On trouve aussi des « web designers », et quelques personnes qu'on pourrait ranger parmi la catégorie des « littéraires ». Les participants du monde de l'open-source sont nombreux, mais on note aussi quelques professions commerciales de type consultant. De nombreux fragments ressemblent à celui-ci, qui est utile pour illustrer :

> "I'm a software engineer by day, airplane builder by night. I started using WordPress (well, actually I started using Nucleus, but WP is far better) to document my progress on building an airplane. I wrote a few simple plugins and noticed that the plugin API documentation has a lot of holes that need filling.
> As I have time, I'd like to add some info on the filter hooks on the Plugin API page."[497]

Ces éléments montrent que la problématique de résolution du bien collectif qui consiste à faire intervenir sur le wiki des développeurs[498], a été traitée dans une mesure significative. Cela peut aussi montrer qu'une expertise significative est nécessaire avant de contribuer au wiki. D'autre part, cela montre qu'il existerait un lien entre contribution à la communauté par des plugins et contribution à la communauté par de la documentation.

Dans une moindre mesure, apparemment, des contributeurs au forum en viennent aussi à contribuer à la documentation. Ces contributeurs à la documentation peuvent recouper la

[496] Nous renvoyons au glossaire pour celles qui ne tombent pas sous le sens.
[497] Source : http://codex.wordpress.org/User:Davepar .
[498] Il ne s'agit donc pas forcément des utilisateurs avancés ne connaissant pourtant rien à l'informatique, ce que laissait pourtant entendre les archives de la liste de discussion de la documentation.

population précédente. Le lien entre contribution à l'un et l'autre des deux SI a été repéré à plusieurs reprises. Nous mentionnons ce lien pour l'administrateur « NuclearMoose »:

> "My name is Craig Hartel. I live in Dawson Creek, BC, Canada, and I have been using WordPress since late in 2003. My first upgrade was to Version 0.72. I decided to contribute to WP through supporting users on the forums and by pushing the documentation effort forward. You can visit my personal WP-powered web site, [http://blog.nuclearmoose.com/index.php nuclear moose candy]."[499]

Sur ce point, une démarche de quantification serait nécessaire. Dans quelle mesure les populations de contributeurs se recoupent-elles ? Observe-t-on dans ce cas une corrélation entre les contributions à l'un et l'autre des systèmes d'information ? Si cela est observé de façon exploratoire, d'autres dépouillements, en particulier du nombre de contributions individuelles au forum, permettraient des précisions plus systématiques. Cela est faisable et n'est guère complexe, mais demanderait de disposer d'un temps significatif pour contrôler les identités ou identifiants et décompter.

Signalons enfin une particularité surprenante, par comparaison avec ce que nous avions mentionné dans la précédente revue de littérature sur les wikis: il n'y a guère de femmes parmi les contributeurs. Nous mentionnons quelques noms : Lorelle, Cena, « Geeksmakesmehot » (sic). En tout cas, si le nombre de femmes est réduit, rien n'indique que les femmes subissent des formes de désincitations à la participation, ce qui peut s'observer dans les communautés open-source très masculines. Doit-on en déduire que la communauté a réussi à solliciter pour la documentation des informaticiens habituellement masculins ? Cela est vraisemblable et il s'agirait d'un indice du bon fonctionnement de la formation du bien collectif.

5.7. Les motivations des contributeurs d'après les données qualitatives

Quelles sont les motivations des participants que nous avons pu relever à partir des données textuelles disponibles ? Les données en question sont les informations qu'ils mentionnent sur leurs pages utilisateurs et d'autres informations que nous relevons sur la liste de discussion de la documentation. Les données étant à l'état de traces, les croiser de façon complémentaire pour avoir une idée satisfaisante des motivations est indispensable. Il est possible de quantifier les fragments en comptant ceux relevant de telle ou telle catégorie, mais

[499] Source : http://codex.wordpress.org/User:NuclearMoose .

cela n'autoriserait pas forcément des conclusions supplémentaires définitives. Il vaut mieux travailler pour le moment de façon qualitative pour comprendre la diversité de la thématique.

5.7.1. Du désir d'aider à une pratique habituelle de don contre don

La générosité, le désir et le goût d'aider ("help"), l'idée de contre-don en retour de l'utilisation du logiciel sont des motivations certes attendues, à plusieurs reprises exprimées. Nous mentionnons ici plusieurs fragments qui aident à comprendre le caractère significatif du phénomène altruiste chez des contributeurs *différents* en matière de qualifications. Nous en citons trois:

> "I am a graphic designer and pixel artist. I have been designing for about 12 years now. I have some knowledge in PHP, not enought to write my own "Wordpress", but enough to modify it to my own needs. I have been using Wordpress since it's first stable release, so I hopefully can help someone in need of it. I enjoy helping people in any possible way that I can."[500]
>
> "Hi all, just another WP user here (who likes helping out with documentation wink). If you'd like to see it, here's my mostly default blog: Eugh Andy..."[501]
>
> "Morgan Doocy lives at http://doocy.net/, and loves to use his leet writoring n editoring skillz to help others out. He's happy to contribute here as much as he can, 'cuz he loves WordPress. Word."[502] (sic!)

Celui-ci, utilisateur à plusieurs reprises du blog WordPress, veut aider afin explicitement d'inciter d'autres à utiliser le logiciel:

> "Trying hard to use WordPress. Trying my derndest. Running into some snags along the way. Just want to contribute to Codex, in case it can help others who follow same roads as me. I'm using wordpress in a number of places, for example: http://japanese.nearlythere.com/ - Heather"[503]

Plusieurs fragments montrent explicitement qu'il s'agit non plus d'aider, mais spécifiquement de « rendre à la communauté » :

> "I decided to become involved with the WP documentation effort in order to give something to the community. I have done a lot of QA work on technical articles, and I have plenty of tools and bandwidth to get the job done!"[504]
>
> "I am a little tired of using the wiki as a resource for plugin reference and downloads only to to find out-of-date URLs. I'd like to know how to become and editor so that I can do my bit for WP.org- the least I can do for using this wonderful blogging software for all these years. Thanks, Carlo »[505]

---

[500] Source : http://codex.wordpress.org/User:Ensellitis .
[501] Source : http://codex.wordpress.org/User:Kobold .
[502] Source : http://codex.wordpress.org/User:morganiq .
[503] Source : http://codex.wordpress.org/User:Nearlythere .
[504] Source : http://comox.textdrive.com/pipermail/docs/2003-November/000009.html .
[505] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-January/001590.html .

La générosité est un phénomène dont les membres de la communauté ont conscience, comme celui-ci, qui rappelle au passage que l'outil wiki lui-même est conçu pour rendre les contributions les plus aisées possible:

> "People have energy, and are eager to contribute. Do it. The wiki makes it supremely easy to cut-and-paste, move, migrate, manipulate and mangle! Use the strengths of the tool as they are."[506]

En conséquence, aider et rendre devient une pratique institutionnalisée, comme si le contre-don était une pratique normale: "...what you give is what you get!"[507]. Lorsque des contributeurs « signent » pour participer à un « wiki week-end » de contributions intensives, le caractère général et normalisé des contributions devient une information partagée : « You got involved to give back to WordPress »[508] On peut se demander s'il s'agit d'une forme d'instrumentalisation du don-contre don, ou alors de l'explicitation par ceux qui sont au fait de cette pratique d'un phénomène naturellement régulier.

A cet égard, on peut se poser justement la question de *l'origine* d'une telle habitude. Les données qualitatives permettent sur ce point d'en savoir plus. Les principales références « institutionnelles » soulignent la permanence de l'idéologie de l'open-source, et moins celle de Wikipedia. De nombreux fragments de présentations personnelles ressemblent sur ce point à celui-ci, qui montre une implication dans plusieurs projets :

> "switched to WordPress from a home-made blogging system in May 2004. […] I'm involved with a couple of other open source projects:
>
> * jEdit - as a plugin author (HeadlinePlugin, an RSS/RDF/Atom reader)
> * JCap - primary developer
> * I'm also building a PHP photo gallery application that I'll open source soon.
>
> I've written a plugin for WordPress 1.2 that, among other things, shows you several readability indexes of your posts: Word Statistics Plugin"[509]

Pour Wikipedia, voici le seul que nous ayons trouvé, à part la présentation de Carthik que nous mentionnerons immédiatement un peu plus loin :

> "After working on wikipedia for a short while, I've a few thoughts and comments on how to make codex more useful and functional.
>
> * Bring conversation about the wiki primarily onto the wiki and off wp-docs. This increases wiki-traffic and creates an internal record of development.
> * Optionally Limit search to codex when in codex, forum threads are less than useful much of the time, and should be incorporated into codex anyway. Plus, it drives users back off of codex, pushing down editting.
> * Easy search for a file like version.php etc and arrive at an article detailing it's function?

---

[506] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000667.html .
[507] Source : http://codex.wordpress.org/User:Olaf .
[508] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000581.html .
[509] Source : http://codex.wordpress.org/User:JohnWatson .

> * List of variables (such as $wp_version) to use in building templates?
> * etc etc. » (retrouver référence)

Cependant, les données sont trop lacunaires ici pour qu'on puisse tirer des conclusions définitives sur ce point.

On remarque par ailleurs en cohérence avec ce qui précède que le don peut aller jusqu'au désintéressement total, en particulier financier. Voici ce qu'un administrateur, doctorant et donc a priori guère aisé sur le plan financier, dit de son travail conséquent sur le Codex, lors de discussion où il est question qu'un opportuniste écrive un livre à partir du travail des autres : « If someone makes a profit out of my work, so be it. If the person were a "good man", he/she'd ask me before doing it, and definitely credit me ;)"[510].

5.7.2. De l'amusement à l'addiction

Aussi curieux que cela puisse paraître, l'addiction est l'une des motivations des contributeurs[511]. Il s'agit d'abord d'un goût prononcé pour les wikis, ce qui peut expliquer les contributions très fortes de l'un des premiers administrateurs, Carthik :

> "I am a Ph.D. student at the UCF, and I find Wikis amazing tools. I love WordPress, and spend a few hours on WordPress each week. As for wikis, I have been a member of Wikipedia since Wednesday, June 30, 2004."[512]

Cet administrateur n'est pas le seul dans ce cas. Lors de la sortie du premier wiki, nous avons trouvé le commentaire suivant : "Awesome..can't wait to play with it."[513] . On observe donc des formes de plaisir associées à l'outil, d'ailleurs explicitées dans les « Guidelines » (« have fun »), mais il ne s'agit pas de ce que mentionnaient LERNER et TIROLE (2002) à propos du code en utilisant le même mot. Il s'agissait alors du plaisir de résoudre des énigmes informatiques. Ailleurs, un goût pour un thème donné demandant complément d'informations qui peut conduire à contribuer: "Hi, I am Olle Jonsson of Copenhagen, Denmark. I like developer topics here at the wiki. So I edit those. I write a personal weblog.[514]" . Cependant, tout le monde dans la communauté n'aime pas forcément ces outils, ce qui limite cette motivation.

---

[510] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000851.html .
[511] Nous rappelons au passage que l'addiction est une notion économique (utilisée en économie de la drogue par exemple), où le consommateur est prêt à payer très cher un bien « indispensable », pour l'obtenir.
[512] Source : http://codex.wordpress.org/User:Carthik .
[513] Source : http://comox.textdrive.com/pipermail/docs/2003-December/000074.html .
[514] Source : http://codex.wordpress.org/User:Olleolleolle .

Carthik, mentionné précédemment, a fini par passer du goût à l'addiction: « I must say it is the most useful thing already, and it's taking over my life already!"[515]. Un autre administrateur notable, Lorelle Van Fossen (« relle » ci-dessous), est sujette à la même motivation. Nous citons un passage amusant, pour une fois, où ses contributions massives sont commentées par deux autres administrateurs :

> « Jun 29 16:44:07 <relle> EEEEEEEK
>
> Jun 29 16:44:08 <photomatt> they can be written collaboratively on the codex
>
> Jun 29 16:44:12 <mdawaffe> someone will have to stop relle from writing a novel each week ;)"[516]

L'addiction expliquerait les effets d'entraînement parmi les contributeurs notables, parce qu'elle serait contagieuse. Il s'agit toujours ci-dessous de la même personne:

> "It went on the list because I was told to put it here a long time ago. Now we have people who are helping, addicted, and the Codex is looking better and more consistent than ever. Lorelle.»[517]

Ainsi, le phénomène en question est confirmé par plusieurs sources qualitatives, sur des points décisifs, à savoir les meilleurs contributeurs. Mais il n'est *par définition* valable que pour eux, pas pour ceux qui ont contribué aux alentours de dix fois. L'addiction est une motivation également observée sur Wikipedia.

5.7.3. Purisme et professionnalisme: le désir du travail bien fait pour lui-même

Le désir de professionnalisme en matière de documentation est l'un des facteurs de motivation à la contribution et à l'amélioration du wiki. Nous tentons une analyse sur ce point. Il s'agit d'abord d'un « amour de la clarté » en matière de documentation :

> "I love WordPress, and I love clear documentation! I use WP to power two sites: Chi Alpha @ Stanford and Glen & Paula Davis. Also, I made one of the Scripturizer plugins."[518]
>
> « Being a programmer who loves clear documentation »[519]

L'effet de ce type de motivation est conséquent, car deux des *administrateurs* au moins, Kafkaesqui et Carthik, sont des puristes qui ne supportent pas les formes d'écriture imparfaite, et aiment « corriger ». Nous mentionnons leurs pages utilisateurs:

> "I'm also often caught in the WordPress support forum, passing on the occasional helpful nugget. And I apparently volunteered to help out moderating on Codex, which involves writing and editing some of

[515] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000537.html .
[516] Source : http://codex.wordpress.org/IRC_Meetups/2005/June/June29RawLog .
[517] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000656.html .
[518] Source : http://codex.wordpress.org/User:Glen_Davis .
[519] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-January/001409.html .

the documentation found on this here wiki. A glutton for tiny needles of thankless activity, that's what I am."[520]

"I'm a Ph.D student in the U.S. The important things to take away from that is that I have odd hours and that I spell things with rigor, not rigour (though I make a lot of typos). I may edit something of yours and, not maliciously, but "unconciously", change your rationalisations into rationalizations. I mean no offense. Please change them back if you prefer."[521]

Par ailleurs, sur le forum utilisateur, nous avons extrait un « post » de l'administrateur Kafkaesqui, également contributeur respecté au forum. Il s'agit clairement d'un « incident critique », une réponse d'une personne vexée (« I take it personally ») à un utilisateur qui avait posté un commentaire soulignant le caractère inutilisable de la documentation. Il s'agit d'un passage insistant, un rappel à l'ordre, qui a le mérite de mettre en perspective de façon *économique* le contexte du caractère gratuit des contributions associées à de telles motivations. Ce post sur le forum mérite d'être mentionné en entier car les propos font système :

"We're well aware of the reasons for hiring people skilled in writing technical documentation. The problem with this would be that:

1. WordPress is an open source, **free** software program. Not a lot of cash floating around to hire these "professionals" we're talking about.

2. Paid technical writers also often need to first learn and/or be trained on the software. This also costs money, again typically not a resource in vast supply for a free application.

3. Many people already involved with writing the documentation are not merely conversant but also well versed with "the field" of WordPress.

Literally thousands of man (and woman!) hours have gone into the "official" documentation for WordPress; not to forget all the various material found in the wild. Each of those hours was taken on by a volunteer. As one of those volunteers, let me be clear that when someone labels it as "truly not usable," I take it personally.

If you think the documentation can be better, tell us how. Constructive criticism with the goal of improving things is something we can all appreciate. Better yet, if you know how it can be improved *and* have the skills to make it so, get involved and start adding to or editing what's there. That's how we've all done it. But telling us "Hire someone, because your work is awful" is not a solution.

Finally, keep in mind that at least a few people writing the WordPress documentation are or have been in the past professional technical writers. And there's nothing they want to hear less than someone telling them they suck at it."[522]

Kafkaesqui, dont les propos sont inhabituellement longs et martelés (1/, 2/, 3/, …), a été vexé des critiques non constructives à l'égard de la documentation ; il a d'ailleurs fermé le fil à des

[520] Source : http://codex.wordpress.org/User:Kafkaesqui .
[521] Source : http://codex.wordpress.org/index.php?title=User:MDAWaffe&diff=next&oldid=6407 .
[522] Source : http://wordpress.org/support/topic/146799 .

remarques supplémentaires. On le comprend lorsqu'on se rappelle ses motivations évoquées précédemment. Il montre l'avantage économique des contributions gratuites, étant donné que le recrutement de professionnels de la documentation occasionnerait un coût d'apprentissage supplémentaire. Il montre et confirme qu'un petit noyau significatif de contributeurs sont des *professionnels* de la documentation, qui sont motivés par le travail bien fait, et, qu'au-delà, la galaxie des volontaires grands ou petits tiennent à la qualité de leur travail[523].

5.7.4. Quelques contributions "au passage"

Pour les contributeurs modestes, dont les interventions, entre dix et trente contributions approximativement sont les plus courantes, il s'agit de contributions « au passage ». C'est le cas de nombreux contributeurs, qui semblent ne pas avoir le temps et qui décident de n'apporter ponctuellement que quelques bribes de documentation, ce qui fait cependant qu'ils peuvent arriver rapidement à plus de 10 éditions. Nous donnons un exemple illustratif de cette classe de contributeurs:

> "So here I am, designing and tweaking code all day long. Then one day I find blogs. Then a few weeks later I find WordPress. Thank you. What a great piece of code, I think I have found my home. I started designing themes and before you know it I have a large blog and several other blogs I either work on or play on.
>
> I have updated all the necessary details on the Themes page so now all the CC/GNU GPL licensing data is on there for everyone to see."[524].

Celui-ci a donc brièvement contribué en mettant ponctuellement à jour des informations. De nombreux contributeurs de ce type sont des auteurs de plugins. Nous faisons cette remarque de façon exploratoire sur notre base de données, elle serait aisée à cerner de façon quantitative.

5.7.5. Contribuer à un projet qui fascine ?

Une autre motivation observée est la fascination pour le projet, le projet de documentation, et pas seulement le logiciel lui-même :

---

[523] Avec le succès de la start-up et la quantité de travail associée au Codex, l'idée de recruter des professionnels de la documentation reviendra sur le devant (source : http://comox.textdrive.com/pipermail/wp-docs/2008-May/001737.html), mais sans jamais que la chose se concrétise (un tel « séisme » aurait été observé dans les archives récentes de 2009 par exemple).

[524] Source : http://codex.wordpress.org/User: Jbbrwcky .

> "I'm new to the world of WordPress and MediaWiki and am just awe-struck by the contributions made by all the talented people. If I can contribute just one iota of what they've contributed… WordPress user at [http://hancockdata.com Hancock Data]."[525]

Ce dernier administrateur est fort intéressant étant donné qu'il occupe un poste significatif dans la société Hancock Data, ce qui est a priori fort prenant. Pourtant, *ce fut par la suite l'un des contributeurs les plus significatifs au wiki*, y compris dans ses tâches les plus humbles. *On voit donc que l'admiration est une motivation potentiellement très puissante*.

Cependant, si la fascination est partagée, elle ne donne pas forcément des contributions concrètes, comme le montre le passage suivant, qui interpelle directement la notion. L'auteur des contributions a été fasciné, mais il n'a pas été bien loin, c'est-à-dire pas au-delà de ce qui était urgent :

> "I find the WordPress Codex a fascinating project, one that ideally I should be actively contributing to but am not. My only contribution has been the WordPress Quick Start Guide page, http://codex.wordpress.org/WordPress_Quick_Start_Guide, which I felt was sorely needed."[526]

C'est aussi le cas de celui-ci :

> "I've been surfing the 'Net since 1996 when I chugged along with my 33.6 modem. I've had a love of creating webpages since that time, and I purchased my domain name http://www.webdiva.co.uk in 1998. I spent quite a considerable amount of time editing for Dmoz and various other human edited directories and so appreciate the passion and commitment that such a venture as the Codex requires and demands.
>
> I have a users knowledge of Wordpress and believe that my experiences as a user can enrich the Codex, especially for those with little technical know-how. To begin with I would like to begin checking documents for errors and slowly build myself up to more involved documents, and perhaps write content."[527]

Pour résumer, cette motivation est potentiellement puissante, mais pas forcément sûre.

5.7.6. Quelques éléments de motivation intéressée

Peut-on observer des motivations intéressées ? Sur ce point, le dépouillement des sources qualitatives ne permet guère de faire de découvertes. Quelques fragments seulement font allusion à ce type de motivations, et ils sont plutôt flous. *Cependant, leur absence ne permet pas de tirer des conclusions définitives*. En effet, dans un projet formé de volontaires, on comprend qu'expliciter des motivations intéressées n'est pas la première idée qui viendrait

[525] Source : http://codex.wordpress.org/index.php?title=User%3AMichaelH&diff=11281&oldid=11280 .
[526] Source : http://comox.textdrive.com/pipermail/wp-docs/2009-January/001876.html .
[527] Source : http://codex.wordpress.org/User:Webdivauk .

à l'esprit de contributeurs à un tel projet. Cependant, une telle absence est aussi cohérente avec les motivations désintéressées qui précèdent. Sur ce point, nous ne pouvons juger définitivement. Dans quelques passages certains tiennent à montrer leurs compétences :

> "While I'm relatively new to WP (first installed a little less than 2 weeks ago), I'm an intermediate to advanced PHP / MySQL coder and consider myself strong in developing end-user documentation. I've already dabbled in creating several plugins and hacks and am becoming increasingly comfortable with what lies beneath the hood of WP."[528]
>
> "If there is a need for documentation that falls into a realm of my expertise, I may pick it up. That's why I'm subscribed."[529]

Un autre passage mentionne des motivations égoïstes en matière de contributions à la documentation développeurs, mais elles ne sont pas claires: « I also had selfish reasons for my involvement with the inline documentation and for it appears the same reasons you have for working on the codex."[530]

5.8. Conclusions, limites, et recherches ultérieures

Nous avons ici obtenu des *résultats* sur plusieurs points : la communauté avait réussi, à la date des dépouillements, à résoudre de façon satisfaisante la problématique de bien collectif qui consistait à contribuer au wiki. Il s'agissait pourtant d'un problème de gestion non trivial. La réussite du projet viendrait d'un noyau significatif et stable de contributeurs qui forme l'essentiel des contributions, discute l'organisation du wiki, travaille en continu. Cependant, d'autres contributeurs donnent plus de 40% du contenu tout en ne prenant pas part à l'organisation. Nous avons donné un tableau exploratoire des motivations, qui sont diverses (plaisir parfois addictif, aide et don contre-don, professionnalisme, participation à un projet prestigieux)[531].

Sur le plan *méthodologique*, nous apportons aussi une contribution en matière d'approche archéologique, qualitative à partir de « traces », associée à des dépouillements quantitatifs significatifs (ce que font aussi les historiens).

Ce travail, au caractère exploratoire sur certains points comporte des limites. Un seul chapitre était peut-être un cadre trop court pour tout traiter. La méthodologie historique, à

[528] Source : http://comox.textdrive.com/pipermail/docs/2004-October/000773.html .
[529] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000885.html .
[530] Source: http://comox.textdrive.com/pipermail/wp-docs/2008-November/001782.html .
[531] On note donc ici un écart par rapport à la règle des 80-20%.

partir de traces lacunaires a pu avoir un rôle contraignant, et trouverait alors ses limites par comparaison avec d'autres approches.

La littérature rassemblée dans la revue précédente, peut-être plus solide du fait de son aspect quantitatif, est arrivée après nos dépouillements, qui datent de l'année 2006, ce qui fait que notre travail, en plus de son caractère exploratoire, est peut-être à présent daté. Mais c'était impossible à anticiper.

Le travail à partir des données de sondage comporte certaines limites, que nous n'avons pas su résoudre : la distribution des articles suit une loi de Pareto (voir début de ce chapitre), ce que nous ne savions pas, l'outil permettant de décompter les articles pour obtenir cette distribution n'existant pas. Or, la théorie des sondages normale fonctionne avec des lois normales. Nous nous sommes aperçus de cela hélas trop tard pour changer ces données. Il est possible que cela pose des problèmes de validité de nos conclusions.

Ensuite, nous avons identifié par extraction de plusieurs variables des populations différentes par leurs comportements, à chaque fois. Cela permet de comprendre les *spécificités* du phénomène de contribution au bien collectif. Mais nous n'avons pas recoupé de façon approfondie ces différentes populations en *une base de données unique*, principalement par manque de temps et par suite des tâtonnements de notre recherche. Nous avions en effet prévu de tenter d'étudier davantage la population des contributeurs, en particulier à partir des historiques des contributions. Mais le travail de dépouillement aurait été démesuré sans outil informatique et nous n'avons pu trouver une méthodologie satisfaisante pour cela. Nous avons pourtant travaillé cette question sans résultat probant. D'autres travaux sur les contributeurs restent à faire et notre travail est exploratoire sur ce point. Il faudrait en savoir davantage sur les contributeurs plus mineurs, car ils peuvent être facteur de diversité informationnelle même s'ils sont moins significatifs quantitativement. Il serait nécessaire d'effectuer davantage de croisements de données entre les contributions aux différentes parties de la communauté (le forum en particulier) et les motivations. Dans une typologie, il faut mettre en relation les informations venant de diverses sources : liste de régulation de la documentation, forum, historiques des wikis, informations personnelles, contribution au code, etc. Une voie de recherche possible serait d'effectuer des typologies des contributeurs, mais leur utilité marginale, en gestion du moins, ne serait guère significative par comparaison avec ce que nous avons fait, et elles supposeraient des données plus conséquentes (par des questionnaires plus détaillés). En particulier, il faudrait tester davantage l'idée qu'il y aurait un « effet système » : les contributeurs au Codex contribuent parce qu'ils ont déjà contribué pour les plugins, ou parce qu'ils ont déjà contribué sur le forum. Ayant déjà

contribué ailleurs, ils peuvent contribuer à la marge dans le wiki, au passage. L'accumulation des actions de cette galaxie de contributeurs mineurs, selon le modèle de la « soupe aux pierre », peut être un aspect crucial de la formation de ce bien collectif.

Un autre aspect de la question est que les motivations ne sont pas « statiques », contrairement à ce que décrivent les modèles d'incitations. Elles évoluent au cours du temps (FREEMAN, 2007), en fonction de l'apprentissage des contributeurs, du contexte, des relations établies, des conventions… Malheureusement, la nature des données par traces ne permettrait sans doute pas de décrire de telles dynamiques, même en datant leur énonciation. Il faudrait aussi composer une liste de l'évolution des motivations en fonction du degré de contributions à telle ou telle date. C'est la raison pour laquelle commencer par les plus gros contributeurs donne au moins les éléments de motivations les plus significatifs et décisifs.

L'étude plus approfondie des motivations en elles-mêmes demanderait des compléments d'enquête, par des questionnaires soumis par voie électronique par exemple. Il n'est pas exclu de construire *ultérieurement* ceux-ci à partir des éléments exploratoires dégagés ici. Il est possible d'explorer d'autres projets dans le domaine de l'Open-Source pour confirmer nos premières conclusions et « saturer » le spectre des motivations. Etudier l'aspect dynamique des motivations supposerait d'étudier les dynamiques de foule, les contributions des uns et des autres n'étant pas forcément indépendantes. Nous étudierons les aspects dynamiques de la collaboration ultérieurement au chapitre 7, certes, mais le lien avec les motivations ne sera pas abordé ; nous ne savons comment faire à partir de données lacunaires.

Après l'étude des caractéristiques de la formation collective du Codex, nous allons examiner dans le prochain chapitre sous une autre perspective les régularités qui la conditionnent.

Chapitre 6 : l'émergence de routines hybrides

6.1. Introduction

La présente sous-question de recherche peut s'énoncer ainsi : comment comprendre les routines qui ont émergé durant l'histoire de la mise en place de la documentation ?

Nous nous focaliserons ici sur certaines caractéristiques structurelles, mais pas forcément saillantes, de « l'embryogénèse » du wiki (pour employer une image parlante). Nous allons documenter les routines émergentes observées et dans ce cadre *examiner dans quelle mesure on observe une hybridation entre des caractéristiques organisationnelles de l'open-source et de l'open-media.*

Nous proposons, pour dire cela autrement, une (nécessaire) analyse institutionnaliste de la régulation informelle de la communauté WordPress, ce que permettent les apports de l'économie évolutionniste et de l'économie des conventions. Les deux approches théoriques mobilisées pour analyser les routines dans notre communauté ont émergé parallèlement de nos lectures et de l'analyse inductive. Une problématique d'économie cognitive par excellence est en effet l'analyse des problèmes d'émergence, de mutation et diffusion des routines. Elle est incontournable pour aborder les problèmes de régulation et gouvernance informelles de la communauté travaillant sur la documentation. L'apport d'une démarche historique est ici utile pour observer et analyser les formes d'adaptation des pratiques, les formes d'exploration/exploitation individuelles et collectives, les solutions adoptées, celles qui auraient pu l'être.

Nous préciserons tout d'abord davantage notre démarche méthodologique compréhensive. Nous approfondirons au préalable dans quel cadre institutionnel juridique ces questions de pratiques se posent et les raisons de l'émergence de cet aspect de la gouvernance. Nous allons ensuite développer l'analyse des routines en les regroupant en deux types, d'une part les routines de structuration de contenu, d'autre part les routines de gouvernance de la population de contributeurs.

6.2. Comment étudier les routines sur notre terrain si particulier?

6.2.1. Justification de l'usage du concept de routine

6.2.1.1. Une thématique a priori paradoxale, d'où quelques exemples concrets

Parler de routines pour un wiki peut sembler paradoxal, étant donné qu'il s'agit d'un projet et d'un ensemble d'actions organisantes marqués par le sceau de l'éphémère. Lorsqu'il est question de wiki, on pense bien souvent à liberté, créativité, voir anarchie. Routine signifie au contraire que nous pouvons observer des régularités significatives de comportement entre individus. Retourner un argument ou rendre statique ce qui est changeant de cette façon est d'ailleurs une approche méthodologique judicieuse de DUMEZ (2009).

Les routines observées d'importance peuvent appartenir à plusieurs registres : modes d'édition des articles, formes de management (sociabilisation, motivations, leadership/autorité, …). Voici un fragment pertinent d'un courriel de Lorelle, l'une des administratrices ("admin") du wiki et l'une des contributrices les plus reconnues à la documentation, où cette dernière tente d'éclairer un contributeur intéressé par l'organisation de la documentation :

> "Thank you for your effort but before you go too far, please see:
> http://codex.wordpress.org/Codex:Guidelines
> http://codex.wordpress.org/Codex:Contributing
> http://codex.wordpress.org/Codex:Maintenance
> http://codex.wordpress.org/Codex:Styles
> As a guide to what is already in place.
> As an organic manual, these have evolved over time."[532]

L'intérêt de ce fragment[533], datant du 7 mars 2008, est rétrospectif : depuis août 2004, date des premières contributions au Codex fonctionnant sur Mediawiki, des « guidelines » ont donc été proposées aux volontaires, écrits par des volontaires, et le fragment en question rappelle que les pratiques collectives, plus ou moins acceptées, ont été sujettes à variations. Un autre courriel rétrospectif de Lorelle[534] datant de décembre 2008 est pertinent pour

[532] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-March/001634.html .
[533] Certains de ces liens sont d'ailleurs copiés-collés par les gros contributeurs sur les onglets de discussions des nouveaux collaborateurs pour qu'ils apprennent et ne mettent pas trop de désordre.
[534] Outre l'aspect rétrospectif de la chose, il faut se rappeler que Lorelle est l'un des administrateurs les plus respectés, et que de tels fragments sont de poids.

compléter l'idée, déjà apparente dans le précédent fragment, *qu'il s'agit de routines, et non de règles*[535] :

> "But don't let that stop you. Dig in. Do what you can. Use your best judgment. There are some core standards on how the Codex works, but all is fixable. Just read through the core Codex guidelines (no "I" stories, link to all panel keywords, etc.) and you should be fine. Let us know which articles you have worked on at first, and we'll double check them for you if you want."[536]

En effet, les deux fragments montrent clairement qu'il n'y a pas de sanctions associées à ces « standards », et qu'ils sont évolutifs, moins rigides que des règles, et soumis à une liberté de jugement. D'où l'intérêt de la notion de routine et de son étude, non pas seulement à partir de ces pages de régulation de la communauté, mais aussi au niveau des discussions au fil des événements. En effet, toutes les pratiques n'apparaissent pas forcément dans les « guidelines » en question, ce qui est attendu dans une organisation informelle : elles peuvent comporter un aspect *tacite*, cependant archivé dans la mémoire collective de la communauté.

6.2.1.2. Définition du concept de routine parmi les définitions possibles et articulation avec d'autres concepts

Le concept de routine a une longue histoire dans plusieurs disciplines, mais il a trouvé une spécificité en économie évolutionniste. Nelson et Winter les caractérisaient par trois principes (pertinents ici): hérédité et permanence, principe de variation, principe de sélection. La définition du concept a fait l'objet de débats, mais les économistes ont fini par se mettre d'accord sur une définition minimale lors de la conférence de Santa Fe en 1995 : « A routine is an executable capacity for a repeated performance in some context that has been learnt by an organization in response to selective pressure. » (COHEN et al., 1996). Une autre définition proche est donnée dans le même travail: "regular and predictable patterns of behavior that are repeated in similar environmental conditions and that have been learnt in response to selective pressures (COHEN et al. 1996)". Nous précisions au besoin qu'il s'agit d'un phénomène emergent qui concerne plusieurs acteurs.

Notre postulat de départ est aussi d'intégrer les remarques de DUBUISSON (1998), un article critique fondé sur des études terrains, qui apporte un éclairage sociologique (ce que ne renierait pas Luigi Marengo par exemple)[537]. Selon la chercheuse, l'approche évolutionniste

[535] Sur la différence entre règle et routines, nous renvoyons à REYNAUD (2004).
[536] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-December/001786.html .
[537] Pour une approche critique plus générale de la notion en économie, nous renvoyons à FELIN et FOSS (2004).

« permet […] d'exclure de la définition des routines, les actions qui mettent en jeu un choix ou une réflexion » et de « [négliger] le traitement politique de la notion de routine, en termes de résolution de conflits et de coordination d'intérêts divergents ». De plus, l'approche classique en économie ne prend pas au sérieux suffisamment le rôle des artefacts comme « moyens de stockage et de maintien de la connaissance dans les organisations ». Par ailleurs, un des contextes où elle a tenté d'observer des actions routinières à l'extérieur des organisations est inhabituel, mais pertinent pour nous : le comportement des usagers dans une gare. Ce terrain a des points communs avec le wiki, où les contributeurs passent de façon éphémère.

Nous avons donc choisi comme définition de départ une approche consensuelle parmi les économistes (la première définition). Mais nous allons en spécifier le contenu de façon plus originale, comme le recommandait cette sociologue. Les routines font l'objet de délibération, révélatrice d'arbitrages et de tâtonnements (et pas seulement de disputes et jugements), et ne sont pas un simple automatisme, contrairement à la vision courante un peu réductrice. Notre matériau, une liste de *discussions*, est justement révélateur de ces processus.

Plus généralement, notre orientation choisie est celle de l'approche historico-culturelle de la cognition proposée par LORENZ (2001), qui introduit une autre façon d'examiner les routines et la résolution de problème, en cohérence avec l'analyse ethnographique de la cognition distribuée, une approche déjà employée pour l'analyse des questions de nouvelles technologies (HOLLAN, 1999). L'analyse des routines « in the field », a déjà été tentée sur ce terrain par quelques travaux (WARGLIEN, 1997). Notre intention est de poursuivre dans cette voie, en particulier par une approche historique assumée. On comprendra qu'une telle approche des routines fait de l'approche historique, plus ouverte à la notion de complexité[538], un outil pertinent.

6.2.2. Spécification des concepts théoriques mobilisés pour comprendre les routines dans ce contexte

Au-delà d'une approche historique qui pourrait se suffire à elle-même sans conceptualisation supplémentaire, nous avons choisi et assumé deux outils conceptuels pour analyser les routines. Dans la perspective que nous avons choisie, une routine peut comporter

[538] La notion de complexité en histoire (celle des intrigues de Paul Veyne) et dans les sciences de la complexité (modélisation des réseaux) n'est pas la même, certainement. La conceptualisation réciproque de chacune, et l'articulation entre les deux traditions de pensées serait certainement un programme de recherche à approfondir. Nous avons discuté de ce sujet à l'occasion avec Paul Bourgine et espérons en discuter encore.

deux éléments complémentaires « internes », pourrait-on dire : les routines comportent d'une part une ou plusieurs heuristiques, d'autre part un principe de justification ou de légitimité qui peut être le fondement de l'évaluation de son efficacité.

Le concept d'heuristique a été développé dans une partie de l'économie évolutionniste hérérodoxe essentiellement en Allemagne, dans le courant dit « fast and frugal heuristics » (GIGERENZER et SELTEN, 2002). Voici cependant une définition d' « heuristique » adaptée à la théorie des organisations :

> « Une heuristique est une méthode qui sert à trouver. Le chemin précis pour parvenir à un but n'est pas fourni, une heuristique donne des indications générales pour atteindre un but qui n'est pas défini à l'avance. Perdu dans la montagne en plein brouillard, si l'on vous dit « essayez de grimper au sommet pour y voir plus clair », cette suggestion est une heuristique. » (SAUSSOIS, 2007)

Parler d'heuristique permet de décrire concrètement les routines. Elles sont adoptées parce qu'elles sont robustes, simples, disponibles, satisfaisantes, qu'il vaut mieux faire y que x auparavant pour des raisons d'adaptation, etc. . Elles caractérisent des solutions économiquement satisfaisantes au sens de Simon dans son fameux modèle.

Par ailleurs, nous utilisons la notion de convention, issue de l'économie des conventions en complément. Nous sommes partis de l'idée, *inductive*, que les formes de *justifications* sont d'importance dans le cadre d'une étude compréhensive des comportements réguliers des contributeurs à la communauté WordPress, en particulier ce qui nous intéresse, les routines. L'idée d'analyser les routines comme heuristiques ne nous semblait pas suffisante pour les comprendre : leur choix peut se faire parce qu'il est légitime, et non par arbitrage coût/avantage. A l'intérieur du courant «institutionnaliste» en économie, nous avons choisi l'économie des conventions[539], qui est le courant économique le plus pertinent pour traiter des questions de justifications. D'autres théories concurrentes seraient moins

[539] Un tel choix n'est pas totalement original, car la thématique du logiciel libre a déjà donné lieu à la marge des travaux conventionnalistes d'économistes (HORN, 2001, DE BESSES, 1999.) ou de sociologues (AURAY, 2006). La théorisation de la notion d'information, dont le formatage et l'organisation sont des questions stratégiques dans cette communauté, a également fait l'objet de contributions théoriques originales (FAVEREAU, 2001). L'application de cette notion parfois controversée à l'analyse des réseaux en général (notre communauté peut se décrire ainsi) en est cependant à ses débuts (FAVEREAU et LAZEGA, 2002). Enfin, l'articulation entre le courant conventionnaliste et les sciences de gestion (GOMEZ, 1994, PARIS, 2003) ne pose pas plus de problèmes que d'autres approches économiques institutionnalistes plus «orthodoxes», en particulier lorsqu'il est question de théorie des organisations. Nous renvoyons sur ce sujet par exemple le cours de Théories des organisations de Jean-Michel Saussois dans le Master 2 « GDO », ou EYMARD-DUVERNAY (2004), etc..Le CRG a plusieurs fois interagi avec des chercheurs conventionnalistes (Favereau, Thévenot). D'autres travaux existent déjà traitant des SI (la thèse de Kevin Mellet), mais c'est encore très marginal. L'économie des conventions s'est intéressée à la notion de gouvernance (on renverra par exemple aux travaux de Batifoulier et Réberioux).

pertinentes à cette fin, par exemple l'approche de Salais et Storper utilisé par HORN (2004) dans le cadre de l'analyse *globale* du secteur du logiciel, ou la théorie néo-institutionnelle des contrats utilisée par Brousseau pour l'analyse du numérique. Le groupe de travail du GDR TICS sur la gouvernance de l'Internet[540] montre que les questions de légitimation et de normes sont incontournables. Mais ce n'est pas pour autant que la question est effectivement traitée dans le cadre des communautés (RUZE, 2009), à quelques exceptions.

Par ailleurs, l'articulation entre analyse des heuristiques et prise en compte des conventions, spécifiquement mobilisés ici est un sujet théorique nécessitant approfondissement, comme l'explicitent GIGERENZER et SELTEN (2002), mais en fait rarement abordé (KECHIDI, 2005, est le seul que nous connaissions) : « In many cases, artifacts, heuristics and conventions coevolve. ». Un bon exemple de coévolution est le feu rouge : un artefact (l'objet matériel avec des couleurs), une heuristique (« s'arrêter si…»), un aspect conventionnel (une régularité arbitraire est préférable, si elle existe, au désordre), et un principe de légitimité (la « cité civique »)[541].

6.3. Conception de la sous-question de recherche

Pour résumer, l'analyse historique des routines suit un protocole. Elle se fait par codage thématique des courriers électroniques, après repérage des passages décrivant la sélection de pratiques, des décisions et arbitrages, des réflexions coût-avantages. Nous cherchons alors quelles sont les heuristiques employées alors par la communauté et les formes de légitimation sous-jacentes. Un travail de hiérarchisation de leur importance relative est par ailleurs nécessaire.

6.3.1. Sélection des sources et précision des biais

Nous avons analysé en historien la liste de discussion de coordination de l'activité de documentation, qui sera notre source privilégiée. Il en est de même des archives des minutes des discussions sur IRC. Nous avons dépouillé les onglets de discussion « user :talk » des administrateurs pour compléter, confirmer ou réfuter nos premières conclusions. Cette source

[540] Source : http://gdrtics.u-paris10.fr/groupes/gouvernance.php?req=detail (consulté le 13/07/09).
[541] Nous rappelons qu'économie des conventions et histoire sont complémentaires, par exemple à l'IDHE.

n’a pas été très riche, contrairement à nos premières attentes. Enfin, certaines pages du wiki centralisent les règles informelles de la communauté, mais elles sont sujettes à caution étant donné qu’il est certainement possible qu’il y ait écart entre les « guidelines » et les pratiques effectives. L’examen des pages du Codex qui traitent de l’auto-organisation de la communauté suppose un travail conséquent de *repérage* et de connaissance du terrain, mais aussi un premier travail sur les « historiques ».

Le *degré de validité* de cette sous-question est significatif, étant donné que nous effectuons une recherche systématique des routines à partir d’une source complète et de formes de triangulation. De plus, étant donné que cette source recoupe les opinions de nombreux participants, sans censure apparente, le degré de validité est renforcé.

Cependant, nous ne pouvons le plus souvent que décrire l’existence de pratiques, il n’est pas toujours possible de déterminer leur extension. La difficulté de ce type d’analyse est aussi de faire la généalogie des routines significatives à partir d’une source différente des pratiques elles-mêmes : l’activité se fait sur le wiki, il est difficile d'observer les routines in situ au sens strict, nous n'avons accès à ces phénomènes que par des archives où elles apparaissent à l'état de traces. En revanche, cette approche a un autre avantage, c'est de "tamiser" les comportements essentiels qui apparaissent dans les discussions centralisées, les faits où il y a décision collective à prendre. La liste de discussion archivée est un organe de régulation qui « filtre » les thématiques abordées, cependant sans censure. C’est là la source indirecte principale, et l’analyse des traces de pratiques qui y sont évoquées est d’emblée de nature historique.

6.3.2. Spécification du codage

Une forme de codage souple, thématique a été effectuée à la seule fin de *s’y retrouver* dans des données et non de faire de la « grounded theory ». La façon la plus économique de traiter les données a été trouvée après *quelques tâtonnements* et après un premier dépouillement exploratoire des données nécessaire pour avoir une vue d’ensemble. Nous avions effectué un premier dépouillement sur fichier Word, hésité à passer sous Excel ou sur un logiciel CAQDAS[542], choisi de passer au codage sur papier[543], et finalement centralisé et finalisé le codage sur Word, car les requêtes pour tel ou tel code et les retours sur l’URL

[542] Une tentative n’a pas été convaincante, car Nvivo ne semblait pas apporter grand-chose de plus par comparaison avec le coût d’entrée requis.
[543] … suite à une crise de tendinite, à l’exemple de ce que fait Hervé Dumez, et par lassitude de l’ordinateur.

d'origine des textes était indispensable et rendaient les résultats plus accessibles pour la rédaction ultérieure. Lors de la rédaction de l'histoire « immédiate » de la communauté, mentionnée dans le précédent chapitre, nous avons eu l'occasion de *relire intégralement* les archives, et de compléter la liste des passages pertinents. Il y a eu trois lectures en tout.

Le codage permet de mettre en évidence des thématiques émergentes, et surtout d'avoir une idée de l'évolution chronologique de notre thématique (quand observe-t-on pour la première fois un comportement donné, et quand s'est-il modifié ?). Les formes d'évolutions peuvent être travaillées et apparaître sans trop de difficulté, car *les dates apparaissent* sur les liens que nous citons.

Concrètement, nous avons regroupé les routines en thématiques générales marquées par des *titres*, issus du codage thématique, ce qui peut laisser parfois une impression d'hétérogénéité. C'est l'un des avantages de la présentation choisie, *qui permet de repérer les codes* et leur état final de conceptualisation. La structuration des titres ici remplace le « codebook », qui n'aurait pas apporté davantage, en particulier dans une perspective historique. Tous les fragments significatifs associés aux codes peuvent être retrouvés au moyen de leur URL.

6.3.3. Ordre des opérations pour l'analyse et l'établissement des routines

La démarche adoptée est d'ordre compréhensif, interprétativiste, et historique.

La démarche est assez facile à comprendre pour l'analyse « conventionnaliste » : il suffit de repérer dans quelle mesure tel ou tel principe de justification et de légitimité soutient telle pratique ou l'amène à être sélectionnée. Nous nous proposons d'identifier autant que possible des formes de justifications explicites (à partir de ce que disent clairement les sources) ou implicites (à partir de notre culture des communautés ou par déduction à partir des sources) des heuristiques observées. Cela dit, nous ne pouvons pas forcément le faire pour toutes les routines.

Nous avons commencé par une phase d'établissement et discussion des faits liés à des thématiques (d'organisation et de management) qu'il s'agit de repérer (exemple : la supervision de l'activité), et d'identification des points d'inflexion dans les pratiques.

Pour identifier la *sélection* des heuristiques, nous avons adopté plusieurs voies possibles par adaptation aux données qualitatives qui permettent de comprendre quelles solutions « satisfaisantes » (au sens de Simon) on été retenues. On peut tenter de mettre en

évidence des formes de « tâtonnements » dans les pratiques observées, les formes d'exploration/exploitation lors de la résolution de problèmes. On peut tenter d'identifier des passages où il était question d'arbitrages. Pour comprendre leur émergence et leur sélection, on peut examiner toutes les situations où, dans le discours des volontaires, il est question d'analyse coût/avantage, de balance avantages/inconvénients, et autres thématiques économiques (aléa moral, sélection adverse, coût d'opportunité, situation de hold-up, appropriation, etc.). L'approche consiste à déduire le processus de sélection à partir de données *argumentatives*.

Il est nécessaire de *hiérarchiser* les routines, c'est-à-dire de déterminer celles qui sont cruciales pour comprendre la viabilité du projet sur wiki et son efficacité, ce qui n'est pas un résultat immédiat.

Un cas de routine où cette démarche a été particulièrement exhaustive est celui des sous-pages utilisateurs, que nous verrons plus loin.

6.4. Deux aspects institutionnels préalables

Avant de nous lancer dans l'analyse précise des routines, nous montrons que deux formes de légitimité entrent en conflit ou coexistent, et que des formes d'adaptation des pratiques ont été perçues comme nécessaires par la communauté. La première permet de présenter l'arrière-plan institutionnel juridique, l'autre permet de comprendre l'émergence des routines.

6.4.1. Le choix de la licence confirme la légitimité significative de la logique du code dans la communauté de la documentation.

La question du choix de la licence est un problème institutionnel que nous ne pouvons passer sous silence. Il a fait l'objet de débats significatifs, qui ont ressurgi après les premières décisions à ce sujet sur le canal IRC et la liste de discussion de la documentation. La licence choisie est la licence GPL[544]. Il ne s'agissait pas de la meilleure pour la documentation, mais

[544] "[17:55] <relle> It is possible to use the GPL for a manual, but the GNU Free Documentation License (GFDL) is much better for manuals. (source: http://www.gnu.org/licenses/gpl-faq.html#TOCWhyNotGPLForManuals).
Voici le choix définitif de la licence, par *autorité* de Matt:

elle convenait mieux à l'intégration de code du logiciel dans la documentation[545]. La légitimité de la *logique du code* a prédominé, car il n'y a pas eu de changement par la suite, malgré des tentations[546]. Cela n'est en effet pas anodin, et a posé par la suite des problèmes pour l'internationalisation de la documentation, comme le montre le passage suivant :

> "What I'll guess is that the people behind the codex decided it was best to use an imperfect license that was compatible than something targeted at written works but couldn't play with GPL code. It's an inelegant solution that causes problems, such as the translation issue, but can head off others such as the inclusion of codex writing in WP files."[547]

Le choix de la licence GPL a donc été effectué pour rendre compatible le code et la documentation, ce qui a posé par la suite des problèmes pour traiter des questions de traduction. Ce passage montre qu'on a adopté une solution « satisfaisante » pour les programmeurs qui a créé ailleurs un phénomène de « lock-in » par la suite.

La tentation d'aligner en matière de licence le Codex sur Wikipedia a pourtant été présente[548] : la licence de Wikipedia est GNU, mais elle n'a pas été choisie[549]. Par ailleurs, c'est Matt Mullenweg qui possède le site du Codex[550], et il peut sembler compréhensible qu'il ait choisi une licence favorable aux informaticiens de ce fait. La GPL est aussi la licence normale pour les plugins[551] : la standardisation en matière de licence est « expansive », et il est donc compréhensible que la documentation ait été englobée.

La licence crée cependant de fait une séparation à l'intérieur d'un système avec une finalité commune: bien que Codex et forum fassent tous les deux parties de l'aide aux utilisateurs, et que les deux SI soient fortement reliés (contenus, liens hypertextes, contributeurs), il y a clairement une frontière juridique entre les deux, le contenu du forum n'étant pas sous GPL[552]. Pour le forum, la forme de « propriété » est conçue vaguement

---

"[18:02] <relle> Can the issue of copyright protection and license on the Codex be settled by next week? Or is totally dead and the GPL stands and no more discussion about it?
[18:02] <photomatt> relle: it's dead
[18:02] <photomatt> GPL stands
[18:02] <relle> Okay."
(source : http://codex.wordpress.org/IRC_Meetups/2005/September/September28RawLog).

[545] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000471.html .

[546] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-February/001611.html. Pour de plus amples explications, voir ici : http://comox.textdrive.com/pipermail/wp-docs/2008-February/001619.html .

[547] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-February/001622.html .

[548] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000995.html .

[549] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/001005.html .

[550] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/001026.html .

[551] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-October/001061.html .

[552] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000711.html .

comme une appartenance à la communauté[553] ; la question ne s'est pas clarifiée, à part la différence entre cet état de fait et la licence GPL.

La question des droits de propriété est un aspect *secondaire* de notre sujet qui a cependant un rôle structurant en matière de légitimité. Les discussions précédentes sur les droits de propriété montrent en revanche *la valeur* du bien collectif produit[554]. Cependant, la réalité de la menace est une question qui doit sans doute être *nuancée*, même si la crainte de l'écriture d'un livre à partir du Codex a été discutée voire envisagée (nous renvoyons au chapitre 4 traitant de l'histoire de la communauté). Le problème de la propriété commune est si complexe, que le problème du « hold-up » semble résolu[555]. C'est peut-être la raison pour laquelle la prise en compte de la question des droits de propriété a été chaotique et dans une certaine mesure tardive, mais WordPress n'a peut-être pas été la seule dans ce cas[556]. Pourtant, le gros avantage du wiki, c'est que les modifications sont plus faciles à suivre que pour le code,[557], ce qui a des implications en matière de droit de propriété même si cela n'a jamais été toujours clair[558].

6.4.2. Davantage qu'un problème d'incitation, on remarque l'émergence de la nécessité de la gouvernance suite aux problèmes d'organisation du premier wiki

Ce que nous mentionnons ici complète ce que nous évoquions dans le précédent chapitre sur les limites du problème d'incitation et motivation du premier ordre dans la communauté et la place prépondérante du problème d'organisation du contenu. Nous montrons ici que la communauté a mis en place une forme de gouvernance et y a réfléchi à un niveau assez général.

Gouvernance signifie « tenir le gouvernail d'un navire » pour reprendre la métaphore sous-jacente au concept d'après DUMEZ (2004). L'image est pertinente, car elle indique qu'un point d'équilibre doit être trouvé, par *tâtonnements*, comme avec un gouvernail, qu'on ajuste de temps en temps, à bâbord ou à tribord. Plus précisément, dans le cas de l'équipe qui s'occupe de la documentation, la mise en place de la gouvernance de la communauté consiste *d'abord* en matière de contributions à maintenir le cap entre deux écueils, à savoir trop et pas

[553] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000798.html .
[554] Voir aussi : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000712.html (« emotive issue »).
[555] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000711.html .
[556] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000981.html .
[557] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000987.html .
[558] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000996.html .

assez de contributions. C'est un problème *évolutionniste* semblable à celui du Bar El Farol de Brian Arthur (ARTHUR, 1994), où il faut trouver un équilibre entre trop et pas assez de clients dans l'établissement en question pour obtenir une sociabilité satisfaisante (au sens de Simon). Ici, il s'agit de contributions à gérer, comme le montre ce fragment fort pertinent datant du deuxième mois du projet de documentation, que nous citons en entier :

> "I think that for now, and perhaps forever, the documentation needs to have some tighter control. Development of code is in the hands of a small group of developers; this is necessary to keep focus and to get the job done. I think that a parallel should be drawn to documentation. Many hands makes small work, but too many hands makes chaos. Certainly there should always be a way for people to contribute their ideas, and perhaps even their own writing, but it needs to go through a smaller group of people for proofing and ultimately for approval by the developers. I'm concerned that "restriction" on access to working on documentation may be perceived by some as "elitist" or that the docs crew are somehow "special" or part of a "clique" or something."[559]

Un autre point d'équilibre à trouver était de s'inspirer de façon raisonnable du modèle de Wikipedia. Suite à l'échec du premier wiki, les solutions adoptées par la suite n'allaient pas de soi[560]. La réaction de la communauté a été assez nuancée : elle s'inspire du modèle ouvert de Wikipedia[561] tout en désirant instaurer une régulation plus stricte de la documentation. Plusieurs fragments montrent une double influence, celle des autres projets open-source et de Wikipedia[562]. A nouveau, la trajectoire de régulation doit trouver le bon équilibre, l'hybridation étant un mécanisme clairement évolutionniste :

> "With regards to the "this wiki or that wiki" comment: I agree that no two wikis are the same, and that it would be incredibly irresponsible of us to just copy someone else. But any decent wiki must at least consider what Wikipedia does, or risk repeating history. Please do not discount what Wikipedia does as a valuable model to learn from: turning a blind eye to the methods of such a successful site would be equally as irresponsible as blindly copying it."[563]

L'enjeu sous-jacent est que le management du wiki a pour but d'éviter le désordre[564]. Cependant, pour nuancer, la nature de l'outil utilisé, le wiki permettrait selon certains participants de la communauté de relativiser le problème[565].

---

[559] Source : http://comox.textdrive.com/pipermail/docs/2003-December/000082.html .

[560] "It's hard to define a set-in-stone format when we don't yet know which formats work well. The chaos of the previous wiki taught us a lot of what NOT to do, but not a lot about what to do." (source : http://comox.textdrive.com/pipermail/docs/2004-August/000512.html).

[561] Le nombre de références à Wikipedia à partir de juin 2004 montre que la référence au modèle open-media est manifeste sur la liste de discussion.

[562] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000661.html et http://comox.textdrive.com/pipermail/wp-docs/2005-March/000087.html .

[563] Source: http://comox.textdrive.com/pipermail/wp-docs/2005-March/000309.html.

[564] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000655.html.

[565] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000667.html.

Les erreurs sur le premier wiki, qui expliquent les réflexions du fragment précédent, ont cependant conduit à la mise en place de routines et de comportements plus réguliers. Nous citons plusieurs passages convergents sur différents supports qui soulignent de façon unanime l'existence d'un problème d'organisation pour le premier wiki ; c'est donc là un fait établi, sur lequel nous avons tenu à insister de façon concrète. Sur le Codex, les pages tentant d'expliciter les pratiques habituelles dans l'équipe comportent le passage suivant :

> "One reason for the decline in use of the previous documentation was a certain user-unfriendliness of the tool used. We've chosen a new one for Codex called MediaWiki — developed by the Wikipedia team — which we think greatly enhances the user experience. Another issue with the previous wiki was in the organization of content: *it developed without a structure*, making it extremely difficult for users to find information. We hope to alleviate this in the WordPress Codex by implementing guidelines for its organization and by keeping a devoted group of moderators to enforce these guidelines."[566] (C'est nous qui soulignons).

Ailleurs, sur les archives des minutes de chat IRC, un administrateur rappelle les inquiétudes à l'esprit des membres de l'équipe :

> "A deluge of random editors creating new pages all over Codex is the primary worry. I think the Codex has enough useful information now, to be linked to :[[User:Carthik|Carthik]] 15:15, 16 Feb 2005 (GMT)"[567]

Enfin, sur la liste de discussions :

> "In my opinion, what we need is a wiki, but with some amount of control of what the page titles and the organization will be. *Nothing new here*."[568]

Les sources convergent donc pour montrer clairement que le problème de gouvernance du *premier* wiki est un problème d'organisation du contenu plutôt qu'un problème d'incitation[569]. Il y eu certes aussi un problème avec le précédent moteur wiki, PhpWiki, mais ce n'est que l'aspect technique du SI qui ne suffit pas ici et plus généralement en SI. De plus, l'aspect technique est partiel, car cette nécessité de management du contenu ne serait pas due à un accident, mais à la nature même de l'outil :

> "The wiki itself. Now wikis are cool tools, but that's what they are 'tools'. I understood a wiki to exist to enhance a collaborative project, to build, to create - but not to keep and maintain once the information is mature enough."[570]

---

[566] Source : http://codex.wordpress.org/Codex:About .
[567] Source : http://codex.wordpress.org/index.php?title=User_talk:Carthik&diff=next&oldid=7861 .
[568] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000087.html. C'est nous qui soulignons.
[569] D'autres fragments explicitent que la cause de l'échec du précédent wiki est une croissance trop rapide et incontrôlée (Source : http://comox.textdrive.com/pipermail/docs/2004-September/000650.html), que les contributeurs ont beaucoup contribué (source : http://comox.textdrive.com/pipermail/docs/2003-December/000086.html) et ailleurs trop (source : http://comox.textdrive.com/pipermail/docs/2004-August/000534.html). Les indices en la matière sont largement convergents.
[570] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000534.html .

Le problème d'organisation se renouvellera d'ailleurs avec MediaWiki, malgré ses qualités, comme on l'a vu lors de l'histoire de la communauté.

6.5. Les routines de structuration du contenu

Tout d'abord, plusieurs routines ayant trait au *contrôle* de la nature du contenu sur le Codex ont pu être mises en évidence. Elles ont été sélectionnées ici parce qu'elles ont un impact essentiel sur la viabilité du projet de gestion du savoir. Les routines montrent à la fois les limites des pratiques usuelles des wikis, mais aussi leur permanence et légitimité.

Nous avons classé ci-dessous les routines en cinq catégories. Il s'agit des routines de « housekeeping », de standardisation des pratiques de rédaction, de développement de l'architecture du wiki, des brouillons d'articles écrits sur des pages utilisateurs, et enfin de contrôle de la nature du contenu informationnel ajouté.

6.5.1. Les routines de la « gouvernante » (« housekeeping », i.e. le contrôle continu de la qualité de la documentation par des tâches ingrates)

La notion de « housekeeping », probablement propre à la communauté WordPress, permet de revenir avec humour au sens original de « management », « mesnager » : le management, selon l'étymologie, c'est « l'art de la ménagère ». Or, il se trouve qu'une telle activité est stratégique[571] dans cette communauté. Elle est essentielle pour permettre le maintien de la qualité de la documentation, surtout dans un contexte de bénévolat. L'un des administrateurs les plus connus, Lorelle Van Fossen, s'en est d'ailleurs fait une spécialité, a contribué à l'organiser à grande échelle, et ses contributions ont été reconnues comme fondamentales, par exemple lors d'un incident critique sur la liste de discussions[572]. Mais elle n'est pas la seule à assumer ces routines[573]. Plusieurs petites heuristiques sont associées aux tâches[574] de « gouvernante » du wiki, des tâches humbles, mais vitales : corriger l'orthographe, catégoriser les pages[575], tisser des liens entre les pages au moyen de "mots

[571] Certaines de ces routines qui peuvent paraître simples furent pourtant l'objet de débats et d'incompréhensions (voir par exemple le cas des balises indiquant des articles en gestation : http://comox.textdrive.com/pipermail/wp-docs/2008-January/001568.html).
[572] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/001045.html .
[573] On pensera à Carthik, MichaelH, MacManX …
[574] Source : http://codex.wordpress.org/Codex:Maintenance .
[575] Sources : http://comox.textdrive.com/pipermail/wp-docs/2007-January/001419.html ou http://comox.textdrive.com/pipermail/wp-docs/2008-December/001788.html). Certaines formes de catégorisation demandent une vraie expertise « locale » (source : http://comox.textdrive.com/pipermail/wp-docs/2008-May/001746.html). De plus, la catégorisation des pages n'est pas une pratique anodine, elle a une

wikis" (qui enclenchent automatiquement la création de liens hypertextes), etc. Ce type de travail est différent de la rédaction d'articles comportant un contenu technique et est observable dans une certaine mesure en déroulant les historiques.

Cette routine qui fonctionne « en continu » est conforme à ce qu'on observe sur Wikipedia par exemple, il s'agit d'une forme de contribution « humble » des plus appréciées, demandée même des plus *amateurs* (on y appelle ces bénévoles « wikignomes »), ce qui n'a *rien à voir* avec les contributions expertes au code. Le *principe de légitimité* sous-jacent est que les contributions nombreuses mais humbles et appréciées sont valorisées.

A l'opposé de cette routine qui a lieu *en continu*, une autre en matière de contrôle de qualité de la documentation qui vient au contraire de l'open-source, *concentrer* les efforts (de documentation et plus particulièrement la maintenance) sur une durée déterminée en mobilisant tout le monde de façon *contrôlée*, n'a pas été routinisée de façon décisive. Le rituel « Codex week », une pratique calquée sur les périodes intenses de contrôle du code par des « chasses aux défauts » (« bug hunt ») [576], n'a été réellement effectué correctement qu'une fois sur les cinq années de vie de la documentation. Lorelle dit qu'une première tentative, un « wikiweekend », a été un échec: elle proposa par la suite une semaine entière[577]. Elle a coordonné l'opération qui a été un succès, comme le montrent plusieurs fragments. Nous en citons au moins un :

> "Jun 22 16:32:06 relle The Cleanup is a major success.
>
> Jun 22 16:32:08 skippy Codex Cleanup ++
>
> Jun 22 16:32:17 relle Over 2000 maybe 3000 edits in the past few days alone.
>
> Jun 22 16:32:18 * jalenack agrees
>
> Jun 22 16:32:19 Firas Nice work guys."[578]

Les besoins pour ce type d'événement semblaient pourtant persistants à cause des faiblesses persistantes de la documentation apparues dès la fin 2005:

> [17:31] <photomatt> relle: you're on, holes in documentation
>
> [17:31] <ringmaster> Darn moths. […]
>
> [17:32] <relle> we need some kind of promo event to kick up documentation and hole filling in the Codex.

---

importance fonctionnelle évidente (source : http://comox.textdrive.com/pipermail/docs/2004-December/001034.html).

[576] "I agree. Once we get a definite format / structure / standard nailed down, I propose we have a "WordPress documentation day", much like the bug-squashing day we had some time ago. Everyone focuses a lot of effort in a short period of time to just document document document ... and then, the next day, the Wiki editors will be required to cruise through each page to adjust for our standard."
(source : http://comox.textdrive.com/pipermail/docs/2004-August/000512.html).

[577] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-May/000546.html .

[578] Source : http://codex.wordpress.org/IRC_Meetups/2005/June/June22RawLog .

[17:32] <relle> In fact, we need some kind of major event to pull everyone together again for the fall."[579]

Une deuxième phase d'effort collectif du même type a été planifiée, comme le montre le fragment qui suit, mais n'a pas été effectuée, d'après les archives disponibles, alors que le précédent est parfaitement visible dans les sources ; il était donc attendu que la coordination d'un second soit aussi visible :

[19:04] <photomatt_aus> "Should we organize an WordPress Codex Update Week for January 2006?"

[19:04] <shep> yes

[19:04] <masquerade> +1 on that idea

[19:04] <skippy> yes. a comprehensive revamp for 2.0 is needed […]

[19:04] <gsnedders> especially plugin hooks"[580]

Ainsi, l'amélioration de la qualité des textes « au coup par coup » en continu a prédominé, la routinisation n'a pas été effectuée, malgré un succès ponctuel et une efficacité certaine. Nous avons cité plusieurs fragments pour montrer cette absence. La cause est peut-être un manque de leadership suffisant, Lorelle ayant dirigé l'opération réussie en permanence. Ce type de routine pourrait demander un coût d'encadrement que la communauté ne peut supporter, ce qui expliquerait son caractère localisé[581] : c'est efficace, mais c'est coûteux, dans un contexte de contributions volontaires. Il n'est pas facile de comprendre définitivement les raisons de la non-venue d'un événement.

6.5.2. La standardisation des pratiques de rédaction : de quelle façon formate-t-on le contenu sur le wiki ?

Plusieurs routines, au-delà du « housekeeping » sont absolument indispensables pour maintenir un standard de qualité suffisant sur le wiki.

La première est celle du « don d'information », qui comporte deux caractéristiques particulières.

Ce type de comportement collectif (l'activité de don se fait à plusieurs) est plus complexe, *beaucoup moins visible que la précédente routine*[582]. Il a fait l'objet de *tâtonnements* typiques des processus de « satisfycing » à plusieurs dimensions de Simon, que

[579] Source : http://codex.wordpress.org/IRC_Meetups/2005/September/September14RawLog .
[580] Source : http://codex.wordpress.org/IRC_Meetups/2005/December/December7RawLog .
[581] Une conséquence au niveau global de la prédominance des pratiques de « housekeeping » au coup par coup pourrait être le caractère linéaire de la distribution des éditions observé dans le précédent chapitre.
[582] C'est sur de tels points que la liste de discussion est un révélateur précieux, cette routine serait très difficile à détecter sur les historiques des articles.

l'on peut cependant décrire. Nous mentionnons les deux dimensions les plus saillantes, qu'on pourrait caricaturer en dimensions « quantitatives » et « qualitatives » :

- *Quel est le degré « satisfaisant » de don de l'information* ? En effet, trop d'information serait contre-productif[583]. La conséquence matérielle de l'existence d'un seuil implicite est la détermination d'une taille satisfaisante des pages. Ce passage montre qu'un arbitrage a été tenté entre concision et exhaustivité :

  > *"Ironically, in Ubuntu we're (slowly) moving in the opposite direction -- fewer book-style manuals*
  > *and guides with chapters, and more individual pages on specific topics.*
  > *<https://wiki.ubuntu.com/TopicBasedHelp> Similar will eventually happen for the Gnome system on*
  > *which Ubuntu is based.*

  I don't think the two options are mutually exclusive. It's a good idea to have small individual pages on individual topics, but those individual pages should also be organised into some kind of book/chapter format.

  I find a lot of the codex pages too long and too wordy, but striking a good balance between conciseness and good explanations is *a tricky balancing act* and depends a lot on your target audience."[584]

Le corrélat d'une telle question est la détermination du degré de précision de la documentation, limitée par la taille des pages. Il s'agit d'une question difficile qui est fonction du lectorat[585]. Sur ce point, le *principe de légitimité* de la cité de l'opinion est limité : tout le monde ne peut pas ajouter tout ce qu'il veut et ce n'est pas toujours valorisé.

- *Quel est le degré satisfaisant de technicité de l'information*[586] ? Le seuil est difficile à cerner ; il s'agit à la fois de ne pas décourager le lecteur, mais aussi sans doute d'encourager les experts à contribuer sur le wiki. La présence de code par exemple peut par exemple décourager les nouveaux venus[587]. Nous citons un exemple qui montre des débats et l'aspect laborieux de la documentation en la matière :

  "> *There is a balance between being ridiculously thorough and not*

---

583 « Trop d'information tue l'information » (source : http://comox.textdrive.com/pipermail/wp-docs/2006-February/001289.html , et http://comox.textdrive.com/pipermail/wp-docs/2008-May/001737.html).

584 Etant donné que cette routine n'est *pas facilement visible*, nous citons plusieurs passages. Ce point a été évoqué de façon comique au débuts de la documentation (Source : http://comox.textdrive.com/pipermail/docs/2003-November/000005.html, voir aussi: http://comox.textdrive.com/pipermail/docs/2003-December/000087.html, http://comox.textdrive.com/pipermail/wp-docs/2006-December/001403.html, http://comox.textdrive.com/pipermail/wp-docs/2008-March/001637.html).

585 Les preuves de la chose sont répétées et visibles, nous avons renoncé à toutes les relever ; voir par exemple http://comox.textdrive.com/pipermail/docs/2003-December/000091.html , ou http://comox.textdrive.com/pipermail/wp-docs/2006-February/001256.html).

586 Source: http://comox.textdrive.com/pipermail/wp-docs/2006-February/001249.html, « Upgrade instruction too hard ? »).

587 Ainsi, ce passage d'un universitaire américain transféré par Matt sur la liste de discussions : "Your motto "code is poetry" says it all. Your material is not accessible to people like me. I have no interest in computers. I use mine to put out a blog. It is only the substance i care for."(source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001464.html).

> *overwhelming people. I think we weigh towards the latter right now.*

With respect Matt - the upgrade docs are as thorough because of two reasons:

1 - How the coders put it all together

2 - the people doing the upgrade

I have certainly written what I have because of forum questions and far too many personal emails at these times. It's not been for fun.

P.”[588]

Ici, au contraire, le principe de légitimité « inspiré » trouve une limitation : les informaticiens ne peuvent pas montrer leur savoir-faire comme ils le veulent en ajoutant du code partout.

Ainsi, en matière de don d’information, on observe un tâtonnement, avec deux seuils (limites) de satisfaction (au sens de Simon) décrits ci-dessus, qui permettent de souligner l’existence d’une routine tacite en matière de rédaction de contenu. Si elle est observée, nous ne connaissons pas en revanche son extension.

Cependant, de telles routinisations ne montrent pas vraiment de tentative de contrôle strict, les discussions montrent que les contributeurs disposent tout de même d’une certaine liberté, et il s’agit plutôt de routines dépendant du jugement de chacun, ce qui confirme ce que nous avons dit plus haut dans ce chapitre.

En matière de standardisation des pratiques de rédactions, une question aussi triviale que le formatage des titres[589] montre même là la présence des principes de justification de l’open-source dans les pratiques de rédaction les plus *banales*. Nous mentionnons pour l’exemple en particulier deux heuristiques :

- Les titres s’écrivent de façon normale : pas de « ChaMots » (« CamelCase »)[590] dans les url internes au Codex, afin de faciliter l’intégration au reste du site wordpress.org par la suite, une idée apparue déjà sur le premier wiki[591]. La pratique a ici été *imposée*, malgré des résistances[592]. Sur Wikipedia, ils sont pourtant obligatoires.
- La répétition du mot « WordPress » dans les titres, qui forment le support de l’architecture du Codex est recommandée (avec un peu de bon sens). C’est l’une des conséquences de l’intégration du Codex dans le SI global de la communauté: il y a un arbitrage entre

[588] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-February/001249.html.

[589] L’organisation des titres est une préoccupation assez récurrente qui a émergé très tôt : (source : http://comox.textdrive.com/pipermail/docs/2004-December/000965.html).

[590] « Do not use **CamelCase** for page titles: The Codex does not use CamelCase like some other wikis do. All page titles and therefore links should be of normal title case. For example, the page about Codex should have the title "About Codex", with the link formatted as: `[[About Codex]]` and not the CamelCase `[[AboutCodex]]`.” (source: http://codex.wordpress.org/Codex_Guidelines). Le fait de coller deux mots est une routine des systèmes wikis pour créer des liens. Notre lecteur aura sans doute deviné l’origine visuelle du mot « ChaMots ».

[591] Source : http://comox.textdrive.com/pipermail/docs/2004-February/000155.html .

[592] Source : http://codex.wordpress.org/Codex_talk:Guidelines .

référencement rapide par les moteurs de recherche par un degré supplémentaire de redondance de l'information contre élégance et parcimonie des titres (au référencement moins efficace) et qualité du style[593] . On constate sur ce point qu'un manuel en ligne diffère d'un manuel ordinaire où ce type de répétitions serait exclu. L'efficacité technique est le principe de légitimation de cette pratique.

Le principe de légitimité est ici industriel : ce qui favorise l'intégration technique et la fiabilité du processus est valorisé.

Ainsi, les contraintes observées pour comprendre l'émergence et les tâtonnements des routines sont d'une part la prise en compte de plusieurs auditoires et d'autre part l'insertion à l'intérieur du système d'information de la communauté[594]. En matière de standardisation des pages, les routines de la cité de l'opinion, « edit mercilessly » ont été largement bornées dans le cadre d'autres pratiques ; la rédaction est contrôlée, mais le contrôle n'est pas toujours strict.

6.5.3. Le contrôle de l'architecture du Codex

La structuration générale du Codex fait l'objet de deux types de routines, l'une rigide, l'autre plus souple.

Tout d'abord, le *contrôle* de l'architecture est l'objet d'une routine (l'heuristique sous-jacente est « déplacer l'article A du point P1 au point P2 ») réservée sur le deuxième wiki aux administrateurs, suite aux problèmes du premier wiki. La structuration du wiki, c'est-à-dire la localisation d'une page à tel ou tel niveau fait l'objet de discussions[595] et/ou de demande d'autorisation[596]. Sur ce point, la pratique observée *s'écarte clairement* du principe d'usage des wikis selon lequel tout le monde peut contribuer, y compris en matière de structuration des pages. Même entre administrateurs, on se consulte avant d'agir en cette matière[597]. Une

[593] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-April/001509.html .

[594] On observe bien entendu d'autres formes de tâtonnements en matière de rédaction, comme le ton du Codex (« friendly voice »), mais ils nous ont semblé plus triviaux et moins stratégiques, tout en complétant encore notre présente idée que la rédaction du Codex fait l'objet d'un degré significatif de contrôle.

[595] La création de sous-pages dans l'espace principal (« main space ») doit faire l'objet de discussions. (source : http://codex.wordpress.org/index.php?title=Codex:Guidelines&diff=prev&oldid=63062).

[596] Par exemple la localisation de la page de documentation d'Akismet, le logiciel anti-spam de l'entreprise Automattic (source : http://comox.textdrive.com/pipermail/wp-docs/2006-February/001264.html).

[597] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000690.html ou http://comox.textdrive.com/pipermail/wp-docs/2005-February/000019.html .

routine extrême, la suppression de page, fait l'objet aussi d'un contrôle, les administrateurs se chargeant de supprimer les pages « taggées » comme devant être détruites[598].

Ainsi, le principe de légitimité sous-jacent à cette routine est industriel, le contrôle du processus est valorisé.

D'autre part, les formes de design de la documentation sont un espace d'évolution des routines, les contributions devant s'organiser selon des « patterns » ; mais elles sont plus souples que dans le cas précédent ; il est difficile de s'y retrouver voire de vérifier si la routine est strictement appliquée. On pouvait ainsi trouver des références en octobre 2004 au modèle des « FAQ », qui n'a pas émergé de façon dominante, mais qui est resté localisé en un endroit du wiki (et non plus ailleurs dans la communauté, un phénomène d'internalisation). La densité requise de documentation et son caractère hypertextuel marqué aurait limité la diffusion de ce « genre ».

On observe en revanche une volonté affirmée de passage à l'organisation « task-based », i.e., en fonction de l'utilisateur et des procédures successives qu'il doit suivre pour atteindre un objectif technique donné en configurant au fur et à mesure son blog[599]. La légitimation de l'organisation « task-based » du wiki est que la documentation composée ainsi résiste mieux aux changements de versions du logiciel[600]. Cette organisation s'apparente au suivi d'une procédure[601]. Le principe de légitimité sous-jacente est clairement d'ordre « industriel », i.e. la valorisation du contrôle du processus pour permettre la fiabilité.

Les évolutions observées ont fait cependant l'objet de débats[602]. L'aspect moins rigide des routines en la matière peut expliquer que des problèmes de circularité dans la structure de la documentation sont à plusieurs reprises observés[603].

Le « genre » (« editorial pattern ») manuel, une routine de composition plus conforme à la cognition des utilisateurs[604] a été davantage souligné. En décembre 2006, vraisemblablement à la suite de la phase de désorganisation constatée en 2006, les débats sur la structuration du Codex ont été relancés et la structure « livresque », plus conforme à la

---

[598] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000684.html .
[599] Source: http://comox.textdrive.com/pipermail/docs/2004-December/000921.html et http://comox.textdrive.com/pipermail/wp-docs/2005-February/000009.html.
[600] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-February/000009.html .
[601] "There is a major bug in the upgrade instructions on http://codex.wordpress.org/Upgrading_WordPress that can cause Mac users to lose their data. This needs to be corrected. The problem is here: » (source : http://comox.textdrive.com/pipermail/wp-hackers/2007-January/010319.html).
[602] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000207.html .
[603] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001484.html .
[604] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000713.html .

cognition usuelle et/ou fruste des utilisateurs, semblait avoir à nouveau la préférence, alors que ce point avait déjà été évoqué depuis longtemps :

> "The FreeBSD handbook is one of the best online documentation efforts I have seen.
> http://www.freebsd.org/doc/en_US.ISO8859-1/books/handbook/index.html
> It is chronological and thorough.
> Virgil Huston, who absolutely hates the completely non intuitive wiki format
> On 7/7/05, Lorelle VanFossen <lorelle at cameraontheroad.com> wrote:
> > I've spent a little time lately going through the Drupal Handbooks examining how they do their
> > documents layout. I think we all need to review a lot of other online manuals to keep a good
> > perspective on what the Codex has to offer. Remember, we are "promoting" the Codex as an "online
> > manual" not a wiki, and this raises the bar on quality."[605]

Pourtant, le fait que le savoir soit répertorié dans un SI sur Internet avec des modalités d'accès au savoir particulières rend dans une certaine mesure la forme « manuel » classique obsolète, et pas seulement pour des raisons technologiques :

> "Most user manuals and guides are, alas, still written as self-contained entities, on the assumption that people will read them from beginning to end, or at least start at their table of contents. (Often, online manuals are merely the HTML versions of something that is optimized for print in the first place.) Even in the physical world this wasn't very realistic (hence "people don't read the manual" and "RTFM"), but on the Web it is even less realistic. The Wiki is a loose collection of pages which will be visited more often via Google than via the front page."[606]

Les utilisateurs font usage de moteurs de recherche, ne lisent pas, et seraient paresseux[607]. Cependant, ce n'est pas parce qu'on cherche par des moteurs que la structure n'importe pas[608]. Savoir s'orienter est important car les problèmes de structuration touchent même les administrateurs[609].

Avec la création des pages utilisateurs comme espace de brouillon, que nous allons aborder juste après, nous avons ici un élément significatif sur *la limite du principe de légitimation de la cité de l'opinion*, tout le monde ne peut éditer à volonté l'architecture du wiki et il est nécessaire de suivre des régularités ; on observe donc une valorisation du contrôle du processus pour permettre la fiabilité. Le principe de légitimité sous-jacent est donc industriel.

---

[605] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000669.html. Regarder le reste du fil de discussions. On observe une reprise des discussions en janvier 2009 (hors de notre période d'étude) sur les inconvénients de la structuration du Codex et goût pour la structure livresque.
[606] Source : http://comox.textdrive.com/pipermail/docs/2004-December/001088.html .
[607] La paresse des utilisateurs est un poncif récurrent parmi les volontaires des activités de support (voir infra).
[608] Source : http://comox.textdrive.com/pipermail/docs/2004-December/001092.html .
[609] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000143.html .

6.5.4. Les brouillons sur page utilisateurs limitent l'application de la routine « edit mercilessly »

Une routine centrale qui a émergé lors du passage d'un moteur wiki à l'autre est celle des brouillons sur pages utilisateurs[610]. L'un des modes de régulation mis en place par la communauté[611] pour contrôler la qualité de la documentation, *c'est la mise entre parenthèses localisée du principe du wiki* (qui dit que tout le monde pourrait contribuer) lors de la *première* phase de rédaction des articles, une phase pendant laquelle l'article se trouve localisé sur la « page » du contributeur et qui suppose qu'un autre contributeur demande l'autorisation avant de modifier ce « draft »[612]. D'après la version *finale* de la routine qui a évolué, ce n'est que sur décision de son auteur et d'au moins l'un des « admins »[613], que l'article en « version bêta » est « publié », comme ce qui se fait pour une version même imparfaite d'un logiciel (« release »[614]). Ainsi, les premières contributions d'un utilisateur sur un nouvel article se font sur leur espace personnel (« user space »), leur page personnelle sur les marges du wiki, et non plus sur l'espace public (« main space ») du système d'information. Les administrateurs décident plus tard de la parution (« release ») sur l'espace public, une pratique qui provient du monde de l'open-source. Ce type de routine clairement *hybride* a été mis en place après l'échec du premier wiki, afin de limiter la désorganisation des pratiques éditoriales et garantir la qualité de l'information.

L'heuristique sous-jacente de la routine est la suivante : « si vous désirez commencer une nouvelle page sur le wiki, écrivez-la d'abord dans votre espace personnel (« user space »). Par la suite seulement elle sera insérée sur les pages du wiki qui font office de documentation (« main space »). C'est une évolution originale absolument décisive pour comprendre la viabilité du wiki.

---

[610] L'origine de l'idée date de septembre 2004 (source : http://comox.textdrive.com/pipermail/docs/2004-September/000668.html) ainsi que la première page de ce type repérée (source : http://comox.textdrive.com/pipermail/docs/2004-September/000678.html). Cela permet de montrer qu'il s'agit d'une routinisation qui a émergé de façon spontanée avant d'être officialisée, tout le contraire d'une règle.

[611] La chose a fait cependant l'objet de discussions (source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000439.html et le reste du fil en question).

[612] Il y a ainsi une différence conceptuelle propre à cette communauté entre un « stub » (« souche » !), un article à l'état de brouillon accessible à tout le monde sur le Codex et que tout le monde peut compléter, et un « draft », un brouillon soumis à une autorisation de l'auteur. Cette différence de conceptualisation est spécifique à la communauté WordPress.

[613] L'ajout de pages demande une autorisation des administrateurs (source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000640.html).

[614] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000446.html (le texte est cité plus bas).

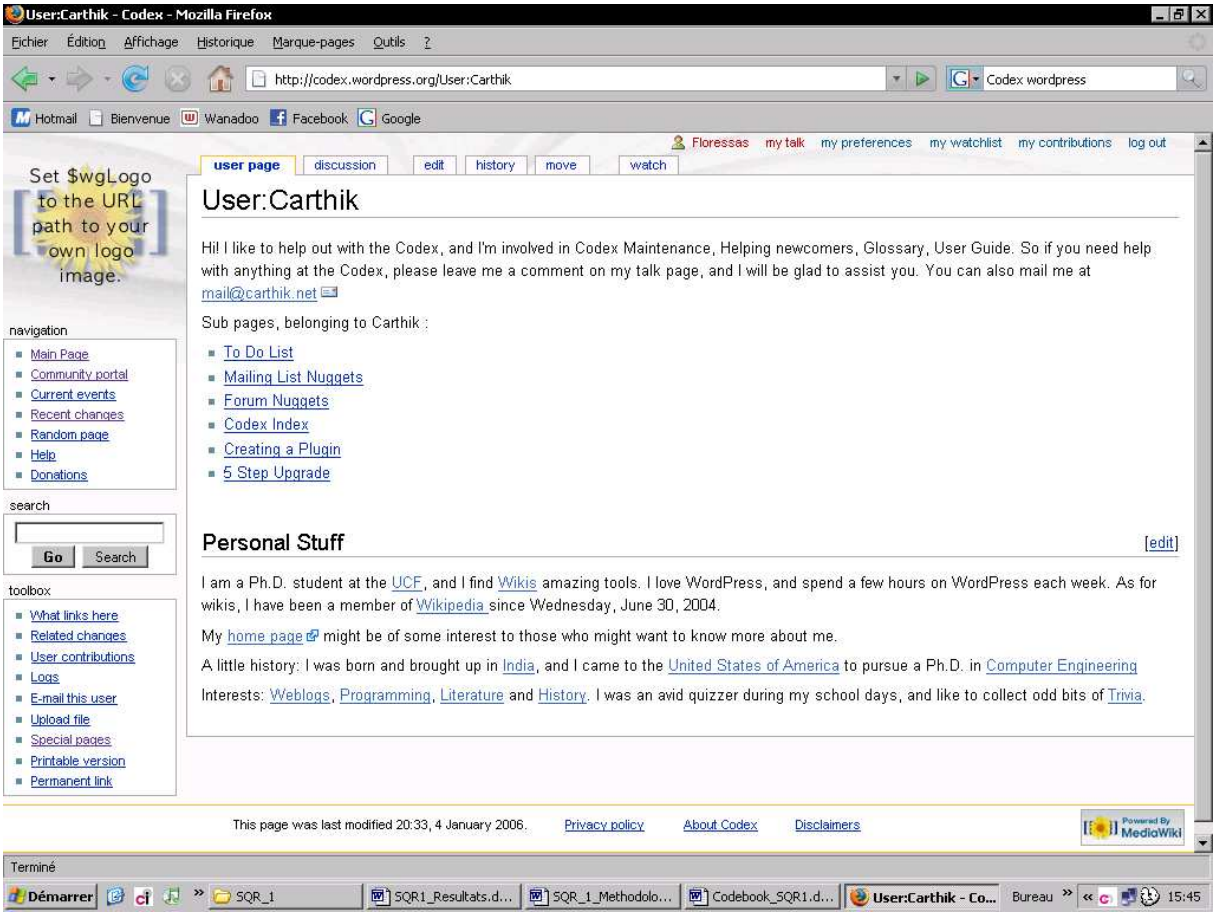


Figure 47 : capture d'écran de la page utilisateur de l'administrateur Carthik où apparaissent les « sub pages » sous forme de liste

Au début, la routine lors de la première proposition était plus stricte encore, la décision revenant aux administrateurs seuls, ce qui donne une idée du degré important de contrôle (nous soulignons le passage en question) destiné à accélérer le processus:

> "I have a proposal regarding the creation of new pages: […] Too many new pages, with titles that are all over the place, and that are not linked-to from other pages. Current Solution: Create new pages as subpages of User: pages, and then move them in later. Proposal:
>
> *1. Sysadmins and Bureaucrats should be able to move in pages as and when they like, without having to check back and forth for speedier development.*
>
> 2. We need to agree on some standards for page titles and stick to the standards.
>
> Benefit: Speed up moving in of useful pages. Don't frustrate doc writers (MDAWaffe comes to mind, among others :))
>
> If you have an opinion regarding these practices, voice them now, within 5 days. Improve the rules above. Decision making threads like this one should have a resolution, summarized in a mail to the

thread, at the end of 5 days (which is yet another proposal), unless there is some special circumstance, which neccessitates extended discussion.

Carthik. »[615]

Un tel degré de contrôle n'a pas été sans inconvénient et a évolué à la marge vers une limitation des prérogatives des administrateurs. La raison en est que certains auteurs ne plaisantent pas avec leur propre production, même si les interventions intempestives sont a priori faciles à corriger. Voici un tel « hurlement »:

« ==Hey, Carthik!==

Why is my new article - DRAFT ARTICLE now sitting in the main stuff????? I'm in panic mode, folks. Please REMOVE IT AT ONCE and all redirects. It is NO WAY ready for prime time!!!!! [[User:Lorelle|Lorelle]] 19:16, 22 Mar 2005 (UTC) »[616]

Cet incident sur le wiki à l'origine de discussions[617] est certainement à l'origine de l'atténuation de la première version rigide de la routine :

"I hear this "release early" garbage over and over, which I think is arrogant, and harmful. Honestly. Release early means putting stuff "out there" with the potential to be very wrong, waiting for someone to pay attention to the material and fix it. Without the huge body of volunteers who are directed and "motivated" to keep track of their areas and maintain timely inspections and edits, an article can sit there for ages before someone pays attention to it."[618]

Cette routine a fait d'ailleurs l'objet de contestations plus radicales d'au moins un administrateur[619], fervent partisan de la routine « edit mercilessly »[620]. Les contestations montrent clairement que le principe opposé est celui de Wikipedia. Cependant, la routine a été définitivement incluse dans le « répertoire » de la communauté[621]. On observe d'ailleurs des

---

[615] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-February/000007.html (remarquer aussi l'emploi du terme « development », qui vient aussi de l'open-source). En fait, il semble que cela existait déjà avant (source : http://comox.textdrive.com/pipermail/wp-docs/2005-February/000010.html), ce qui montre que les pratiques ont été officialisées progressivement, ce qui est caractéristique de l'auto-organisation sans règle officielle. Pour la datation de la codification de cette pratique sur le wiki, voir : http://codex.wordpress.org/index.php?title=Codex:Guidelines&diff=next&oldid=9149 .

[616] Source : http://codex.wordpress.org/index.php?title=User_talk:Carthik&diff=next&oldid=9862 .

[617] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000329.html, qui reprend la plainte précédente de Lorelle.

[618] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000446.html .

[619] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000118.html et http://comox.textdrive.com/pipermail/wp-docs/2005-March/000126.html . C'était peut-être aussi de lui dont il était question ici antérieurement hors liste (source: http://comox.textdrive.com/pipermail/wp-docs/2005-March/000071.html).

[620] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000234.html .

[621] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000146.html .

formes de rappel de la nécessité de cette pratique[622], ou même une auto-correction[623]. L'usage des pages utilisateurs est un *fait massif* observé sur le wiki et sur la liste de discussion[624].

Il est clair qu'une telle pratique s'écarte de ce qu'on peut trouver sur Wikipedia et certainement d'autres wikis. Sur la fameuse encyclopédie, ce type de routine se nommerait « walled gardens », et serait proscrit[625] ; elle ressemble davantage à ce qui se pratique dans le milieu de l'open-source: on parle de « release » pour un article[626] , ce qui est semblable à l'idée que l'insertion d'un morceau de code dans le cœur du système doit être approuvée…. Ce qui est valorisé et légitimé est donc à la fois le contrôle du processus pour permettre la fiabilité et la valorisation du créateur de contenu, l'auteur, au début. *Il y a donc ici deux principes de légitimité, industriel et inspiré.*

L'avant-dernier fragment de Lorelle cité ci-dessus n'aurait *jamais* pu être énoncé sur Wikipedia[627].

Cependant, cette routine semble comporter des exceptions pour des pages de taille modeste ou dont le contenu est répétitif. C'est ce que montre ce passage suivant, que nous avons eu la chance de repérer :

> "Sounds like a great plan! Future WordPress developers will be thankful for any effort you can put into documenting functions in WordPress.
>
> You definitely don't have to mark your new pages as Stub or Draft if your first version is more or less the final version. :)
>
> Regards,
>
> Jennifer »[628]

Il est possible que cette limitation *locale* de la routinisation de l'édition sur les pages utilisateurs soit due aussi à une *confiance* supplémentaire envers les programmeurs, par comparaison avec ceux qui en savent peu en informatique. La question mériterait une plus ample étude, étant donné qu'un nombre significatif de pages sont concernées, étant de taille

---

[622] Source : http://comox.textdrive.com/pipermail/docs/2004-October/000776.html .

[623] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000871.html.

[624] Il manquerait là une chronologie des usages, nous n'avons pas eu le loisir de la mettre en œuvre, mais elle serait envisageable. Nous nous concentrons dans ce chapitre sur les aspects structurels et les ruptures, et non sur une chronologie fine.

[625] Sur Wikipedia, une nouvelle page se crée directement sur l'encyclopédie (source : http://en.wikipedia.org/wiki/Wikipedia:Tip_of_the_day/September_14,_2006).

[626] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000330.html . Cependant, le degré de contrôle des articles rend cette règle de contrôle encore plus sévère apparemment (source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000330.html).

[627] Sur Wikipedia, la chose est cependant apparue seulement comme une *suggestion* de comportement possible, et encore très *tardivement*, le 29 mai 2008. C'est ce que montre une analyse archéologique par tâtonnement de l'historique de l'article de régulation pertinent : http://en.wikipedia.org/w/index.php?title=Wikipedia:Starting_an_article&diff=next&oldid=215653295 . Nous disposons d'une preuve que ce n'était pas le cas sur Wikipedia avant fin 2005, et la communauté en avait conscience (source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000076.html).

[628] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-December/001793.html .

brève[629]. Ce passage montre que ce type de routinisation utilisant les pages utilisateurs comporte à coup sûr des limites, mais qui sont compréhensibles.

Par ailleurs, la routine wikipédienne « edit mercilessly » est conservée sur l'espace principal (« main space »)[630].

Plusieurs raisons et avantages attendus peuvent permettre de comprendre l'émergence historique de cette routine :

- *Cette routine est semblable à une autre qui a déjà été observée lors de la phase de documentation avant juillet 2004* : beaucoup de contributeurs commençaient à écrire des morceaux de documentation sur leurs propres sites, qu'ils contrôlaient, avant de les intégrer sur le premier wiki par exemple[631] en se coordonnant sur la liste de discussion. L'habitude est la même, excepté que les projets de chacun sont désormais écrits via Mediawiki. Il y aurait là un *phénomène de dépendance par rapport au chemin*, une continuité dans les pratiques qui a probablement favorisé l'émergence de cette routine (nous ne pouvons cependant le prouver définitivement)[632].
- L'auteur de l'article peut contrôler son travail en début de rédaction au moins[633], et le publier (« release ») lorsqu'il le désire[634]. Cette routine permet une certaine souplesse : elle permet aussi de cacher certains travaux[635], ou au contraire de permettre un travail collectif[636]. Cette personnalisation de l'espace numérique est appelée ponctuellement de façon révélatrice « desktop »[637].

---

[629] La distribution de la taille des pages suivrait une loi de puissance, comme nous l'avons montré.
[630] Par exemple http://comox.textdrive.com/pipermail/docs/2004-October/000777.html ou lire les guidelines.
[631] Voir par exemple : http://comox.textdrive.com/pipermail/docs/2004-May/000271.html . La chose s'est poursuivie à la marge pendant un certain temps, comme le montre le site de l'administrateur Podz qui a conservé longtemps des éléments de documentation sous format PDF.
[632] Nous avons trouvé un passage où la *transition* entre les deux pourrait apparaître : « Write an entry anywhere (on your blog, or as a sub page of a user page) » (source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000417.html).
[633] "If the user's can't have a "protected" (whatever that means) area so they can sandbox their ideas and work on them in order to maintain the intergrity of the Codex, then we need to get more structured with the Codex volunteers." (source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000446.html). « Sandbox » signifie « bac à sable », et fait allusion à une page du wiki où chacun peut s'entraîner à écrire (pour apprendre la syntaxe wiki par exemple) sans danger pour les autres contributeurs, un peu comme un enfant écrit sur le sable qui s'efface facilement.
[634] Voici un exemple de contrôle de l'utilisateur, qui donne une autorisation aux admins : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000120.html.
[635] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000568.html .
[636] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000769.html .
[637] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-March/001309.html .

- L'artefact choisi est implicitement valorisant et motivant[638], car l'url désigne le contributeur de l'article et le gratifie instantanément en montrant son travail, ce qu'illustre par exemple la capture d'écran qui est la mémoire des projets de chacun. Ceux qui ont écrit des articles conservent un lien vers leur production après introduction de celle-ci sur le « main space ». Le phénomène de valorisation est observé non seulement sur la liste de documentation où chacun dit ce qu'il fait, mais aussi dans les citations de user :page sur la liste de discussion des hackers[639]. Il est valorisant car il existe une différence subtile entre éditeur et auteur, l'auteur étant celui qui écrit l'article, l'autre celui qui l'améliore[640]. Dans la notion de « draft », *la notion d'auteur*[641] *réapparaît*, ce qui n'est pas le cas des wikis en général, où la rédaction est collective.
- Cette routine de rédaction permet un plus grand contrôle des *administrateurs* sur le contenu du wiki[642].

En revanche, l'émergence de cette routine comporte des coûts :

- Tout d'abord, la hausse de la qualité sur le wiki a pour contrepartie une augmentation du risque de duplication des efforts[643] . Si ce risque existe, il est assez limité[644], mais peut-être potentiellement dangereux. Ce n'est pas parce que rien n'est arrivé précisément à cause de cela que ce risque n'a pas d'importance.
- Une autre contrepartie à l'émergence de cette heuristique est que celui qui décide de l'installer au départ sur sa page gagne en cohérence, car il contrôle ce qui est écrit, mais il perd en richesse d'information[645], car tous les lecteurs ne peuvent apporter une formulation différente de sa rédaction ou des retours[646]. Certes, ils peuvent le faire après,

---

[638] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000446.html .
[639] Sur la liste de discussions des hackers, l'essentiel des interventions de l'administrateur MichaelH sont de cet ordre, mais il n'est pas le seul à orienter vers ce type de pages.
[640] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-September/001560.html .
[641] L'en-tête standardisé en haut de chaque brouillon le montre d'ailleurs clairement :
"This article is a **rough draft**. The author is still working on this document, so please do not edit this without the author's permission. The content within this article may not yet be verified or valid. This information is subject to change. ».
[642] "I understand the desire to have people create pages in their User space; it gives some extra control to the sysops etc. in what gets included." (source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000099.html)
[643] « Also, under the "User space" system, duplication of effort is very easy. If this User space system continues, I would suggest asking people to suggest ideas for new pages on this list _before_ starting one in their User space (rather than the other way around) so that duplication can be avoided." (source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000099.html).
[644] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000978.html .
[645] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000076.html . On y mentionne aussi le risque de duplication de l'information.
[646] C'est l'un des avantage reconnus généralement du principe wiki (source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000444.html en 2005 et http://comox.textdrive.com/pipermail/wp-docs/2009-January/001874.html début 2009).

mais ceux qui avaient des informations utiles ne repasseront pas forcément, et la structuration de l'article peut créer un effet de « dépendance par rapport au chemin » qui peut orienter les contributions futures dans un sens qui n'est pas optimal. Il est difficile d'estimer correctement les inconvénients d'une telle pratique sur ce point, nous nous contentons de mentionner sa probabilité.

- La mise en place d'une telle routine nécessite donc un surcroît de discipline et de coordination, et les nombreuses mentions des nouvelles pages utilisateurs sur la liste de discussions montrent aussi la nécessité d'une coordination centralisée[647]. Cette routine suppose un coût, un effort de contrôle de la structure par les administrateurs et la liste de discussions. En particulier, le système ne fonctionne que si on annonce sur la liste de discussions ces sous-projets[648]. Certains membres de la communauté ont conscience du coût de coordination induit par un tel système hybride qui s'éloigne des pratiques usuelles des wikis, et qui occasionne une division des tâches entre éditeurs (sur leur pages utilisateurs) et administrateurs (qui s'occupent de la structuration du wiki)[649].
- Dès le début, l'application de ce système n'aurait pas été sans mal, même si les raisons ne sont pas claires malheureusement[650].
- Enfin, un tel système a un autre coût, il suppose une intégration technique avec le reste du système d'information de la communauté, au-delà de la liste de discussions. La routine des pages utilisateurs suppose un artefact technique d'élimination des sous-pages utilisateurs des résultats des requêtes effectuées sur des moteurs de recherche pour être viable[651]. Une page wiki du Codex est référencée par Google[652] en *deux ou trois jours*, et les utilisateurs qui cherchent à se débarrasser d'une panne arrivent vite sur les pages du Codex. Certes, un « release » précoce pourra dépanner davantage d'utilisateurs, mais avec le risque que l'information soit erronée, malgré les corrections que les contributeurs à la

---

[647] Contrairement à Wikipedia, on dit ce qu'on va faire (source : http://comox.textdrive.com/pipermail/wp-docs/2008-September/001761.html).
[648] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000099.html .
[649] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000083.html .
[650] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000110.html. Il s'agirait peut-être d'un mauvais départ du processus de vérification des nouvelles pages (source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000078.html).
[651] Source: http://comox.textdrive.com/pipermail/wp-docs/2006-August/001336.html . Voir aussi:
"17:59] <photomatt> wrt the search
[17:59] <mdawaffe> (one more redesign point: the screencaps should be from firefox, not safari)
[17:59] <relle> and codex user pages need to be out of the search, too.
[17:59] <photomatt> I think it's been good in that people are using it more, and I'm seeing more very old threads pop up when people have the same problem" (source :
http://codex.wordpress.org/IRC_Meetups/2005/September/September07RawLog).
[652] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000442.html .

documentation peuvent apporter[653]. Cette routine suppose aussi la création d'une catégorie « draft » qui date d'avril 2005[654] : elle permet le repérage des travaux en cours. Les drafts doivent être signalés, étant donné qu'ils sont détectés par les moteurs de recherche[655]; ils peuvent contenir des informations fausses préjudiciables à la communauté et à la diffusion du logiciel.

6.5.5. Les limites de la parabole de la « soupe aux pierres »[656] est qu'il faut choisir les ingrédients : que met-on de façon routinière dans le Codex ?

Une des routines importante lorsqu'il est question de contribution à un bien collectif est *la nature du contenu apporté* : que met-on habituellement dans le Codex, que ne met-on pas, et pourquoi? Dans quelle mesure cela évolue-t-il ? Le degré d'intégration de l'information dans le wiki donne une idée de la *viabilité* du processus, en particulier sa *crédibilité* dans un monde de développeurs. Cela ne s'est pas fait sans tâtonnements ni interrogations.

Le Codex était destiné à fournir de la documentation aux utilisateurs les plus novices en informatique, cela n'a jamais fait débat, bien que l'efficacité du système d'information puisse être discutée. Cependant, la routinisation de la documentation du contenu comporte *une limite*. En matière de contenu pour les utilisateurs novices, les volontaires doivent expliquer souvent des éléments triviaux, mais il n'est pas possible de *tout* exposer, la tâche serait trop ardue, en particulier pour un système d'information reposant sur le volontariat. L'internalisation des FAQ dans le Codex, par exemple, semblait assez logique[657]. En revanche, dans quelle mesure la communauté doit-elle expliquer ce qu'est un blog[658] et aider au-delà de WordPress[659], dans quelle mesure la communauté doit-elle accepter de documenter l'utilisation de Filezilla[660], qui est un autre logiciel fréquemment cité par les utilisateurs? Dans quelle mesure doit-on expliquer toutes les étapes d'installation du logiciel en fonction de tous

[653] Nous renvoyons au précédent fragment alarmiste de Lorelle qui illustre cette idée.
[654] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000460.html .
[655] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000383.html .
[656] L'explicitation de la parabole a été effectuée au chapitre 2.
[657] Source : http://comox.textdrive.com/pipermail/docs/2004-December/001025.html .
[658] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000625.html .
[659] Source : http://comox.textdrive.com/pipermail/docs/2004-October/000781.html .
[660] Par exemple : http://comox.textdrive.com/pipermail/docs/2004-September/000628.html; ou encore : http://comox.textdrive.com/pipermail/docs/2004-October/000733.html .

les systèmes d'exploitation[661] ? En ce qui concerne les variantes du produit, WP-Mu a été intégré au Codex, des discussions montrent que ce n'était pas le seul choix possible[662]. L'idée de séparer la communauté et l'entreprise wordpress.com peut aussi amener à contrôler la nature du contenu inséré[663].

De la même façon, l'information concernant la documentation pour les utilisateurs plus avancés, c'est-à-dire les développeurs, doit-elle être de façon routinière placée dans le wiki ? Ce point a été sujet à débats, comme le montre par exemple le fragment suivant :

> "[17:36] <ringmaster> I'm slowly coming to the conclusion that dev docs shouldn't be in the Codex at all.
>
> [17:36] <relle> If there was some exciting event coming up, then we can use that to spur on enthusiasm for the docs.
>
> [17:36] <trevorturk> i like having dev docs in the codex!
>
> [17:36] <relle> ringmaster: a lot of people really like them there and want more.
>
> [17:37] <relle> Certainly *not everything* but the stuff that helps with plugins and stuff are good."[664]
>
> (c'est nous qui soulignons)

Cependant, la routine s'est maintenue malgré les limites de cette documentation[665], et ce maintien s'est confirmé lors de la reprise de cette partie de la communauté par J. Hodgdon en janvier 2007[666]. Pour le code, on ne documente les erreurs de langage PhP que dans la mesure où elles sont récurrentes, pas au-delà[667]. Cependant, l'insertion de la documentation des fonctions[668] dans le wiki et non plus à l'intérieur du code-source (« inline documentation »), en est un aspect important[669], bien que cela n'aille pas de soi[670]. Pour les fonctions, les tentations d'automatiser le processus existaient[671] . Une réplique de Jennifer Hodgdon,

---

[661] "WP support and documentation are such difficult problems in general due to the scope that we define for ourselves. A user who does not know the basics of hosting a website and adminstering a user account (her own) on a server should not expect to get lessons about the same from the documentation for a web application! […] There is no reason to feel bad about saying "sorry but that is outside our ken". The sooner we learn that the better. At least, the official set of documentation resources should not be crippled by this. Setting a modest goal (for documentation) is not a crime. »
(source : http://comox.textdrive.com/pipermail/wp-docs/2006-February/001288.html) .

[662] Source : http://codex.wordpress.org/IRC_Meetups/2005/September/September21RawLog .

[663] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-April/001318.html .

[664] Source : http://codex.wordpress.org/IRC_Meetups/2005/September/September14RawLog . Voir aussi http://comox.textdrive.com/pipermail/wp-docs/2005-February/000005.html .

[665] Ainsi, la documentation des fonctions n'a pas toujours bien marché (source : http://comox.textdrive.com/pipermail/wp-docs/2008-October/001769.html).

[666] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-January/001410.html .

[667] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000466.html .

[668] Il s'agit d'une tâche conséquente (source : http://comox.textdrive.com/pipermail/wp-docs/2008-December/001811.html).

[669] La chose n'avait rien d'évident, et plusieurs solutions avaient été successivement retenues (source : http://comox.textdrive.com/pipermail/wp-docs/2008-December/001809.html).

[670] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-October/001068.html .

[671] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-February/001444.html et http://comox.textdrive.com/pipermail/wp-docs/2007-February/001452.html .

coordinatrice de la documentation de la programmation, montre que les avantages du wiki en la matière sont de trois ordres : l'essentiel de la documentation du code s'y trouve déjà (« path dependency »), rapidité et simplicité (contrairement au logiciel qui suppose de sortir une nouvelle version mineure lorsque modifications il y a), enfin, la documentation n'est pas forcément mieux réalisée par les développeurs[672] . La mise en place et le maintien d'une telle routine suppose cependant de supporter *une contrainte non triviale*, la synchronisation avec le reste de la communauté, en particulier la programmation[673]. Ce maintien de la routine malgré ces contraintes est une preuve que le wiki comporte une légitimité sur des questions qui intéressent les programmateurs[674]. Un autre phénomène de légitimité semblable peut être observé lorsqu'on examine ce qui, d'une façon générale, est codé et ce qui n'est pas codé, i.e. ce qui n'est pas automatique, donc, qui doit expliqué et dont la *procédure* doit être documentée. Ainsi, la communauté refuse par exemple la mise à jour (« upgrade ») automatique et donc *sous-traite* la problématique à l'équipe de la documentation[675]. Sur certains points, comme les questions de sécurité, délicates et controversées, un consensus parmi les spécialistes est nécessaire[676].

Pour les thèmes[677] et les fonctionnalités additionnelles (« plugins ») en revanche, la routinisation n'a pas été permanente. Ils ont été ôtés du Codex en juillet 2006 et juillet 2007 respectivement par suite de décisions de Matt Mullenweg[678]. La routine de centralisation de l'information sur le Codex ne fonctionne plus en la matière à partir d'une certaine date, même si dans les faits la routine s'est poursuivie un temps[679] alors que ce n'est plus officiel. La raison originale de lister les plugins sur le wiki était d'abaisser la barrière à la contribution des nouveaux programmeurs au minimum[680], ce qui est conforme à l'esprit du wiki, i.e. maximiser le nombre de contributions et contributeurs. Un fil de discussion sur le wiki[681] confirme que la décision d'en changer est assez mystérieuse, montre les limites de l'aspect

---

[672] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-February/001447.html .
[673] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-October/001769.html .
[674] Certains passages (source : http://comox.textdrive.com/pipermail/wp-docs/2008-November/001785.html) montrent que la qualité de la documentation des fonctions sur le Codex est satisfaisante, mais ce n'est pas l'objet de cette partie que de regarder cela en détail.
[675] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001467.html .
[676] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001491.html .
[677] On peut trouver assez facilement des preuves que les thèmes étaient conservés à l'origine sur le Codex, par exemple : http://codex.wordpress.org/IRC_Meetups/2005/August/August17RawLog" .
[678] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-January/001598.html pour les thèmes, la raison serait le refus de duplication de contenu et l'infériorité qualitative de l'interface wiki. (Source : http://comox.textdrive.com/pipermail/wp-docs/2007-July/001551.html).
[679] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-July/001551.html .
[680] Source : http://comox.textdrive.com/pipermail/hackers/2004-June/000296.html .
[681] Source : http://codex.wordpress.org/Talk:Plugins#Removing_the_list_-_not_a_wise_move.3F .

démocratique de la décision prise[682], et qu'on a empêché le listage « non-officiel » des plugins[683]. Les raisons de la fin de cette politique ne sont pas claires d'après les archives, mais on peut penser au minimum à davantage de contrôle[684]. Il est nécessaire car on ne doit lister sur le Codex finalement que ceux officiellement associés au cœur du logiciel (ceux qu'on télécharge en même temps que lui)[685]. Pour les plugins, la raison de leur disparition du Codex était qu'ils étaient difficiles à maintenir[686], de même que les thèmes et les templates[687]. Pour les thèmes, c'était un travail difficile, et cela ne semble pas être la vocation première du Codex[688]. Faire le tri entre les thèmes autorisés ou non par Matt et les autres aurait été une charge trop importante pour la communauté[689]. Pour tout ce qui est fonctionnalités additionnelles programmables, qualifiées de « cauchemar »[690], le caractère « provisoire » du Codex était nettement plus pertinent et assumé[691]. Ainsi, certaines routines absolument essentielles ont évolué en défaveur du wiki, mais, sur ce point, *nous ne savons pas tout à fait clairement pourquoi malgré de fortes présomptions*. Il est cependant probable qu'il y ait une prise de décision « politique » derrière l'évolution de cette routine, mais nous ne savons pas le détail d'après nos archives. Ce type de décision est certainement proche de ceux observés dans l'open-source. Ce type de routines « macro » serait plus susceptible de décisions « dictatoriales », contrairement aux routines micro. Est-ce étonnant ?[692]

Par ailleurs, l'insertion de contenu dans le wiki suppose l'émergence d'une routine de collecte des données, questionnements et informations dans le forum des utilisateurs[693]. Le processus ressemble à une « récolte »[694].

---

[682] La non-insertion des plugins dans le Codex est regrettée par la suite (source : http://comox.textdrive.com/pipermail/wp-docs/2008-January/001594.html). La principale raison de tels regrets serait la perte de liberté des programmateurs qui pouvaient répertorier (et non stocker) comme bon leur semble ces fonctionnalités additionnelles sur le wiki, au lieu de les stocker sur un site officiel (http://wp-plugins.net/beta/) appartenant à une partie de la communauté (source : http://comox.textdrive.com/pipermail/wp-docs/2008-January/001596.html).

[683] La présence de ceux-ci dans le Codex s'est cependant poursuivie (source : http://comox.textdrive.com/pipermail/wp-docs/2008-January/001590.html).

[684] Source : http://comox.textdrive.com/pipermail/docs/2004-November/000814.html .

[685] Source : http://comox.textdrive.com/pipermail/docs/2004-November/000812.html .

[686] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000318.html .

[687] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000320.html .

[688] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000392.html .

[689] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-July/001540.html .

[690] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000645.html .

[691] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000321.html .

[692] Sur d'autres points, c'était attendu, il n'y a pas de routine, et pleine liberté est laissée aux contributeurs ; par exemple, il n'y a pas eu de politique officielle pour la documentation des anciennes versions (source: http://comox.textdrive.com/pipermail/wp-docs/2007-January/001424.html).

[693] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000576.html ou http://comox.textdrive.com/pipermail/wp-docs/2005-August/000834.html .

Pour résumer, en fonction du type de contenu, on observe soit une approbation de l'ajout de contenu sans contrôle, soit une valorisation du contrôle du processus pour permettre la fiabilité. On a donc selon les cas deux principes de légitimité, « cité industrielle » et « cité de l'opinion ».

6.6. Les routines de gouvernance de la population de contributeurs au Codex

Nous venons d'examiner cinq ensembles de routines ayant trait à la structuration du contenu. Cependant, plusieurs ensembles de routines doivent être pris en considération pour comprendre la viabilité du projet. Ils concernent cette fois la régulation de la sous-communauté, et non plus le contenu, même si une bonne régulation de la communauté a un impact sur le contenu, et qu'il y a interdépendance entre les deux classes de routines. Encore une fois, on constate une forme d'hybridation entre les pratiques de l'open-source et de l'open-media, qui cependant se maintiennent.

Nous examinerons donc successivement cinq types de routines, celles ayant trait à l'ouverture aux contributeurs, à la communication, au traitement des conflits, la supervision et à la motivation.

6.6.1. « Massivement » ? Quel est le degré d'ouverture du wiki?

Quelle est la routine d'inclusion des nouveaux contributeurs potentiels et le degré d'ouverture associé ? Cela ne va pas de soi et n'est pas toujours implicite : certains wikis limitent l'entrée des contributeurs d'une façon ou d'une autre. Les communautés open-source ont également des façons sélectives d'accepter les contributeurs informaticiens, et c'est certainement le cas aussi de celle-ci.

Cette routine a un impact direct sur la démographie de l' « équipe » (« team ») de documentation et donc sur la viabilité de celle-ci. *Sur ce point, la routine montre la domination quasi-totale du principe de légitimation d'ouverture à tous de la cité de l'opinion : on valorise tout le monde*. Une tentative de bannir les comptes des « lurkers » a suscité une protestation, autant des inactifs que de certains administrateurs qui trouvent

[694] On parle de « fourrager » (« fodder », source : http://comox.textdrive.com/pipermail/docs/2004-December/000923.html), de « moissonneur du forum » ("forum harvester », source : http://comox.textdrive.com/pipermail/docs/2004-December/000886.html).

qu'une telle évolution serait contre l'esprit du wiki[695]. Il a ensuite été confirmé que le simple fait de s'enregistrer est utile[696] et la routine s'est maintenue[697]. La communauté accepte même les contributions des utilisateurs non-identifiés[698]. L'opposition à toute « barrière à l'entrée » en matière de contributions et de contributeurs est confirmée quelques années après dans les archives[699]. Il ne s'agit pas seulement du wiki, mais de l'ouverture de la liste de discussions aux « lurkers »[700] , qui sont en effet utiles pour les critiques et compliments[701].

Cette routine a été *partiellement limitée* par une décision dont la légitimité provient de l'open-source. Il s'agit du « blanchissage » du Codex fin décembre 2005[702] : la décision a été prise d'homogénéiser l'apparence extérieure[703] du wiki de sorte que son esthétique soit conforme au reste des sites de la communauté. Tout le problème pour cette évolution, qui est le corrélat de l'officialisation de la documentation[704], est de faire face à des exigences contradictoires : d'une part l'intégration[705] dans une communauté en tant que site web et système d'information[706], qui nécessiterait la prise en compte de l'esthétique, un concept absolument central pour ces programmateurs (ce qui a causé le « blanchissage »), d'autre part la facilité d'utilisation pour les contributeurs. Le « monobook skin » classique de MediaWiki est beaucoup plus pratique, plus facile d'utilisation pour les lecteurs. La routine d'accès au wiki avec sa nouvelle apparence (identification et modification du « skin ») est plus

---

[695] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000257.html . Dans les versions récentes du logiciel MediaWiki, les utilisateurs « actifs » sont dénombrés.

[696] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000312.html . Le problème s'était posé peu avant, lors de la décision finale d'acceptation d'un compte avec un pseudonyme comportant bizarrement le mot « gay » et qui avait suscité la perplexité des chasseurs de spammeurs (source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000307.html).

[697] Source : http://comox.textdrive.com/pipermail/docs/2004-August/000534.html. Ou encore plus tard :
"[5:35] <masquerade> photomac, wouldn't stop spam, especially when its things like images
[5:35] <photomac> I haven't investigated it
[5:35] <skippy> We could violate the principles of a wiki, and require some sort of manual account creation process for would-be contributors. ick
[5:35] <photomac> skippy, I'd prefer not to do that "
(source : http://codex.wordpress.org/IRC_Meetups/2005/November/November02RawLog). C'est nous qui soulignons. « Photomac » est ici un pseudo « écorché » pour Matt Mullenweg.

[698] Source : http://comox.textdrive.com/pipermail/docs/2004-November/000812.html .

[699] Une malencontreuse proposition en la matière a trouvé une opposition résolue de la part d'un contributeur respecté au Codex. (source : http://comox.textdrive.com/pipermail/wp-docs/2008-March/001640.html).

[700] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000904.html .

[701] On citera par exemple : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000342.html .

[702] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-December/001173.html .

[703] On appelle cela le « skin », « peau », c'est-à-dire configuration visuelle modifiable. La première configuration apparaît sur la précédente capture d'écran qui affiche la fleur de tournesol du moteur Mediawiki.

[704] Source : http://codex.wordpress.org/IRC_Meetups/2005/May/May18RawLog (le texte complet cité est placé en annexe, étant donné que les « minutes » de l'IRC sont assez longues.

[705] C'est certainement aussi le cas dans d'autres communautés faisant usage des wikis, comme celle du célèbre Mozilla, qui offre une documentation avec le moteur MediaWiki modifié (https://wiki.mozilla.org/, consulté le 20/06/09).

[706] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-November/001356.html .

complexe, même si la difficulté est assez anodine : des contributeurs ont eu des difficultés, mentionnées sur la liste[707]. Plus sérieux, la nouvelle esthétique du wiki cache qu'il s'agit d'un wiki : les liens « éditer les pages » apparaissent en effet en grisé quasi transparent, ce qui cache la fonctionnalité pour les nouveaux bénévoles. La chose a été décidée longtemps à l'avance et effectuée alors que les archives montrent que la documentation laissait à désirer et que des débats ont eu lieu sur cette modification[708] , au point que manifestement beaucoup de contributeurs ont préféré configurer leur accès à MediaWiki avec le « skin » précédent, qui offre plus de confort[709]. La routine antérieure s'est donc maintenue pour une partie de la population des contributeurs.

La convention sous jacente à cette évolution pour des raisons d'esthétique, la « cité inspirée » va à l'encontre de celui de la cité de l'opinion (désir de contribution sur le wiki) : on limite dans une certaine mesure l'ouverture car on valorise l'esthétique. On observe un compromis entre les deux cités, qui n'aurait pas existé si l'artefact de la technologie n'avait pas facilité ce changement (tous les wikis n'offrent pas techniquement la possibilité de choisir un « skin »).

6.6.2. La routinisation des outils de coordination de l'activité : quelle coexistence ?

L'existence même de l'archive de la liste de discussion de la documentation montre clairement que la coordination de l'activité de documentation ne se fait pas seulement sur les pages du wiki. Un artefact technique essentiel qui a fait très tôt l'objet d'un processus de routinisation lors d'une phase de tâtonnements, d'essais et d'erreurs dans la sous-communauté

---

[707] Voici un dialogue entre deux personnes pour lesquelles la chose a posé problème: "On Wed, 2008-01-16 at 11:31 -0800, Carlo wrote:
> Scrap this thread. I am actually able to edit the wiki. I am an idiot,
> I didn't see the edit link on the RHS.
I don't blame you. It's hard to find, and if you're logged in with my account, it doesn't even show up. Nobody believes me, though, so I don't think anybody's bothered to fix it. That's probably what's driven me away from the Codex." (source : http://comox.textdrive.com/pipermail/wp-docs/2008-January/001593.html).

[708] Ici un mécontentement sur la nouvelle apparence du Codex : http://comox.textdrive.com/pipermail/wp-docs/2006-November/001347.html .

[709] « The old codex was just fine. This new reliance on the wordpress stylesheet is kind of cumbersome when you're trying to parse a lot of information […] I do think that the monobook feel of the wiki is something that people are comfortable with and works well. I don't see in this case a reason to have the new wordpress formatting used on it. I'd highly suggest going to monobook as default and letting people select the wordpress layout if and when they log in." (source : http://wordpress.org/support/topic/53224). L'ancienne routine pour éditer, désirée, n'a pas été retenue. Un autre fil montre que le blanchissage affecte négativement la structure « techniquement parlant » (source: http://wordpress.org/support/topic/110931.

de la documentation dès fin 2003 est *l'usage de la liste de discussion, une routine observée sur plus de 5 ans*.

La liste permet beaucoup d'opérations, la « socialisation », la mise en commun des projets et initiatives, rendre « connaissance commune » les « trous » dans la documentation, « signaler les duplications d'efforts », rendre « connaissance commune » un plan de rédaction d'article[710], demander la révision d'articles[711], signaler les demandes de contributions expertes[712], contrôler ce qui doit être ajouté ou non à la documentation[713], etc. En revanche, certaines possibilités de la liste n'ont pas été routinisées, comme la publication d'un journal de ce qui se passe dans l'équipe par Lorelle[714].

L'usage de la liste de discussions comporte *des bénéfices dont la prise en compte permet de comprendre l'arbitrage effectué en faveur de cette heuristique* :

- La liste de discussions a été utilisée bien avant le canal IRC ou les onglets de discussions, et elle serait donc davantage utilisée par phénomène de « path-dependency » : continuer à en faire usage est plus commode et devient conventionnel.
- La liste de discussions sert de mémoire centralisée à la communauté, permet une forme d'expertise et d'apprentissage collectif, des routines, des erreurs faites[715]. Les onglets de discussions ne sont pas du tout faits pour cela de par leur caractère dispersé. Les archives d'IRC ont en revanche cette propriété, mais sont plus pénibles à utiliser et demandent un compte-rendu.
- Un autre avantage de la liste de discussions est de rendre connaissance commune (« common knowledge ») rapidement aux participants le flux d'activité de la documentation et les décisions prises[716], ce qui est difficile voire impossible sur un wiki de grande taille, à moins d'y être actif à plein temps. Cela est utile par exemple pour éviter les duplications d'activités. Cela permet de rendre « common knowledge » les messages adressés à plus d'une personne[717]. Nous mentionnons un passage représentatif de cette

[710] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000931.html .
[711] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000794.html .
[712] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000662.html .
[713] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/000967.html .
[714] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000795.html .
[715] Par exemple : http://comox.textdrive.com/pipermail/docs/2003-November/000026.html, http://comox.textdrive.com/pipermail/wp-docs/2008-December/001794.html, http://comox.textdrive.com/pipermail/docs/2004-May/000264.html et http://comox.textdrive.com/pipermail/docs/2004-September/000642.html.
[716] On observe que les décisions pour modifier le codex sont suffisamment complexes pour ne plus être prises de façon décentralisée… Le wiki ne fonctionnerait donc pas comme outil de décision (contrairement à ce que mentionne par exemple le pamphlet *De la supercherie wiki*).
[717] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000569.html .

activité de coordination, à la fois *temporelle* et « *géographique* » (dispersion des zones d'activité) :

> "On Thu, Sep 04, 2008 at 09:31:43AM -0700, Lorelle on WordPress wrote:
>
> *> Thanks. I thought there was an article in the Codex on Widgets. Hmmm, have*
>
> *> to catch up and recheck things. ARGH. Thanks for the catch!*
>
> I've started moving the API documentation over. There was already a document here: http://codex.wordpress.org/Plugins/WordPress_Widgets_Api
>
> Which I've modified heavily and created a bunch of sub-pages off of.
>
> Everything from here: http://automattic.com/code/widgets/api/ has been copied over. The undocumented functions I'll add later."[718]

En revanche, l'usage d'un tel mécanisme de coordination peut comporter des coûts ; voici ceux qu'on peut relever :

- La coordination sur la liste a un impact négatif sur l'efficacité de composition de la documentation: « I say that (a) design by committee sucks, and (b) the ratio of actual documentation to discussion about documentation is *already* far too low. :-)"[719] . Il est difficile cependant de le quantifier.
- La coordination sur la liste est plus « bruyante » dans les boîtes aux lettres électroniques des contributeurs : les discussions sur le wiki font « moins de bruit », n'étant pas centralisées et redirigées; la liste de discussions comporte cette caractéristique, est potentiellement plus « bruyante », avec par exemple 333 messages par mois en mars 2005. Tout le monde ne désire pas forcément suivre la liste[720].
- Ce canal de communication aurait une dimension « politique », mentionnée ici plus bas parmi les raisons qui font que certains préfèrent les onglets sur le Codex.

En ce qui concerne la coordination par « chat » IRC hebdomadaire, on observe une phase de coexistence, puis une éviction, d'après les archives. L'analyse de l'historique des archives montre que la centralisation des discussions et des décisions s'est concentrée sur la liste de discussions, et non plus sur IRC, après une période de coexistence ; on observait en effet beaucoup de discussions sur IRC au début à propos du wiki :

---

[718] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-September/001761.html . Ainsi, contrairement à Wikipedia, on dit ce qu'on fait sur la liste.

[719] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000966.html. Nous pouvions effectivement observer ce décalage dans l'histoire de la communauté, entre la trajectoire des éditions et celle des interventions sur la liste de discussions, qui n'obéissent pas exactement aux mêmes rythmes. Le passage mentionné ici montre que certains membres de la communauté avaient observé la chose, qui devient « connaissance commune ».

[720] Source :http://codex.wordpress.org/Codex_talk:Community_Portal#centralized_wiki_meta_discussion (16 février 2006).

> "The Wiki documentation of the IRC meetups is exactly what was suggested and implemented to address this issue the last time I complained about it. That's good, and should continue, but its a transcript of a one hour IRC meetup dominated by the discussions of the codex team (that's not a bad thing, I see decisions being made by the codex team just as I do on the wp-docs mailing list)."[721].

Le ralentissement des activités sur le wiki a probablement contribué à évincer l'un des trois canaux de discussion, celui-ci. L'usage de l'IRC a servi lors des périodes d'activité soutenue sur le Codex, et pas aux autres périodes. Deux passages montrent par exemple que l'IRC est moins important :

> 14:39:39 <relle> We've tried all sorts of things and this is what works BEST for what we want to do.
> 14:39:44 * Firas dimunitively notes that codex formatting issues are best left for the docs list or something
> 14:39:47 <photomac> okay"[722]

Une approche quantitative à partir du dépouillement des archives des minutes montre que l'usage de l'IRC a été localisé dans le temps :

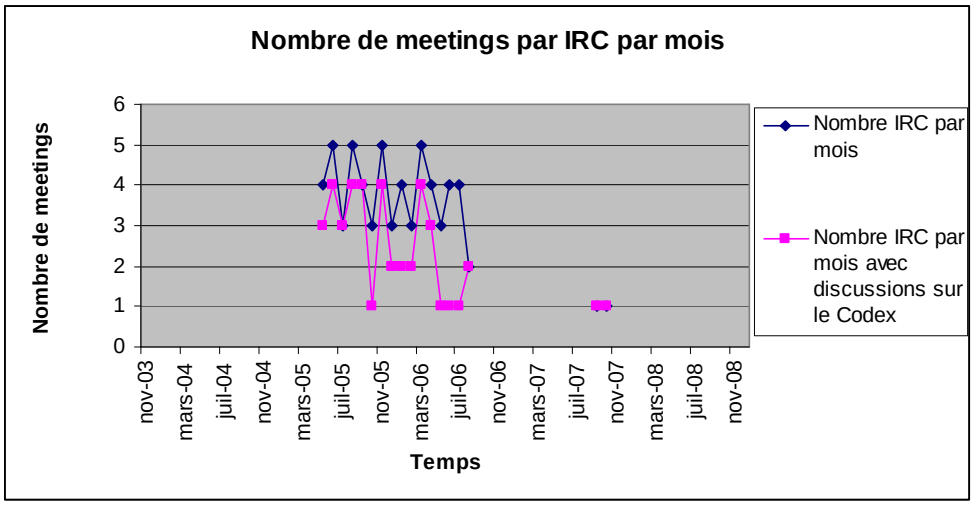


Figure 48 : série chronologique des meetings IRC hebdomadaires[723]

On observe une complémentarité entre les deux autres canaux de communication, la liste de discussions et les onglets du wiki, et donc une hybridation des pratiques par coexistence des routines de communications, même si la liste de discussions tend à évincer les

[721] Source : http://comox.textdrive.com/pipermail/wp-hackers/2005-June/001430.html . Cette importance, observée effectivement dans les minutes (au moyen de recherche par mots-clefs dans les pages en question) mais pas forcément au point de parler de « domination », est un élément à relever, qui montre l'importance du phénomène d'organisation de la documentation.
[722] Source : http://codex.wordpress.org/IRC_Meetups/2005/June/June01RawLog .
[723] Ces données ont été tirées du dépouillement de http://codex.wordpress.org/IRC_Meetups . La localisation des meetings où il est question du Codex a été effectuée par repérage de mots-clefs avec la commande clavier adéquate.

pages utilisateurs (« user pages »)[724]. Le passage qui suit montre qu'il y a complémentarité entre les discussions sur le wiki et l'usage de la liste de discussions; et la nécessité d'une liste de discussions serait une limite des wikis en général, et pas seulement un aspect de notre communauté.

> "I return now to codex after a decent wiki education via wikimedia projects and find it frustrating that there is apparently zero conversation about this wiki on the wiki itself. I see that it is partially due to low edit #'s in general... but I wish there were something similar to the Village Pump here on the wiki? The email list is a much higher barrier to entry (and committment) than simply asking a question on a meta-discussion wiki page. Further, bringing conversation into the wiki allows for easy reference within the wiki to past discussions -- and even policy.
>
> I see potential aversion due to forking discussions between the wiki and the list -- but in general I think this is a better idea than bad. The wikimedia projects I have participated in work well with a combination of in-wiki discussion alongside fundamentally important email lists. Often individual parties will see only one -- but it all seems to work quite well. In my current situation -- I am unlikely to follow the *wp-docs* email list, but check the codex periodically."[725]

En revanche, certains passages montrent que les principales discussions ne doivent pas s'y trouver pour des raisons pratiques, comme celui-ci, où Matt Mullenweg, le chef de projet, prend la parole :

> "May 11 16:38:07 photomatt maybe if the http://codex.wordpress.org CodexContributing page could have a better call to action
>
> May 11 16:38:08 Firas i almost just suggested that the codex should have a wiki
>
> May 11 16:38:18 RandyWalker Firas: uhhh
>
> May 11 16:38:19 photomatt discussion should probably *mostly* happen over the docs list
>
> May 11 16:38:23 westi Firas: but that wiki would need a codex!
>
> May 11 16:38:27 photomatt as following discussion on the wiki is horrid
>
> May 11 16:38:35 relle It puts a big YOU HAVE MESSAGES on the site when I visit. Always post messages in the User discussion area."[726]

En ce qui concerne les onglets, les statistiques exploratoires pour les discussions sur le wiki sont les suivantes. Au sortir de l'été 2006, on décompte 1500 éditions de discussion sur les articles, et 3300 éditions de discussions sur les pages utilisateurs. Par comparaison, fin août 2006, on dénombre 2402 courriels dans la liste de discussions. Le temps de dépouillement des éditions des articles n'a pas permis de compter à nouveau par la suite. En revanche, il est plus aisé de décompter le *nombre* de pages concernées par les discussions.

---

[724] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-May/000523.html .
[725] Source :http://codex.wordpress.org/Codex_talk:Community_Portal#centralized_wiki_meta_discussion (consulté le 16 février 2006).
[726] Source : http://codex.wordpress.org/IRC_Meetups/2005/May/May11RawLog .

Ainsi, le décompte permet de dire que début avril 2009[727], on constate que 387 articles sont concernés par des discussions, dont nous ignorons l'importance. En ce qui concerne les onglets de discussions des utilisateurs, 4537 ont été impliqués mais avec un biais conséquent : plus de 9 interventions sur 10 au moins sont des messages de bienvenue (« welcome »), c'est à dire des copiés-collés qui donnent une liste des pratiques de comportements sur le Codex et qui sont essentiels pour la bonne marche de la communauté. On peut donc estimer au maximum à 500 les contributeurs discutant ainsi. Estimer ainsi l'importance relative des discussions sur le wiki et des discussions sur la liste est impossible. On peut en revanche affirmer que des discussions significatives ont eu lieu sur les onglets de discussions du wiki, et qu'en conséquence les deux modes de communication se sont juxtaposés. L'outil wiki MediaWiki a d'ailleurs été sélectionné précisément à cause des fonctionnalités « onglets de discussions »[728]

Quelles peuvent être les raisons de l'usage à la marge de ces onglets ? Deux passages sont pertinents sur ce point. Le premier est une remarque de l'administrateur Lorelle repérée dans les strates des archives de discussion d'un utilisateur (Boneill), le 6 juin 2005, c'est-à-dire un an après le début du wiki, et qui montre l'intérêt de la coordination décentralisée :

> "And also, to reply back, you need to click on my name and edit my discussion page to leave a message so the "you have messages" flag will be triggered. Other people might not check your user page for your response. It's just how we "do" things around this crazy place. It makes following a conversation near to impossible except for the two parties involved. So major discussions happen in the mailing list or the IRC chat #wp-docs."[729]

Ce passage montre que les onglets ont une utilité « locale », semblable à la communication privée et dyadique par courriels, mais qui est susceptible d'être vue, y compris dans les archives des historiques. Un autre fragment d'un administrateur confirme que toutes les interactions ne se font pas sur la liste de discussions, qui serait un « lieu de pouvoir », et que les discussions peuvent être cachées sur les onglets :

> "Carthik, I just wanted to mention [[USER:MDAWaffe/Administration Panels]] here, since I have no faith in the docs mailing list <code>:)</code>"[730]

On observe donc un classement d'importance des routines de communication : liste de discussion, onglets de discussions, « chat IRC ».

---

[727] On peut obtenir des statistiques à partir de la page spéciale du wiki http://codex.wordpress.org/Special:SpecialPages répertoriant par catégories le contenu en manipulant liens et menus déroulants … et en s'armant d'un peu de patience lors des décomptes.

[728] Source : http://comox.textdrive.com/pipermail/hackers/2004-July/000990.html .

[729] Source : http://codex.wordpress.org/index.php?title=User_talk:Boneill&diff=next&oldid=13608 .

[730] Source : http://codex.wordpress.org/index.php?title=User_talk:Carthik&diff=next&oldid=7163). On se rappellera d'ailleurs que, dans un fragment précédent mentionné ci-dessus, le même administrateur préférait que son travail à l'état de brouillon reste caché.

On observe donc une *juxtaposition* entre les routines justifiées par la cité industrielle (valorisation du contrôle central du processus), et de la cité de l'opinion et une coexistence des deux types de routines principaux, même si l'une est plus importante que l'autre. Nous avons d'une part un modèle plus décentralisé, moins visible, plus informel, accessible, d'autre part un autre plus centralisé, suivi par tous et difficile d'accès, poster dans la liste pouvant intimider.

Nos conclusions, d'après les archives disponibles, sont qu'une routine de communication n'est plus employée, et que les deux autres se sont maintenues et sont complémentaires. Au-delà de ces arbitrages inconvénients/avantages et de ces formes de complémentarité, la communication a été efficace dès le début, comme le montre ce passage de la liste de discussions des « hackers », mais il ne fait pas forcément référence à la liste de discussions, et n'est pas forcément valable pour toutes les périodes, même si nous n'avons pas trouvé de fragment problématisant sérieusement le fait que les problèmes de communication ont un impact négatif décisif sur l'activité de la communauté de la documentation :

> "If you compare the process and volunteers around the Codex there is a marked contrast. There are a number of really good players working on the Codex. Things get planned and discussed in advance. It is very clearly democratic and is working brilliantly. Occasionally some discussions have floundered and ended in a stalemate, then someone has taken the decision and implemented one of the solutions. The codex is a valuable resource. The quality is on the whole outstanding. »[731]

6.6.3. « No one bites here » : les routines de résolution des conflits sont assez marginales

Une pratique vitale routinisée, ce qui est assez compréhensible du fait de son caractère répétitif et simple, est celle de l'exclusion des indésirables. Nous ne connaissons pas l'heuristique exacte employée, n'étant pas administrateur, mais elle ne doit pas être manuellement compliquée. Voici une série chronologique de ce type d'interventions :

[731] Source : http://comox.textdrive.com/pipermail/wp-hackers/2005-June/001430.html .

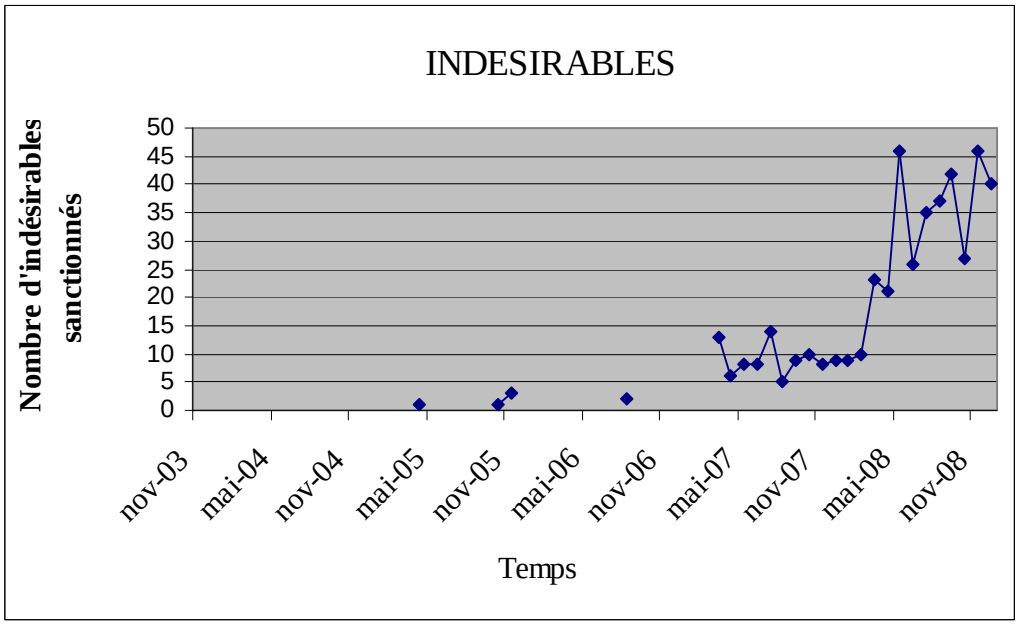


Figure 49 : série chronologique du nombre d'indésirables sanctionnés[732]

Ce graphique permet de confirmer une hypothèse attendue, mais guère originale : avec le succès de la communauté, le wiki est de plus en plus « attaqué » : on relève 459 indésirables en tout fin décembre 2008[733]. Les exigences de qualité de la documentation ne souffrent pas que des pages soient détruites par malveillance (les spammeurs détruisent les pages pour y insérer une série de liens en général tendancieux).

Même cette routine simple a connu une évolution : ainsi, l'observation des sources dans la liste de discussion à propos de la routine d'élimination des indésirables montre qu'aux premiers incidents il y avait centralisation des demandes et décision d'élimination, avec en général peu de discussions (quelques exclusions par erreur). La procédure a été ensuite automatisée et décentralisée sans passer par la liste de discussions. On observe donc une forme d'apprentissage collectif, même si l'administrateur MichaelH a assumé à lui seul l'essentiel de cette tâche ingrate[734].

Nous avons relevé un seul cas où la routine a « bloqué » (i.e. n'est plus automatique), un petit incident ayant fait l'objet d'un gros fil de discussion dans la liste de régulation : on a

[732] Cette série chronologique est issue du dépouillement de la page « système » : http://codex.wordpress.org/Special:BlockList début janvier 2009). Il est possible qu'il y a un léger biais peut-être, il nous semblait qu'il y avait plus de cas relevés dans la liste de discussions. Mais cela ne modifie guère les conclusions.

[733] Nous ne voyons pas actuellement la nécessité d'en faire une typologie dans le cadre de cette question de recherche, mais c'est techniquement faisable. Nous n'avons pas distingué ici de nuances dans les sanctions, qui relèvent le plus souvent de l'exclusion définitive pour cause d'insertion de liens vers des sites n'ayant rien à voir avec la communauté.

[734] Source : http://codex.wordpress.org/Special:BlockList .

exclu 10 jours un nouveau participant qui n'aurait pas eu un comportement « professionnel » en conditionnant son aide au départ de la « clique » dirigeant le projet[735]: la justification de la routine implicite sous-jacente a alors été explicitée [736]. Pour éviter les conflits, une autre routine observée est d'éviter qu'ils ne se résolvent sur la liste : cela se fait hors de la communauté[737].

Ainsi, bien que cette routine de résolution des conflits soit *vitale*, elle est d'une importance assez localisée, et on ne trouve pas beaucoup de conflits dans l'histoire de la communauté[738]. Nous les avons tous repérés lors du chapitre sur l'histoire de la communauté. C'est probablement l'une des raisons de son excellente réputation en tant que telle. On ne trouve pas dans les archives de routine typique de l'open-source en matière de régulation des conflits : personne ne s'est jamais fait « flammer ». L'heuristique observée ponctuellement est plutôt de *calmer autant que possible au plus vite*: « no one bites here »[739], "no personal attacks"[740], excuses spontanées[741]. Cependant, des contributeurs notables, des bénévoles, sont assez sensibles aux critiques quant à la qualité de leur travail, critiques qui doivent être motivées et constructives, ce qui peut créer des périodes d'irritations[742]. Ce n'est pas parce qu'il n'y a pas de conflits que les causes potentielles de ceux-ci n'existent pas.

A quelle cité ou principe de justification avons-nous affaire ici ? L'idée de recherche et de valorisation de consensus relève de la cité de l'opinion. L'usage ponctuel de l'humour[743], ce qui est une pratique observée dans l'open-source correspond plutôt à une valorisation du principe de création (cité inspirée).

Par comparaison, les discussions sur la façon de formater l'information dans le Codex sont quantitativement beaucoup plus importantes[744].

---

[735] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-January/001235.html .
[736] Source : http://comox.textdrive.com/pipermail/wp-docs/2006-January/001244.html .
[737] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000396.html .
[738] Source : http://comox.textdrive.com/pipermail/wp-docs/2009-January/001837.html.
[739] Source : http://comox.textdrive.com/pipermail/docs/2004-April/000178.html .
[740] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000394.html .
[741] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000396.html et http://comox.textdrive.com/pipermail/docs/2004-December/000919.html .
[742] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000564.html .
[743] Par exemple : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001470.html .
[744] Des routines aidant à la résolution de contentieux n'ont pas pris, à savoir faire une liste des spammeurs (source : http://comox.textdrive.com/pipermail/wp-docs/2005-May/000526.html), et mettre des images à la place des images tendancieuses, à cause de l'excès d'effort manuel demandé : (sources : http://comox.textdrive.com/pipermail/wp-docs/2005-August/000824.html).

6.6.4. Les routines de supervision, entre autorité et responsabilisation : l'autonomisation de l'équipe de documentation

La question sous-jacente à l'analyse de ces routines est : quel est l'équilibre entre autorité et autonomie dans cette communauté en ce qui concerne la documentation ?

On observe trois niveaux de régulation informelle : l'autorité des hackers chefs du projet WordPress, la régulation par l'équipe de documentation et les administrateurs, et la responsabilisation personnelle (autour de « projets »), les trois catégories d'acteurs concernés pouvant se recouper de façon complexe. On le voit, même dans un projet de cette nature, les routines ne furent pas uniques.

La mise en place d'une supervision bienveillante et diversifiée dans l'équipe de documentation s'est faite en réaction à l'échec du précédent wiki[745]. La question de la « supervision » de la documentation, et, plus avant, celle de l'autorité, est en effet absolument essentielle pour comprendre l'efficacité du processus de documentation. Très tôt émergèrent des discussions sur le *degré* de contrôle de la documentation, une idée qui a pour inspiration les pratiques des développeurs :

> "Development of code is in the hands of a small group of developers; this is necessary to keep focus and to get the job done. I think that a parallel should be drawn to documentation. Many hands makes small work, but too many hands makes chaos. Certainly there should always be a way for people to contribute their ideas, and perhaps even their own writing, but it needs to go through a smaller group of people for proofing and ultimately for approval by the developers."[746]

On observe dans certains fragments pertinents à cet égard des formes de routines d'autorité et de supervision qui dessineraient les contours d'un « habitus » (des réflexes de comportement) ; elles vont à l'encontre de justifications de l'autonomie dont la nécessité est reconnue dès le début de la création de la documentation :

> "Craig,
>
> I know what you mean about not wanting to duplicate efforts, etc, as well as your earlier point about this time of year not being conducive to getting much done. I think part of the problem is that we (the non-dev listmembers) are probably waiting for someone (like Ye Olde Grande WordPress PooBah) to

[745] "Part of the reason for the decline of the previous wiki was the user-unfriendliness of the wiki tool we used. We've chosen a new tool for Codex called MediaWiki — developed by the Wikipedia team — which we think will greatly enhance the user experience.
The second issue with the previous wiki was in the organization of its contents: it developed without a structure, making it extremely difficult for users to find information. We hope to alleviate this in the WordPress Codex by implementing guidelines for its organization and by keeping a devoted group of moderators to enforce these guidelines." (source : http://codex.wordpress.org/Codex:About).

[746] Source : http://comox.textdrive.com/pipermail/docs/2003-December/000082.html.

say, Craig, do this, Cena do that. However, I think that as 'regular users' we're in a better position sometimes to tell what needs documenting.”[747].

Un tel réflexe d’autorité continue à être admis par la suite, y compris par les contributeurs les plus chevronnés comme Lorelle van Fossen :

“[5:52 PM] <photomatt> new front page for the codex?

[5:52 PM] <relle> http://codex.wordpress.org/User:Lorelle#Codex_TOC_New_Look

[5:52 PM] <relle> Was hoping to have it up by now but Carthik has gone missing.

[5:53 PM] <photomatt> what's the holdup?

[5:53 PM] <relle> a decision.

[5:53 PM] <relle> permission.

[5:53 PM] <relle> chain of command.

[5:53 PM] <relle> say when photomatt and it's done.

[5:54 PM] <photomatt> I think the page looks great.”[748]

Voici au moins un exemple de l’autorité de Matt envers le même contributeur :

“Jun 14:11:41 <photomatt> I would strongly object to such an article on the codex

Jun 14:11:41 * relle peers at everyone and they turn away with guilty looks.

Jun 14:11:52 <relle> Why? Ads on blogs is a problem?”[749]

On peut aussi citer la décision *sèche* de ne pas utiliser de « CamelCase » pour le formatage des liens hypertextes à l’intérieur du wiki, toujours sur décision de Matt[750]. Ce type de pratique *autoritaire* n’a rien à voir avec ce qui se fait sur Wikipedia : le principe de valorisation du *créateur* de l’œuvre d’art informatique (qui contrôle d’autorité sa création) correspond à la « cité inspirée ».

Ainsi, on observe nettement aux débuts de la documentation, en 2004-2005, des formes d’autorité marquées. *L’heuristique serait « obéir aux instructions des principaux développeurs »*. L’autorité porte sur ce qu’il faut faire, la manière de faire. Il ne s’agit pas d’ordres cependant, mais l’autorité a un sens dans un espace fondé sur le volontariat.

On note cependant une évolution, i.e. une autonomisation et une prise d’indépendance significative de l’équipe de documentation pour certaines décisions importantes, leur travail légitimant leur pouvoir de décision. Nous mentionnons par exemple la décision de changer de CMS (logiciel de gestion de contenu) pour la documentation, une décision lourde, qui avait la préférence de Matt, à l’origine du nouveau CMS, Blicki. Ce projet, qui est devenu un « serpent de mer » pendant un an dans l’équipe de documentation, ne s’est jamais concrétisé, sans doute par *habitude* de la routine d’instruction précédente. Sur le point qui nous intéresse,

[747] Source : http://comox.textdrive.com/pipermail/docs/2003-December/000053.html .
[748] Source : http://codex.wordpress.org/IRC_Meetups/2005/June/June08RawLog .
[749] Source : http://codex.wordpress.org/IRC_Meetups/2005/June/June15RawLog .
[750] Source : http://codex.wordpress.org/Codex_talk:Guidelines .

nous mentionnons un passage qui montre clairement des discussions et une indépendance de l'équipe de documentation :

> "[19:43] <photomatt_> finally - New Codex. [...]
>
> [19:44] <masquerade> Considering a different wiki software?
>
> [19:44] <photomatt_> ryan and I (mostly ryan) have been working on a new plugin for similar purposes, called blicki [...]
>
> [19:47] <Podz> it's a massive amount of info to transfer over
>
> [19:47] <photomatt_> ultimately the decision is up to the docs team, they do the work, so it's whatever tool their prefer
>
> [19:47] <photomatt_> Podz: it's a ton of data, but an importer isn't *too* hard
>
> [19:47] * MichaelH shivers at that prospect. [...]
>
> [19:47] <Podz> unless you make the new one 2.0 only and freeze the codex
>
> [19:48] <photomatt_> but the rest is worried that we're putting hundreds of hours into a system which is highly fragile. [...]
>
> [19:50] <photomatt_> ultimately I'm not going to change anything without the docs team on board."[751]

C'est donc l'équipe de documentation qui reçoit le pouvoir de changer de logiciel, une responsabilité de taille. Le passage suivant *résume* bien l'impression de celui qui parcourt les archives, en matière de prise de décision dans les fils de discussions de la liste : on observe une routinisation de la prise de décision démocratique dans l'équipe de documentation :

> "Jun 15:25:59 <tunicwriter> I see no problem with it unless Matt does. Obviously we can't make any real decisions, but we can ablest have a general consensus on the issues ready for him when he gets back"[752] .

On observe des formes de délégations et de coordination[753], et la routine a commencé à émerger assez tôt, par exemple en matière de management des contributions :

> "May 11 16:41:56 photomatt relle: would you like to be the person who assigns incoming ~~victims~~ volunteers to the codex?
>
> May 11 16:42:18 relle Hmmmm.
>
> May 11 16:42:33 relle I'll think about it.
>
> May 11 16:42:49 MCincubus http://codex.wordpress.org/Put_me_in_coach ;-)
>
> May 11 16:43:05 RandyWalker I want to work on the Codex too

---

[751] Source : http://codex.wordpress.org/IRC_Meetups/2006/February/February08RawLog .

[752] Source : http://codex.wordpress.org/IRC_Meetups/2005/June/June15RawLog ; voir aussi :
"Jun 22 16:09:16 Strings28 WikiPress is a nogo
Jun 22 16:09:16 relle Wow. That's democracy in action.
Jun 22 16:09:20 skippy motion carried. Now Matt can ignore us all, as usual. =)
Jun 22 16:09:26 westi skippy: :P " (Source :
http://codex.wordpress.org/IRC_Meetups/2005/June/June22RawLog).

[753] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-December/001564.html; voir http://comox.textdrive.com/pipermail/wp-docs/2008-January/001588.html qui prouve que la coordination fonctionne.

> May 11 16:43:11 relle I think that anyone could push someone towards an area IF THEY KNEW WHO DID WHAT."[754]

La conséquence de cette routinisation et forme de délégation est que le management de la documentation donne du pouvoir, et provoque *une mise à distance*, ce que montrent plusieurs passages. Ainsi, Lorelle est qualifiée de « goddess »[755] ; ailleurs, on craint de déplaire aux dieux (« displease the gods »)[756] ; « star » s'est transformé en « gods ». Lorelle se dit « god-like » de part son statut d'admin qui lui permet de détruire du contenu définitivement[757]. En creux, l'absence d'un administrateur de la première heure peut faire différer des décisions[758]. Il est difficile d'avoir une idée précise de l'extension d'une telle représentation des relations entre contributeurs, dont le chef de projet a conscience[759], qui ne concerne pas que les contributeurs secondaires[760]. Cependant, Lorelle admet que « ruling the world » est le rôle de Matt Mullenweg[761].

Une conséquence de la routinisation de ce mécanisme de coordination, entre autorité et autonomie, est que si le mécanisme de coordination est trop fort, il peut *dissuader* d'autres contributeurs : "but I'm a n00b to the Codex, so I dare not displease the gods."[762]. « n00b » signifie vraisemblablement « nobody » ; le sens n'est pas totalement clair, mais il est certainement péjoratif, ce qui montre l'existence de deux catégories de participants, et que la seconde catégorie se limite dans ses initiatives. Ainsi, sur la liste de discussions des « hackers », ce passage pertinent a posteriori de « Skippy », pourtant administrateur, montre qu'il existerait un phénomène général *d'éviction* des contributions :

> "I was keen on documentation. Along with Podz and Carthik and others, we got the initial wiki going (after we convinced Matt LOL) and then made the transisition over to MediaWiki and Codex[…]. Soon we had lots of good folks jumping in to help out, and they were indeed wiki savvy. However, what happened was that soon these people were basically doing everything, while people like me wound up being relegated to the sidelines.
>
> I could have learned on my own how to do a lot of things, but the point I'm trying to make is that had I been paired up with someone who was more advanced than I, we could have had two "experts" in time. Then we train two, and they train two etc. You get a CSS-effect in your documention team development. The same would be true for support.

---

[754] Source : http://codex.wordpress.org/IRC_Meetups/2005/May/May11RawLog .
[755] Source : http://comox.textdrive.com/pipermail/wp-docs/2009-January/001838.html.
[756] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-May/001744.html .
[757] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-March/001683.html.
[758] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000717.html .
[759] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000643.html .
[760] Au début, Lorelle qualifiait Carthik de « master of the Codex » (source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000698.html).
[761] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-May/000538.html .
[762] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-May/001744.html .

> Right now the support and Codex are dominated by a small group. Again, this is not an attack or judgement, it's simply the reality of a lot of work to be done, and only a core group of folks with the occasional help from a lot of others.
>
> I believe that WP has grown to the point where there needs to be some "middle management" and those folks who would so choose could become involved towards managing support and managing documentation almost as distinct open-source projects of their own. Craig. ”[763]

Nous observons donc les caractéristiques d'une deuxième heuristique, « prendre une décision par consensus » parmi les plus importants contributeurs. La légitimité de cette pratique qui consiste à valoriser l'opinion de chacun par la recherche du consensus (« cité de l'opinion ») est conforme à ce qui se passe sur Wikipedia, où cette notion de consensus est fondamentale[764].

Enfin, nous observons l'émergence d'un troisième niveau, celui de l'autonomisation des projets individuel, qui induit une limite des pratiques de supervision de l'équipe de documentation dès le premier wiki :

> “The docs team has never worked on the assumption that WE will tell YOU what to write, by god. If there's something you want to address, go for it. Just drop a note to the list so we can try to eliminate duplicated efforts.”[765]

On observe une routinisation du lancement de projets, qui correspond à une autonomisation des contributeurs importants[766]. L'heuristique sous-jacente à cette routine est « X prend la responsabilité de tel projet et y contribue ». Ainsi Lorelle a coordonné la partie « WordPress Lessons » et Jennifer Hodgdon la documentation pour développeurs[767] . Le *corrélat* de cette routine organisationnelle est la possibilité de délégation à d'autres, i.e. se décharger d'une quelconque responsabilité ; elle est observée dans certains fragments :

---

[763] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-October/003788.html .*Ce passage montre donc les limites de cette routine de supervision au point de poser des problèmes de gouvernance en particulier de « middle management », mais il est difficile d'aller plus loin avec ce seul fragment. Nous reviendrons sur ce point dans la conclusion générale de la thèse en développant une piste de recherche.*

[764] Si la légitimité de l'équipe est réelle, et se différencie de celle des informaticiens, on observe cependant que la chose va même *au-delà* par rapport à ce qui se passe sur Wikipedia, où les administrateurs ont surtout un rôle plus «policier » de respect des règles et des conventions *complexes* de cette gigantesque communauté, et ne se préoccupent pas de la nature du contenu (il n'y a *aucune* raison qu'ils soient qualifiés pour intervenir avec pertinence sur tel ou tel article en sus de ce rôle de surveillance.). Ici, en plus, ce sont des experts, des interlocuteurs, et des gens qui contrôlent l'architecture du Codex, en plus de la qualité. Ici, la qualité est une donnée stratégique, préoccupation *que n'ont pas* les administrateurs de Wikipedia (à part le spam), mais qui est compréhensible dans le cas de WordPress, qui est un *produit* et une *marque* même s'il ne s'agit pas seulement d'une entreprise.

[765] Source : http://comox.textdrive.com/pipermail/docs/2004-April/000202.html.

[766] “May 11 16:14:42 Firas Possibility of assigning "positions or areas of responsiblity" for coordindation of the various Codex activities and assignments. Lorelle 22:39, 6 May 2005 (UTC) » (source : http://codex.wordpress.org/IRC_Meetups/2005/May/May11RawLog).

[767] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001488.html .

> “I'm leaving the Developer Docs and Advanced Topic section to you guys for now, since I want to concentrate on structure, TOC, and tutorials”[768]

Ce fragment correspond à une délégation de la part de Lorelle. Il n’est cependant nul besoin d’être un administrateur pour avoir des responsabilités :

> “May 11 16:34:04 Semiologic relle: i think a better approach would be to simply making it more obvious who is maintaining the codex
> May 11 16:34:08 relle Already people like Macmanx handle themes and plugins - so why not say so on the list.”[769].

Le contributeur notable « PotterSyst », qui avait 15 ans (d’où sans doute « Potter ») lors de sa prise d’initiative a décidé d’assumer la responsabilité du projet en espagnol[770]. Ce cas limite montre que ce niveau de décision est différent du précédent, où on n’aurait certainement pas accepté un jeune homme de 15 ans.

On peut en effet se demander si ce type de pratique parfois intéressée[771] n’est pas lié de façon significative à une logique de prise de pouvoir locale, ce qui *limiterait* naturellement l’extension de ces zones (plus de place « supplémentaires » pour les ambitieux) et donc de facto la diffusion de la routine en question[772] . Ainsi, ce passage montre bien à la fois que désir d’émergence des responsabilités et appropriation sont liés:

> “May 11 16:32:16 relle The Codex is really tightening up and I'd like to see some responsiblity assigned to volunteers to coordinate efforts.
> May 11 16:32:28 relle After the name of the admin, I'd like to see areas of responsibility.
> May 11 16:32:38 twistedraisin what do you need the volunteers to do?
> May 11 16:32:50 relle So if someone comes in and is a plugin expert, I would know to send them to X who is in charge of plugin information on the codex.
> May 11 16:33:02 relle If someone wants to help with Lessons, they would know that this is my baby.”[773]

Ce développement localisé des *zones de responsabilités*[774] est certes informel; mais certains tiennent tout de même à contrôler la leur ; il peut exister des intérêts divergents et Lorelle par exemple tient à ce qu’on respecte son travail comme « séparé » du reste du Codex à l’intérieur

---

[768] Source : http://comox.textdrive.com/pipermail/wp-docs/2007-February/001455.html .
[769] Source : http://codex.wordpress.org/IRC_Meetups/2005/May/May11RawLog .
[770] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-March/000389.html .
[771] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-April/000496.html ou http://comox.textdrive.com/pipermail/wp-docs/2005-July/000698.html pour l’expression « master of the Codex ».
[772] Source : http://codex.wordpress.org/index.php?title=User_talk:Carthik&diff=next&oldid=11442 .
[773] Source : http://codex.wordpress.org/IRC_Meetups/2005/May/May11RawLog.
[774] Il est difficile d’en déterminer l’origine précise, on peut la dater d’après les archives des minutes d’IRC de mai 2005, mais Lorelle pourrait être à l’origine du constat de manque de la chose et du mouvement de routinisation de cette pratique (source : http://codex.wordpress.org/index.php?title=User_talk:Carthik&diff=next&oldid=10463).

de celui-ci[775]. Par ailleurs, une intervention dans une zone suppose de se coordonner avec le responsable[776].

Là encore, on valorise des contributions de chacun (cité de l'opinion), sans qu'il soit besoin d'être expert.

Ici encore, cette routine dépend d'une autre, la coordination se fait par la liste de discussions, comme le montre la capture d'écran suivante de la page du wiki dédiée à la coordination du projet :

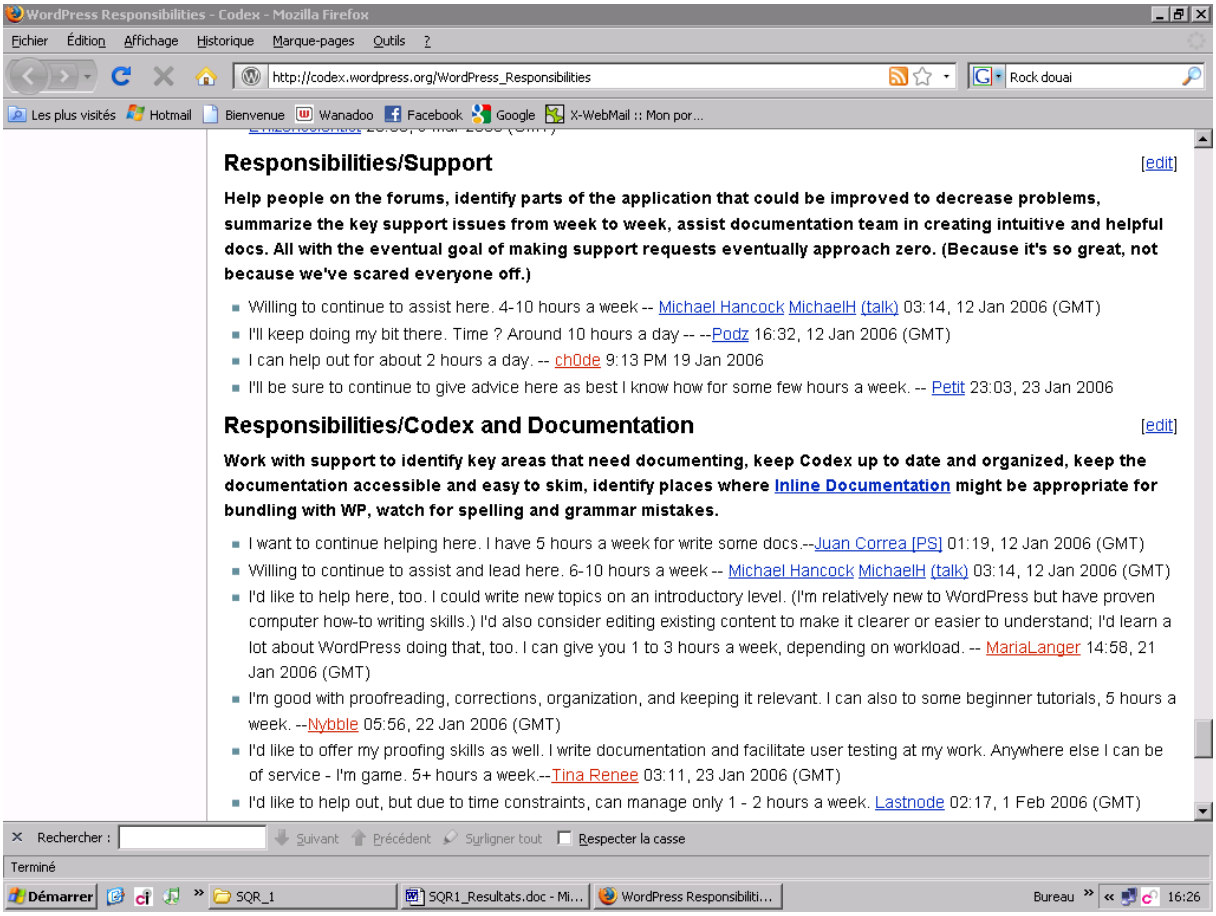


Figure 50 : capture d'écran de : http://codex.wordpress.org/WordPress_Responsibilities

En effet, cette page *dit peu de choses* (six « puces ») par comparaison avec les archives de la liste de discussions à propos de l'organisation des responsabilités. Manifestement, avec la

[775] Ainsi, elle s'élève par la suite contre des usages de son travail contraires à son intention initiale : « The WordPress Lessons were never intended to become replacements for the technical manual information. Due to the lack of articles on some of these subjects, WordPress Lessons are now filtered into the technical side of the Codex." (source : http://comox.textdrive.com/pipermail/wp-docs/2007-March/001490.html).

[776] On citera par exemple : http://comox.textdrive.com/pipermail/wp-docs/2005-June/000585.html ou http://comox.textdrive.com/pipermail/wp-docs/2005-April/000476.html .

croissance du projet, il a cependant fallu tenter de rendre claires les responsabilités sur le wiki. Cela montre aussi que le Codex sert de façon limitée à coordonner le projet de documentation.

6.6.5. Une routine de remerciement emphatique suffit à traiter la question de la motivation dans l'équipe de documentation

Nous nous focalisons ici sur l'existence ou non de mécanismes de motivations extrinsèques (incitations)[777]. L'analyse historique est pertinente pour mettre en évidence « en creux » des phénomènes qui n'ont pas émergé, mais qui auraient pu apparaître[778].

Les archives IRC montrent une volonté de supprimer dès le début des échanges les incitations monétaires, pour éviter un « effet d'éviction » (« crowding-out effect »), une idée que ne renieraient pas des économistes plus chevronnés :

> "May 04 16:17:02 shep i don't see why people would have to pay
> May 04 16:17:05 skippy the other problem with paid support is that the folks who contribute for free have a real disincentive for continuing. (I'm specifically thinking of my own contributions to the codex)
> May 04 16:17:11 masquerade many users would not pay for support, but the ones that would be willing to pay would get a guarantee of a support staff's time to help diagnose and fix the problem
> May 04 16:17:12 jefte I think because of the nature of the product (free, open-source) I don't expect to pay, to get the product to run [….].
> May 04 16:18:49 photomatt free support would NEVER be eliminated, the codex would never go away, etc"[779]

La tentation de rémunérer les contributeurs à la documentation était pourtant bien présente dans l'esprit de certains administrateurs[780]. La mise en place d'une rémunération monétaire pour la communauté sous forme de dons (« donate »), telle qu'elle est observée ailleurs sur la Toile a aussi été envisagée[781], mais n'a pas été institutionnalisée par la suite, dans la mesure où nous n'avons rien trouvé sur ce point. Il n'est pas sûr toutefois que les contributeurs à la documentation aient été concernés.

L'examen suivi des archives de la liste de discussions de la documentation confirme qu'il n'existe pas de pratiques incitatives notables, à part des encouragements, l'adoubement comme « star » de la documentation, et la qualification « d'administrateur », à la portée

[777] La question des motivations intrinsèques a fait l'objet d'un chapitre précédemment.
[778] Veyne (1971) insistait sur ce type de diagnostic.

[779] Source : http://codex.wordpress.org/IRC_Meetups/2005/May/May04RawLog .

[780] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000884.html .
[781] Source : http://comox.textdrive.com/pipermail/docs/2004-December/000887.html .

symbolique, politique, et technique. La motivation reposerait donc plutôt sur une routine systématique de remerciements emphatiques. Nous citons deux fragments parmi d'autres:

> "Cheers, John! You and Podz have really been churning it out. Your hard work is much appreciated!
> Craig."[782]

Ou:

> "Great, however, I was just about to rename the WordPress Widgets Api to WordPress Widgets API. That's a lot of pages to rework, so I'll leave it for now. Great care has to be taken in creating titles like that, especially with subpages.
> This looks fantastic and I'll go through them in the next couple days to clean up any little oddities and such. Thank you so much for doing this. It's incredible!
> Lorelle"[783]

Citer en même temps un article ou un travail intéressant les utilisateurs comporte une double fonction de motivation : c'est à la fois honorer, et motiver d'autres contributeurs pour d'autres contributions localisées sur le même point du Codex[784].

Il est cependant difficile d'évaluer l'efficacité réelle de cette routine largement visible dans les archives : un contributeur travaillant de façon intensive pendant même une courte période se voit loué de façon emphatique[785], ce qui peut suggérer à l'observateur extérieur que cette routine est assez dérisoire. C'est cependant un moyen à coup sûr non négligeable d'acquérir une réputation locale.

Cette routine n'a rien à voir avec ce qui se pratique dans le domaine de l'open-source, qui est plus exigeant pour la qualification technique et le niveau de contribution : c'est l'artiste en matière de programmation qui est valorisé. Le mode de justification serait plutôt celui de la « cité de l'opinion », où celui qui est « grand » est celui qui *devient populaire*.

De telles considérations permettent encore une fois d'infirmer l'idée préconçue qu'il aurait existé un problème d'incitation non-surmontable à la constitution du bien collectif, et montre la viabilité des pratiques de la « cité de l'opinion » qu'on observe aussi sur Wikipedia.

---

[782] Source : http://comox.textdrive.com/pipermail/docs/2004-September/000717.html .
[783] Source : http://comox.textdrive.com/pipermail/wp-docs/2008-September/001766.html .
[784] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-July/000750.html .
[785] Source : http://comox.textdrive.com/pipermail/docs/2004-October/000772.html .

6.7. Synthèse : les liens entre routines

Nous proposons une synthèse des routines précédentes sur un seul tableau, ce qui rend beaucoup plus allusif tout ce qui a été mentionné précédemment. Il en reprend en revanche la *structure*, tout en mettant en évidence d'une part la présence d'un processus de sélection économique par « seuil de satisfaction » des heuristiques, d'autre part les principes de légitimités[786] sous-jacents aux routines qui ont émergé et qui permettent aussi de comprendre la sélection de celles qui sont les plus légitimes. On observe en particulier à droite du tableau que le principe de légitimité sous-jacent à la collaboration massivement distribuée et à Wikipedia se maintient partiellement :

[786] Pour les caractéristiques des cités, nous renvoyons à BOLTANSKI et THEVENOT (1991).

	Routines	Présence de tâtonnements par coûts/avantages permettant la sélection de l'heuristique	*Principes supérieurs de légitimité*		
			Cité industrielle : valorisation du processus, ce qui est performant et fiable.	*Cité inspirée* : créativité, se soustrait à la mesure, le jaillissement de l'inspiration.	*Cité de l'opinion* : émergence de la célébrité, la réputation.
Routines de structuration du contenu	Housekeeping	Le housekeeping était moins coûteux en terme de coordination.	-	-	+
	"Codex week"		+	-	-
	Limitation de la précision du don d'information	Choix qui a permis d'éviter les effets de saturation pour les utilisateurs et de mettre en avant l'intégration du SI	+	-	-
	Limitation de la technicité de l'information		-	-	+
	Standardisation des titres		+	-	-
	Contrôle de l'architecture	Eviter tout désordre et adaptation à la cognition fruste des utilisateurs.	+	+	-
	Routine de design du Codex		+	-	-
	Edit mercilessly	Comparaison approfondie en matière de coûts/avantages du contrôle du processus éditorial.	-	-	+
	Brouillons sur page utilisateur		+	+	-
	Conservation et limitation des types de contenu ajouté	Eviter de trop demander aux volontaires. Coût de coordination avec les informaticiens	+	+	+
Routines de gouvernance de la communauté	Ouverture massive routinisée	Coûts anticipés de la fermeture trop importants	-	-	+
	Complexification de l'édition par "Blanchissage"		-	+	-
	Liste de discussions	Equilibre entre les avantages et inconvénients de chaque canal.	+	+	-
	Onglets de discussions		-	-	+
	Recherche du consensus et de non-conflictualité	Reprise peu coûteuse de ce qui existe par imitation.	-	-	+
	Faire de l'humour		-	+	-
	Routine d'autorité	Equilibre entre les gains de la valorisation individuelle des uns et l'éviction des autres	-	+	-
	Prise de décision par consensus		-	-	+
	Routinisation des projets individuels		-	-	+
	Remerciement emphatique facile	Peu coûteux et évite un "effet d'éviction".	-	-	+

Tableau 3 : synthèse des routines.

Les « + » signifient que la routine observée est cohérente, compatible et favorisée par le principe de légitimité énoncé en colonne. Les « - » signalent que le principe de légitimité est neutre, voire contraire à la routine observée.

6.7.1. Synthèse concernant les cités mentionnées

Un passage des archives qui confirme nos intuitions du premier chapitre de cette thèse montre qu'au moins une personne dans la communauté a pris conscience de l'existence d'une différence de logique d'action entre open-source et open-media. Nous le citons intégralement:

> "This comes up all the time with open media (as opposed to open source) folks. Open source folks have been at it for a good while and have tread this ground. Advertising clauses, vanity links, and such are a no no. It's about the code being free, and such links impinge freedom. Open source folks worry more about maintaining open distribution and keeping the code free. Open media folks, on the other hand, are preoccupied with credit. I think there was a conference recently where open source and open meda folks got together to talk about these cultural differences. It can sometimes be hard to get open media that is compatible with open source. That's becoming a bigger problem as the two intertwine more."[787]

Cependant, une telle citation méritait des approfondissements concrets et des vérifications. Les deux terrains de l'open-source et de l'open-media *ne révèlent pas spontanément leurs logiques de légitimation*, et encore moins leur cohérence interne et leurs oppositions point par point. L'analyse des routines permet de mettre en évidence *trois* principes de justifications sous-jacentes aux faits observés, dont la cohérence renvoie à trois « conventions » au sens de l'économie des conventions. Toutes les routines se rapportant à Wikipedia ou au monde de l'open-media se justifient par le principe de légitimité de la « cité de l'opinion », alors que celles qui proviendraient du monde des « hackers » se justifient par le principe de légitimité de la « cité inspirée »[788] et/ou par celui de la « cité industrielle ». C'est là ce qui donne une cohérence aux logiques d'action observées. Nous proposons un tableau synthétique qui permet de voir la cohérence de l'ensemble par rapport aux principes de justification mentionnés en haut du tableau en question :

[787] Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/001024.html .
[788] « Code is poetry » est le slogan de WordPress. Les programmeurs du libre conçoivent leur activité comme un art et un amusement, aussi exotique que cela puisse paraître.

	Cité de l'opinion	Cité inspirée	Cité industrielle
Principe sous jacent	Opinion de l'autre, célébrité.	Génie créateur, esthétique, beauté.	Professionnalisme et efficacité
Terrain d'application	Wikipedia, encyclopédie.	Open-source, logiciel.	
Membre de la population	« Star », amateur, vedette.	Créateur, ici « hacker »,	Expert valorisé.
Légitimité	Problématique, renommée, notoriété à conquérir.	Non problématique si expertise et contributions	
Principe de décision	Consensus, « pensée dominante », « éviter conflits ».	Dictateur bienveillant, primat de l'artiste, flamming au besoin.	Indéterminées
Principe de sélection	Ouverture à tous, toutes contributions.	Contributeurs « sélectionnés », élitisme.	
Motivations	Problématiques, contributions ingrates	« Fun ».	Indéterminées
Critère de qualité	Accessibilité pour tout utilisateur.	Esthétique des sites et du code	Code qui fonctionne toujours
Type de communauté	Communauté de pratique.	Communauté épistémique.	

Tableau 4 : synthèse des conventions

L'articulation entre deux voire plusieurs cités est l'un des axes d'analyse incontournable des routines car les principes de justifications cités permettent leur sélection. Nous constatons que les routines justifiées par les principes de la « cité de l'opinion » se sont révélées économiquement viables (« edit at will »), moyennant adaptations et compromis.

L'apport analytique de l'économie des conventions est essentiel : notre cas permet de documenter les formes de *légitimité* sous-jacentes aux formes organisationnelles dites « open-source » et « open-media » : ces derniers concepts n'orientent pas spontanément l'analyse vers la question de la légitimation, ce sont des phénomènes plus *macrosociaux* dont la qualification demanderait approfondissements ; l'analyse des modes de justification est en revanche un aspect *microsocial*. Enfin, le principe de cohérence « inspiré » va au-delà du « bon et beau» code : il s'étend à l'esthétique en général, qui est ici importante pour des blogueurs fort attachés à la beauté sur leur blog comme dans leur communauté. Enfin, la mobilisation de deux cités autres que la cité de l'opinion permet de constater qu'il existe *deux* formes de passage d'une pratique à une autre, qu'il fallait distinguer pour bien comprendre le processus.

Signalons que nous ne connaissons aucun travail « conventionnaliste » sur l'articulation entre les cités mobilisées, la cité inspirée et la cité de l'opinion par exemple[789].

Le caractère limité des conflits dans les archives de cette communauté, en particulier la liste de discussions de la documentation, suggère que des *compromis* ont été trouvés alors que les principes de justifications entre les trois cités ne sont pas les mêmes.

6.7.2. Schéma de synthèse des routines

Nous avons donc adapté l'idée de modèle de MILES et HUBERMANN (1994) pour proposer une synthèse des routines et rattacher celles-ci aux problématiques de management. Le but est d'avoir une vue « de haut » de ce qui se passe, pour comprendre comment ces éléments s'articulent. Une routine peut en entraîner une autre.

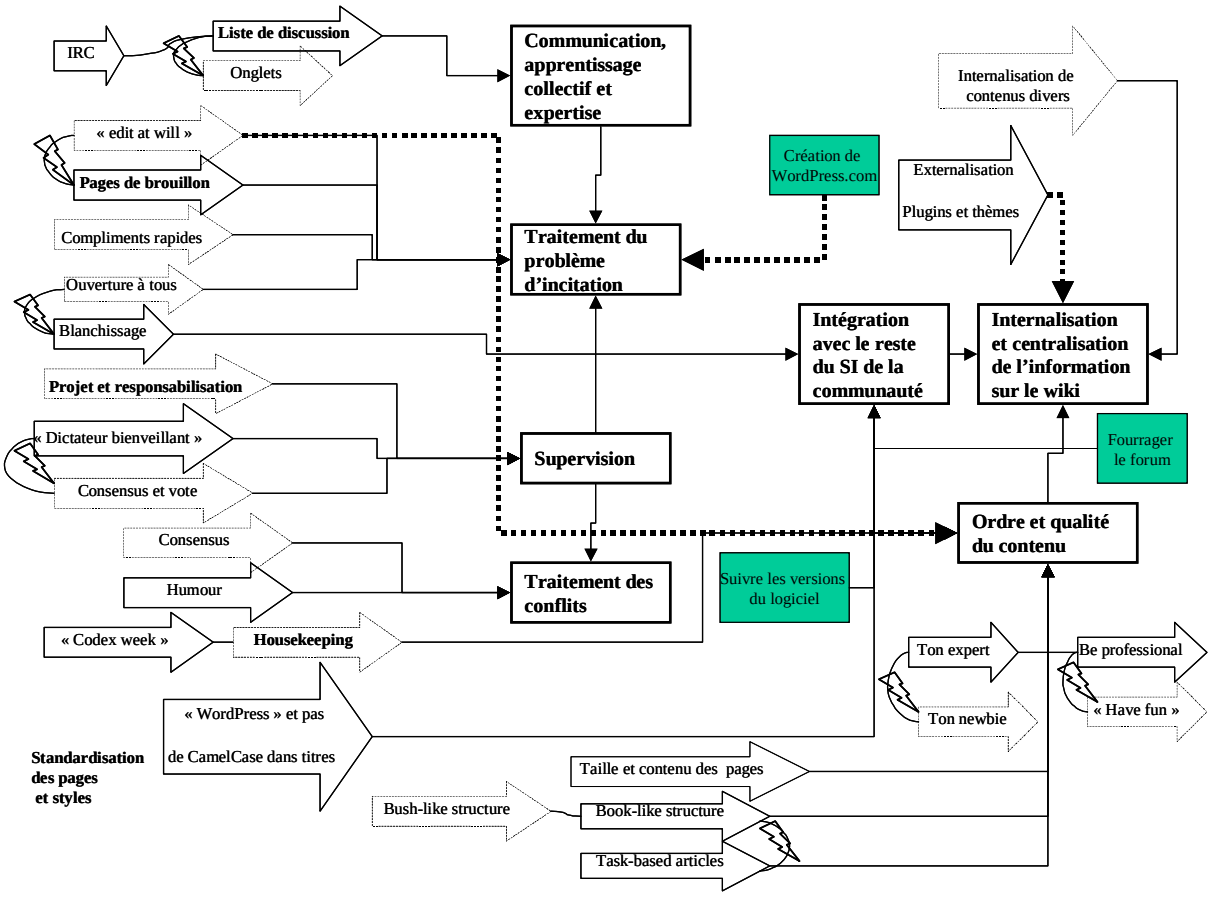


Figure 51 : articulations entre routines et problèmes de management

[789] En revanche, l'articulation des deux cités pertinentes pour la programmation peut poser des difficultés. Mais cela n'entre pas dans notre sujet et il aurait mieux valu pour étudier la question utiliser des sources directement pertinentes, comme les listes de discussions des informaticiens. Un tel travail de raffinement n'est pas à exclure, car il pourrait approfondir ce que propose HORN (2004).

Légende :

- Grosse flèche à trait plein : routine venant de l'open-source.
- Grosse flèche à trait pointillé : routine venant de l'open-media.
- Eclair : zone de tension entre routines.
- Carrés à traits forts : problème de management traité par une routine.
- Flèche à trait continu : action positive.
- Flèche à trait pointillé : action négative.
- Carré foncé : articulation avec le reste du SI.

6.8. Conclusions

Nous avons décrit la morphologie d'une hybridation et tenté de la comprendre. Nous avons montré que les routines observées s'écartaient de façon significative de ce qui était observé dans l'archétype de la « collaboration massivement distribuée », Wikipedia. Nous l'avons observé en particulier en matière de standardisation des pratiques de rédaction, de rigidification de l'architecture du wiki, avec la création des « drafts » sur les pages des utilisateurs, l'exclusion de certains types d'informations, la coordination centralisée sur une liste de discussion. *Cependant*, certaines d'entre elles, comme le « housekeeping », la routinisation de l'ajout de contenu « expert », l'ouverture quasi-totale du wiki aux contributeurs, la routinisation de la décision consensuelle et de l'initiative décentralisée, et la motivation par louanges des contributeurs montrent que les routines associées à ce mode de collaboration restent viables.

Voici quel est l'apport empirique de ce chapitre, d'un point de vue imagé. Ce qui s'est passé est semblable à une plaque tectonique qui se détache d'une autre ; elles finissent ensuite par se heurter lorsque la seconde revient vers sa plaque-mère, des reliefs se forment progressivement, des lignes de subduction locales. Sur certains points, la plaque-mère reste en surface et la plaque-fille disparaît. Ici, ces éléments structurants que sont les routines se maintiennent « en surface » ou disparaissent, si elles ne sont plus viables. Ce qu'on observe, quel que soit le niveau des routines, c'est un équilibrage entre celles qui sont justifiées par la cité « inspirée » et la cité « de l'opinion ». On observe la juxtaposition des deux modes de justification, plutôt que des points de tensions irréconciliables.

Nous avons étudié, pour simplifier, le volet « production » d'une économie de l'attention : la structuration de l'information et les routines de gouvernance qui a une raison

d'être économique ; permettre aux lecteurs d'arriver à trouver l'information le plus rapidement possible (coût de recherche de l'information). Des travaux supplémentaires sur le « management des communautés » étaient bienvenus (O'MAHONY, 2007).

Aborder la gouvernance sous l'angle des routines permet d'observer la collaboration massivement distribuée sous un angle original, dans l'optique de ce que proposait DUMEZ (2009) ; l'une des conclusions est que la liberté d'édition concernant l'architecture du site est *utopique*. Une bonne partie de l'histoire du wiki décrit bien cette impossibilité. L'approche historique des routines permet de comprendre pourquoi. En ce qui concerne à présent la littérature scientifique significative, notre travail permet par exemple de nuancer le passage suivant :

> « Wiki enable instant publication of content. As soon as an author saves the new content, it becomes immediately visible to all readers viewing the page. No intermediary, coordinator, or editor is involved in the publication process" (WAGNER et PRASARNPHANICH, 2007).

Cette citation ne correspond pas à ce qui est observé dans notre communauté[790], où émergèrent justement des routines d'édition (« editorial pattern ») et des formes de gouvernance. Cela ne veut pas dire que l'édition spontanée (« edit mercilessly ») ordinairement associée au wiki n'existe pas, mais notre communauté aurait mis en place une routine qui est un compromis satisfaisant entre deux façons de faire, chacune ayant ses avantages et ses inconvénients.

Notre présent travail peut aussi comporter des apports théoriques et méthodologiques. C'est la première fois qu'on documente l'articulation entre cité inspirée et cité de l'opinion, à notre connaissance, et peut-être la première fois qu'on traite ainsi de telles archives sur le plan qualitatif. Nous avons également tenté une conceptualisation de la notion de routine telle que la demandaient Gigerenzer et Selten.

Notre présent travail malgré son caractère descriptif approfondi comporte des limites, mais à la marge. Ainsi, la documentation des pratiques effectives mériterait des approfondissements. Cette partie ne comporte par exemple que peu d'analyses quantitatives, ce qui peut poser des problèmes pour comprendre l'étendue réelle des comportements des agents sur le wiki, même si nous en avons mentionné. C'est là la contrepartie de l'insistance sur les sources « délibérative », à la frontière de l'écrit et de l'oral, et sur les aspects plus structurels que sont les routines. Ici, nous avons surtout un système et une structure, ce qui est

[790] Cela correspondrait certainement davantage à ce qu'on a pu observer, au moins à une époque donnée à définir par d'autres recherches, dans Wikipedia, mais ce n'est pas notre sujet.

quand même prioritaire (et on voit l'aspect *collectif* de la routine). Nous ne pouvons ici avoir une idée complète de l'efficience des routines ; nous n'en avons une idée que par un retour incomplet d'expérience sur la liste de discussion.

Des travaux ultérieurs consisteraient à savoir dans quelle mesure les routines observées se retrouvent dans d'autres communautés open-source et au-delà. On peut en effet sérieusement se demander si la forme d'hybridation que nous avons observée ne correspond pas à un *cas général*, et non pas à un cas particulier : *en effet, les routines observées ne sont pas là par hasard, c'est ce que nous avons documenté, il s'agit d'une forme d'adaptation à un environnement qui est assez courant, les populations d'utilisateurs des logiciels dans un contexte communautaire. Il est donc probable qu'elles émergent dans d'autres communautés de façon semblable*. C'est là, nous l'avions mentionné, un point crucial en matière de conceptualisation de notre thèse, qui demanderait des recherches supplémentaires. Les idéaux-types se rencontrent peu fréquemment, et des formes diversifiées d'adaptations existent certainement; restent à savoir lesquelles sont viables par sélection. Certains faits vont dans cette direction (émergence de pratiques de gouvernance des wikis, usages des listes de discussions dans les projets…). Il est difficile de répondre à cette question avec notre seul matériau et d'en dire davantage à ce stade. Si c'est cependant le cas, il s'agit d'un apport en théories des organisations (légitimité, hybridation, pratiques…)., i.e. la description *réaliste* d'une forme organisationnelle nouvelle au-delà des concepts exploratoires discutés dans notre second chapitre et des idées reçues.

Chapitre 7 : analyse des formes de collaboration permettant la gouvernance du Codex

7.1. Introduction

L'objectif de ce chapitre est de comprendre davantage la nature de la *gouvernance* de la *collaboration massivement distribuée* par l'étude structurelle de la liste de discussion qui coordonne l'activité sur le Codex. C'est l'une des voies qui permet dans cette thèse de se faire une meilleure idée du fonctionnement de la partie de la communauté WordPress qui contribue au Codex.

La « collaboration massivement distribuée » sur un wiki se caractérise en effet par une population d'individus qui interagissent entre eux soit directement (formation de liens), soit indirectement (via un artefact comme une page de wiki). Ainsi, l'étude structurelle de la liste de discussions de coordination de la documentation (« wp-docs ») peut paraître incongrue : l'essentiel ne serait-il pas dans le wiki ? Ce n'est pas le cas : l'étude qualitative des routines dans un chapitre précédent montre l'importance *centrale* de ce TIC pour la régulation de la communauté. Se concentrer de façon significative et en premier lieu sur cette liste est donc parfaitement justifié. Une thèse de management des systèmes d'information peut se permettre sans difficulté d'étudier un canal de communication (la liste de discussions) de gouvernance d'un SI (le wiki)[791]. Peut-on nous accuser d'éviter la difficulté, étant donné qu'étudier une liste de discussions serait relativement plus aisé que d'étudier un wiki ? On nuancera au vu du travail que demandent un traitement manuel des archives et leur analyse. Enfin, un tel travail n'est jamais apparu dans la littérature rassemblée.

L'étude d'un processus et de ses ressorts cachés est donc ici l'analyse d'une *série chronologique*, celle qui supporte la gouvernance d'un wiki. Nous nous focaliserons sur les aspects quantitatifs, incontournables pour en analyser les aspects structurels qui n'apparaissent pas forcément dans les données qualitatives, malgré l'intérêt de certains fragments. De plus, lorsqu'il est question d'étudier l'auto-organisation sur le Web, il est opportun de tenter une analyse de réseau, ce qui ne va pas techniquement de soi pour un wiki, comme on l'a vu dans la revue de littérature.

[791] Par ailleurs, la nature d'une liste de discussion aurait d'après certains travaux un impact sur ses propriétés (DORAT et al, 2007).

De nombreuses thématiques sont en effet pertinentes pour décrire, comprendre, voire expliquer les caractéristiques de la gouvernance des wikis[792]. Voici les résultats que nous obtenons, que nous proposons de résumer en une intrigue que nous avons volontairement écrite de la façon *la plus simple possible*[793] :

La gouvernance du Codex est un *processus* qui se déroule en deux phases, l'une, avec une forte activité d'une population conséquente d'intervenants, l'autre, d'activité plus faible sur ces deux points et marquée par l'émergence de cycles. C'est un processus aux caractéristiques parétiennes (i.e. avec des distributions exponentielles), marqué par la dispersion et la dissymétrie sur l'essentiel de ses dimensions (démographiques, temporelles, nombre de discussions, concentration des discussions) mais aussi par des corrélations simples entre elles. On observe des phénomènes d'entraînement dits « boule de neige » entre les deux types d'intervenants, les administrateurs et les autres, les non-administrateurs, ce phénomène étant concomitant avec les caractéristiques parétiennes précédentes, en particulier dans la première phase. Les administrateurs ont des caractéristiques d'intervention originales, comme d'interagir plus fréquemment entre eux (homophilie) ou de concentrer sur des périodes significatives l'essentiel des interventions. Le réseau formé par ces interactions est faiblement connecté, à l'exception des pôles multiples que sont les administrateurs qui sont les uns pour les autres des interlocuteurs maniant la réciprocité et crédibles, ce qui n'exclut pas d'autres formes significatives de réciprocité dans le reste de la communauté[794].

Le processus de gouvernance comporte donc des propriétés structurelles marquées caractéristiques des systèmes dynamiques en économie cognitive, et ce, sur une population pourtant assez réduite : il n'est *aucunement vraisemblable* qu'un tel processus *nécessaire à la viabilité du projet* se soit concrétisé sur les parties du wiki prévues pour la discussion. *La compréhension d'un tel processus permet donc de renforcer notre thèse.*

Un plan est une autre façon de structurer ce qui précède. Après avoir présenté notre base de données, nous commencerons par examiner la démographie de la liste de discussion. Nous proposerons un l'examen en coupe (« cross section ») de la série chronologique, poursuivrons par un examen des formes de structuration des interventions, examinerons l'organisation des fils de discussions pour terminer par une analyse de réseau.

---

[792] D'un point de vue simplement factuel, l'étude d'un tel processus permet d'aller au-delà d'une description en « structure en oignon » (CROWSTON et al., 2006) proposée pour *l'ensemble* des participants à un projet : nous apportons ici une description plus précise, sur une *partie* du projet, et prenant en compte sa dynamique historique.

[793] Sur ce point, nous avons suivi les recommandations d'Alain Jeunemaître, qui a aimablement relu ce chapitre.

[794] Au-delà de ce résumé des résultats, nous n'hésiterons pas à évoquer les points où l'échec en matière de recherche de résultats présente un intérêt, en creux.

7.2. La constitution d'une base de données à partir du dépouillement des archives

La liste de discussions « wp-docs »[795] a été analysée de façon quantitative (ailleurs de façon qualitative), de novembre 2003 à décembre 2008, sur plus de cinq ans. Nous avons arrêté le dépouillement pour achever la présente thèse, bien que l'activité de la liste se poursuive en 2009. Elle comprend de novembre 2003 à décembre 2008 inclus 2872 courriers électroniques, 680 fils de discussions, 980 kb d'information, 173 discutants ayant parlé au moins une fois. Pour une sous-question de recherche, le matériau est d'une quantité honorable.

Par comparaison avec d'autres travaux (DORAT et al., 2007 par exemple), c'est une liste de taille plus moyenne. Une originalité immédiatement visible, sur une période plus longue de 5 ans, est de comporter des plages significatives d'activité réduite, quoique ses contributeurs potentiels soient toujours présents, et que l'activité du wiki se poursuive.

L'étude proposée est à la fois statistique et d'analyse des réseaux sociaux. La partie statistique a demandé le dépouillement systématique des archives en ligne de la liste de discussion : quoique pénible et guère ergonomique par comparaison avec ce qui peut se faire avec des outils d'extraction automatique de données, cela n'a occasionné aucune difficulté particulière. La partie d'extraction du réseau a en revanche causé plus de soucis, et nous avons sollicité Jean-Philippe Cointet, thésard au CREA qui travaille sur des problématiques proches des nôtres pour lui demander de concevoir quelques lignes de code pour extraire les données des fichiers-source.

Une capture d'écran de la base de données et les commentaires associés ont été placés en annexe de ce chapitre. Elle peut servir à d'autres chercheurs.

Ces précisions méthodologiques achevées, nous présentons des éléments descriptifs essentiels de la liste de discussion.

7.3. Caractéristiques démographiques de la population des discutants à la liste de discussions « wp-Docs »

Nous abordons la question de la structuration de la liste de discussions en commençant par celle de la démographie. Au-delà de la confirmation de l'existence de deux phases distinctes dans l'histoire de la communauté qui dessinent des contrastes entre deux périodes

[795] Nous incluons les anciennes archives « docs » dans cette liste.

en matière démographique, elle montre qu'il existe au contraire une forme de permanence du ratio administrateurs/non-administrateurs et un niveau d'encadrement d'une stabilité significative. C'est ce que nous examinerons successivement.

7.3.1. Statistiques descriptives

On observe que 173 individus parlent au moins une fois sur la liste sur une période de 5 ans. Nous précisons que d'autres personnes sont abonnées à la liste de discussion, mais restent silencieuses[796]. Si des personnes interviennent sur la liste, il est impossible de les manquer, et il n'y a pas de biais sur ce point.

Voici l'évolution du nombre de participants par mois, que nous différencions entre administrateurs et non-administrateurs :

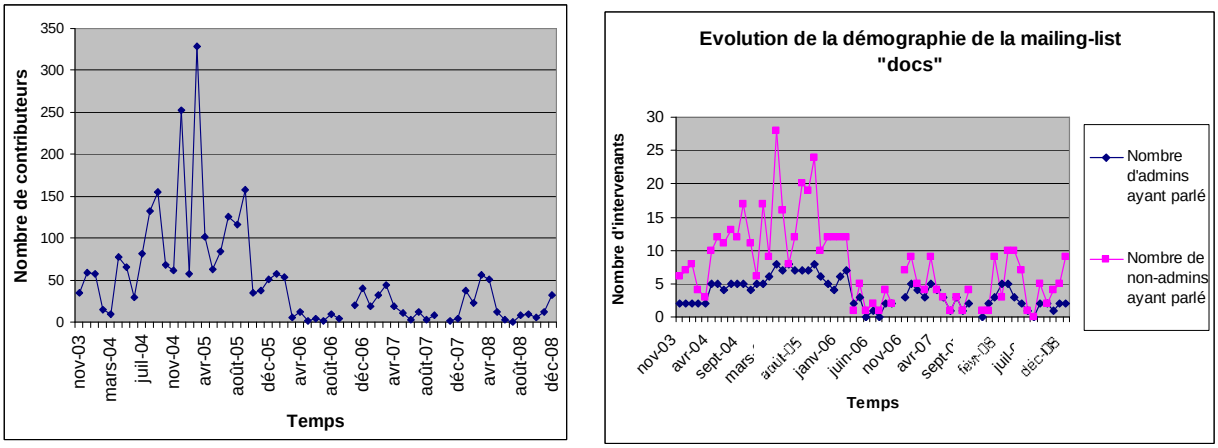


Figure 52 : nombre de contributeurs, différencié entre administrateurs et non administrateurs

On observe des *régularités conformes aux premiers résultats* du chapitre de contextualisation (chapitre 4) :

- La démographie de la liste de discussion comporte deux phases nettement différenciées, la première allant de novembre 2003 à février-mars 2006, la seconde de mars-avril 2006 à décembre 2008. On peut discuter de la périodisation et des points de rupture à quelques mois près, mais elle différencie nettement deux périodes, visuellement parlant, *l'une avec une*

[796] On peut accéder régulièrement à la liste en devenant soi-même un « lurker », mais nous n'avons pas surveillé l'évolution de cette population, ce qui présentait un intérêt relativement limité. On pourra nous le reprocher si cela affecte les conclusions que nous tirons ici ; nous n'avons pas de raison de le penser.

courbe haute fortement indentée, l'autre avec des *cycles plus petits mais nettement visibles*. Mathématiquement parlant, la rupture correspond au point où la série chronologique passe sous le « trend » décroissant sur les cinq années, qui est aussi le point de passage des moyennes mobiles (sur quatre et six périodes) sous ce trend (avril 2006)[797]. Si on découpe les données en ces deux périodes, on observe que la variance et la moyenne du nombre d'intervenants sur la liste sont significativement différentes, 16,89 contre 6,5 pour la moyenne, 7,94 contre 4,18 pour la variance. La discontinuité est donc suffisamment caractérisée.

- Par ailleurs, on peut observer que la rupture entre le premier wiki et le deuxième en juillet 2004 n'est pas spécialement marquée négativement en matière démographique. En revanche, la rupture entre la première liste de discussions et la deuxième qui date de février 2005 se traduit juste par un creux notable en ce qui concerne les non-administrateurs : le noyau d'administrateurs se trouve toujours là. Il est difficile de savoir si cela est significatif.

7.3.2. Formes d'encadrement

La relation entre administrateurs et le reste de la population sur la liste, les non-administrateurs, est un élément a priori pertinent de la structuration de la collaboration. Nous avons donc différencié la *proportion* d'intervenants de chaque sous-population, qui montre une forme de *stabilité* de la relation entre administrateurs et non-administrateurs :

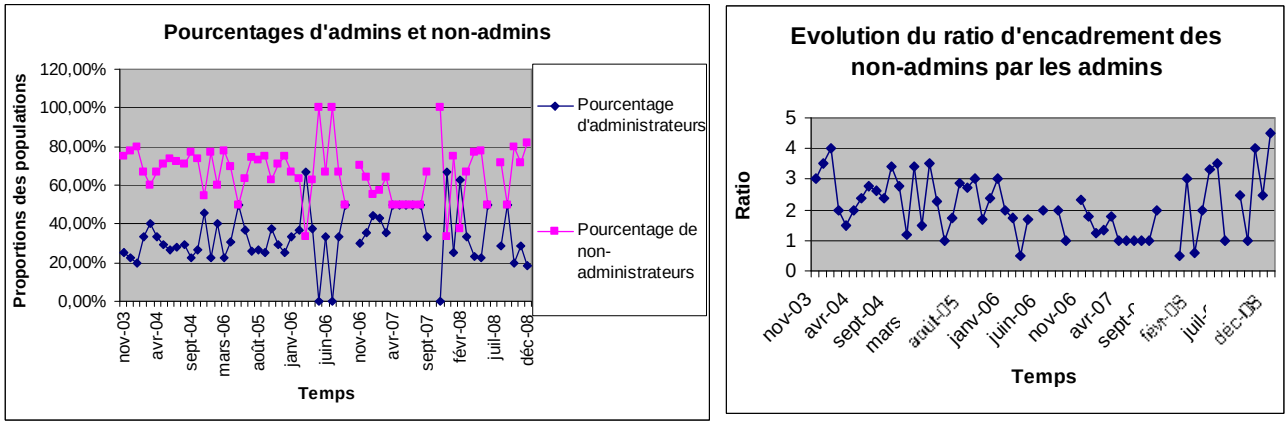


Figure 53 : présentation du rapport entre le nombre d'administrateurs et de non-administrateurs.

[797] Nous n'avons pas surchargé les graphiques avec des courbes supplémentaires.

La moyenne et la variance du taux d'encadrement qui reste assez stable, entre 2 et 4, sont de 2,13 et 0,96. On observe ici la rupture en deux phases mentionnée précédemment avec un ratio qui tombe en dessous de 2, mais elle n'est pas pérenne, le ratio repartant à la hausse. Cependant, même si le ratio comporte d'après ce qui précède une forme de stabilité, la répartition des administrateurs et non-administrateurs suit, comme le montre le graphique ci-dessous, une corrélation qui n'est pas définitivement concluante, ce qui montre cependant l'existence d'un effet d'entraînement dont *le sens* resterait alors à déterminer (nous n'y sommes pas parvenu) :

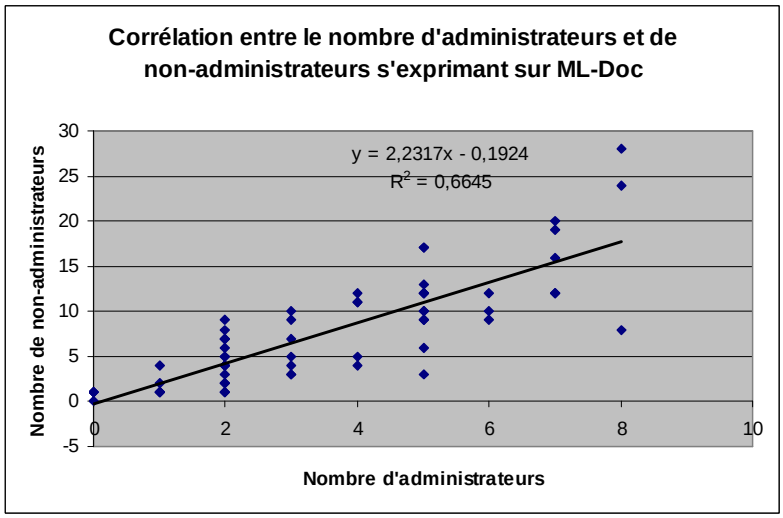


Figure 54 : relation linéaire entre le nombre d'administrateurs et de non-administrateurs

Ce graphique qui permet de mieux visualiser la dispersion entre les effectifs mensuels des deux populations montre d'une autre façon que précédemment qu'il n'y a pas de sous-encadrement significatif : on n'observe sur ce graphique aucune valeur élevée des non-administrateurs en ordonnée correspondant à des valeurs faibles des administrateurs en abscisse (il n'y a pas de points en haut à gauche du graphique).

7.4. Les caractéristiques en coupe (« cross-section ») de la population qui intervient sur la liste de discussion :

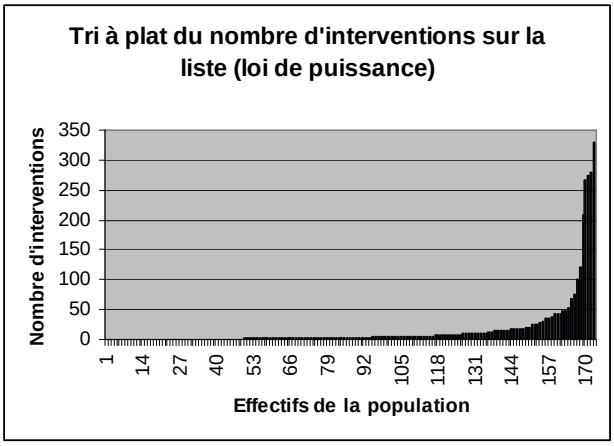
Les statistiques en coupe (« cross section ») de la population d'intervenants sur la liste de discussion sont utiles pour se faire une idée du point d'aboutissement de la série chronologique. Les analyses « en coupe » à partir de la sommation de l'activité de la liste sur

plus de cinq ans (en « bout » des séries chronologiques) offrent des informations indispensables, mais comportant des limites dans une certaine mesure du fait de l'absence de caractère dynamique.

Elles permettent de montrer que l'essentiel des variables descriptives de cette population suivent des lois de puissances, synonymes de fortes disparités (dispersion et dissymétrie) ; de plus, ces distributions sont corrélées ; les individus concernés par l'émergence de ces lois sont en général les mêmes. *Le processus d'intervention sur la liste est donc caractérisé par l'instabilité davantage que la continuité.*

7.4.1. Statistiques descriptives

Cette population se caractérise par des distributions en forme de *loi de puissance* autant en ce qui concerne le nombre d'interventions, le nombre de mois d'interventions (i.e. avec au minimum une intervention), la variance des interventions de chaque individu, et le nombre moyen de courriel par mois, comme le montrent les quatre graphiques qui suivent ci-dessous :



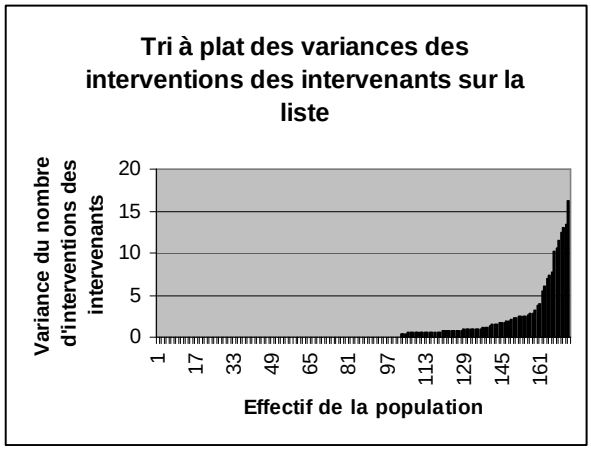
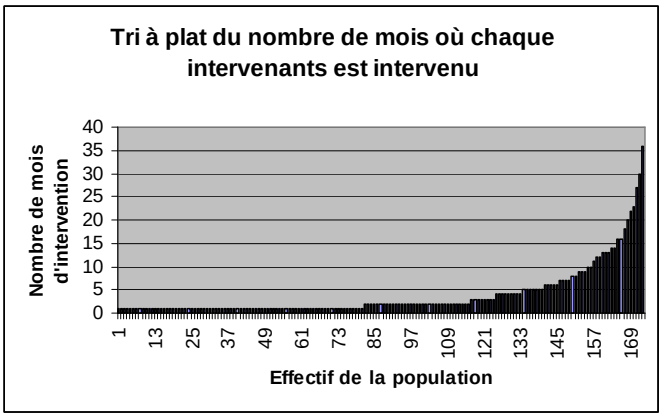


Figure 55 : tris à plats du nombre d'interventions sur la liste et des variances des interventions sur cinq ans

En ce qui concerne le graphique de gauche, qui permet une répartition des 173 intervenants sur la liste, on constate sur les données chiffrées, que l'essentiel des intervenants sont intervenus moins de 10 fois (75%) et qu'un petit noyau (5,8%) l'a fait plus de 50 fois, seuil de

rupture *graphique* d'où part l'emballement exponentiel observé[798]. Sur le graphique des variances, on observe un seuil de rupture graphiquement semblable ; et effectivement, 6,9% des intervenants ont une variance supérieure au seuil de rupture de 4, un chiffre semblable au précédent. Pour les autres graphiques ci-dessus reprenant les dernières variables, on a 10,98% pour le nombre de mois[799], et 8,09% pour le nombre de messages par mois.



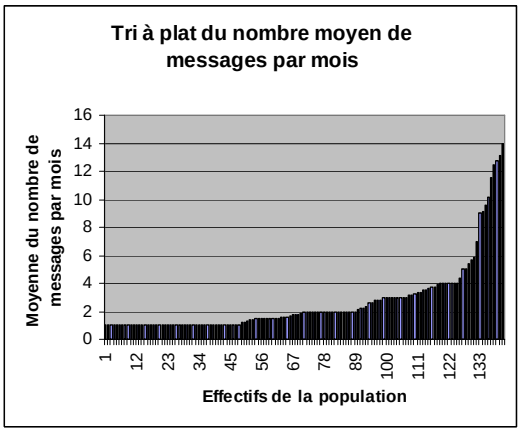


Figure 56 : tris à plat du nombre de mois où chaque intervenant est intervenu, et du nombre moyen de messages par mois

On observe donc à chaque fois, pour plusieurs indicateurs statistiques, une rupture, sous forme « d'accélération de la montée » visuellement, autour de 6-10 % de la population. Une variable, celle du nombre de mois d'interventions comporte une évolution plus douce, ce qui peut se comprendre, étant donné qu'il suffit de participer une fois un mois donné pour faire croître le chiffre de la variable en question.

7.4.2. Des formes de dissymétrie à l'intérieur des historiques d'interventions des membres de la liste de discussion.

Les graphiques précédents montraient des contrastes entre les intervenants répartis selon des caractéristiques exprimées par des variables usuelles. Ils permettaient de constater des différences plus ou moins marquées entre deux populations différentes. Examinons maintenant les contrastes, i.e. les dissymétries à l'intérieur des historiques des interventions des uns et des autres. On part pour cela de la médiane pour arriver à l'écart moyenne-médiane :

[798] Ce type de seuil peut se discuter, mais il est pratique visuellement.
[799] Le point de rupture ne va pas de soi visuellement.

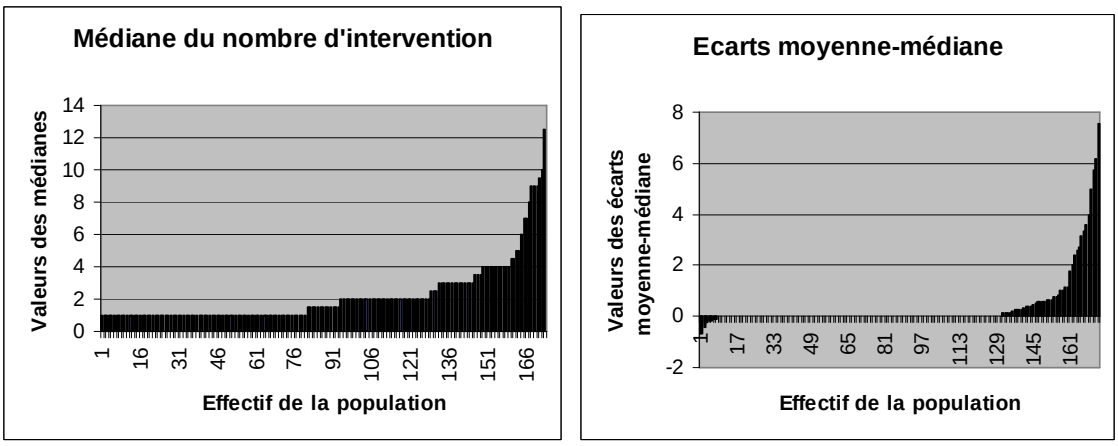


Figure 57 : médiane et écart moyenne-médiane dans la population des intervenants

Là encore, on observe ci-dessus que la population totale comporte une petite sous-population avec des dissymétries significatives au niveau du nombre de leurs interventions.

7.4.3. Quelques corrélations significatives

Les statistiques descriptives précédentes issues du classement des intervenants, indispensables, ne disent cependant absolument pas si les variables en question sont corrélées ou pas, et implicitement, et ne permettent pas de dire si la population avec des valeurs fortes est la même pour chaque variable observée. C*ertaines* des variables étudiées ci-dessous sont corrélées, *ce qui renforce le principe général de dispersion que nous décrivons dans cette partie* :

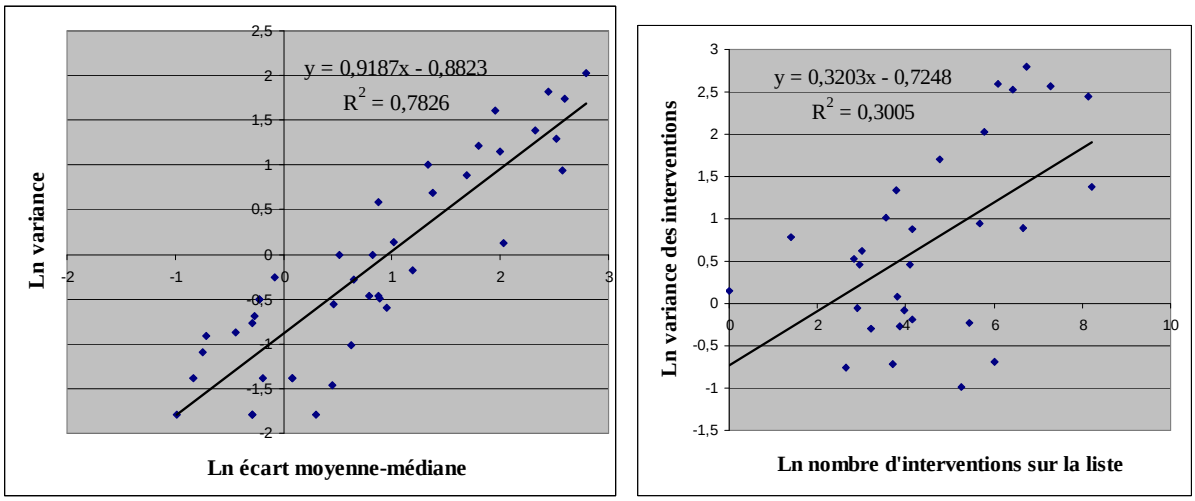


Figure 58 : corrélation entre l'indicateur de dispersion (variance) et l'indicateur de dissymétrie (écart moyenne-médiane) et entre variance et nombre d'interventions

On observe ci-dessus à gauche une corrélation entre l'indicateur de dissymétrie et l'indicateur de dispersion, et à droite entre le chiffre brut des interventions et l'indicateur de dispersion. Il existe aussi une relation significative entre la moyenne des interventions et la variance, avec un R2 de 0,68, mais nous ne donnons pas le graphique ici. Plus un intervenant intervient sur la liste (nombre d'intervention brut ou moyenne), plus la dispersion (variance) de ses interventions augmente. Nous avons par ailleurs cherché des corrélations entre écart moyenne-médiane et chiffres bruts et moyens du nombre d'interventions, mais cela n'a rien donné de concluant. En revanche, on trouve une régularité entre le nombre de mois d'intervention et le nombre d'interventions, avec les valeurs à l'échelle logarithmique, ce qui redonne quelque peu de stabilité au processus.

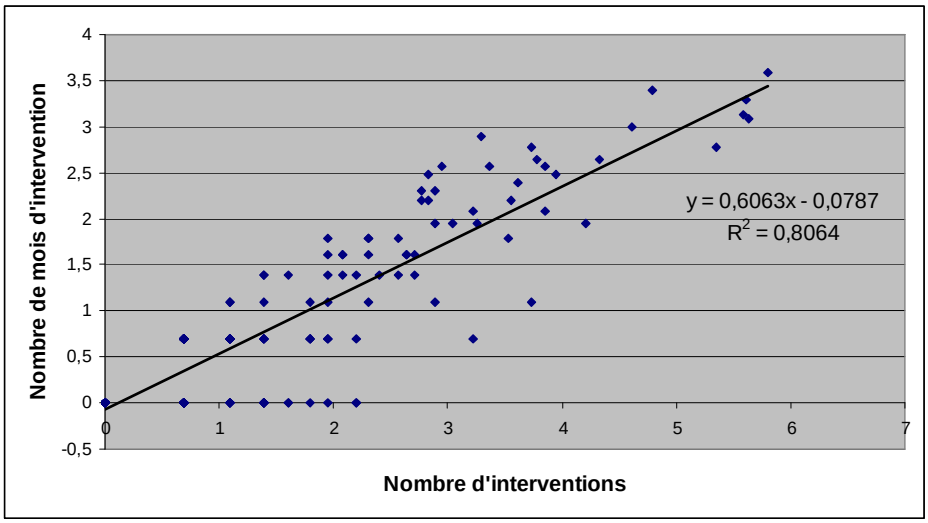


Figure 59 : le nombre d'interventions de chacun est corrélé au nombre de mois d'intervention (échelle logarithmique)

7.5. La structure dynamique des interventions sur la liste de discussions

Après avoir étudié les caractéristiques démographiques et les caractéristiques en coupe de la liste de discussion, nous tentons à présent de comprendre la dynamique quantitative des interventions de cette population d'intervenants.

D'une part, nous montrons ici les spécificités des interventions des administrateurs et non-administrateurs, les premiers étant les intervenants les plus prolifiques, même si leurs interventions moyennes diminuent sur les cinq années étudiées.

D'autre part, nous décrivons un phénomène d'entraînement de type « boule de neige » ou d'avalanche, et ses présentes spécificités : les interventions suivent une loi de puissance, les administrateurs entraînent les autres interventions, l'augmentation du nombre d'intervenants augmente l'intensité des interventions, les interventions des administrateurs et non-administrateurs sont corrélées, et ces caractéristiques se traduisent par de fortes dispersions et dissymétries conjointes… comme dans une avalanche qui, en grossissant, créerait en son front avancé des boules de neige de taille de plus en plus imposantes et une multitude croissante de boules plus petites. C'est là l'un des aspects dynamiques du processus de collaboration.

7.5.1. Statistiques descriptives

On peut montrer, à partir d'un classement par catégorie des intervenants en fonction du nombre de leurs interventions, que la forme des contributions à la liste de discussions suit, comme sur le wiki, une loi de puissance ; un petit nombre de contributeurs significatifs produit l'essentiel des interventions :

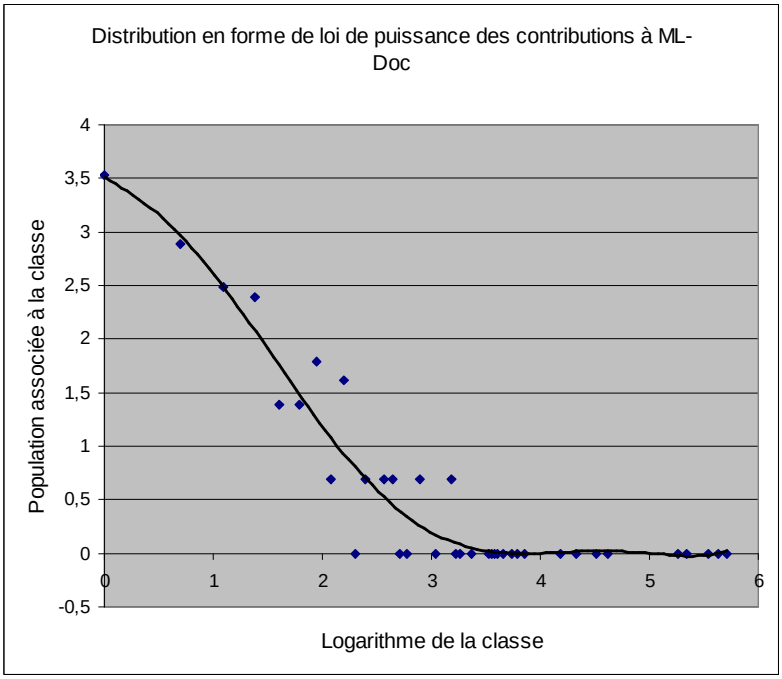


Figure 60 : distribution en loi de puissance des interventions sur la liste de discussions[800]

[800] Le graphique avec des points permet de mettre en évidence les valeurs où ln=0, ce qui n'est pas possible avec des bâtonnets. Cela explique cette présentation peu usuelle.

Si on traçait sur ce graphique une droite qui irait de 3,5 en abscisse à 3,5 en ordonnée, ce qui correspond à la définition de la distribution de la loi de puissance, on observerait qu'une telle droite minimise l'essentiel de la variance de la partie à gauche du graphique ; en revanche, la partie droite s'écarte de façon significative du modèle, comme si un phénomène de « course » au plus grand nombre d'interventions s'observait. On pourrait pour raffiner utiliser comme indicateur de l'écart entre la distribution idéale de la loi de puissance et la distribution observée la différence entre le R^2 de la courbe polynomiale ici présente sur le graphique et celui de la droite décrite précédemment.

L'hypothèse de l'existence d'une distribution en forme de loi de puissance est donc encore une fois pour l'essentiel vérifiée (elle n'est pas « parfaite »), alors que ce type de distribution est récurrent entre autres sur le Web. Cependant, il était nécessaire de souligner qu'on l'observe *déjà* pour une population de taille réduite (173 individus). Le nombre relativement limité d'individus de la population rappelle aussi qu'observer une distribution parfaite n'est guère probable.

Il n'est cependant pas mécaniquement nécessaire que les interventions ici se structurent de la même façon que sur le wiki, où le même type de distribution est observé: en particulier les plus gros contributeurs ici et là-bas ne sont pas forcément les mêmes par exemple.

Nous différencions à présent en matière d'interventions entre les populations des administrateurs et non-administrateurs :

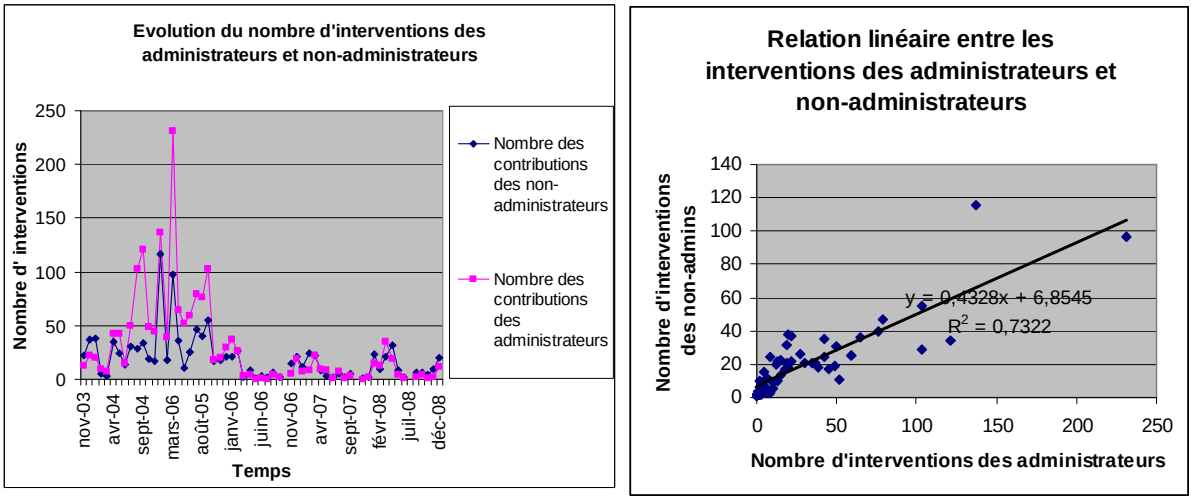


Figure 61 : séries chronologiques de l'évolution du nombre d'interventions des administrateurs et mise en évidence d'une relation linéaire entre les interventions des deux populations

Nous avons déjà évoqué l'évolution globale agrégée des interventions sur la liste de discussions. On peut, comme les séries chronologiques de la démographie de la liste, séparer son histoire en deux phases. L'évolution des interventions sur la liste de discussion comporte d'un point de vue quantitatif deux phases nettement différenciées, séparées par une période de rupture délicat à situer, mars à juin 2006 ; la première irait de novembre 2003 à mars 2006, la seconde de juin 2006 à décembre 2008. Si on découpe les données en ces deux périodes, on observe que la variance et la moyenne sont significativement différentes, 84,75 et 16,09 pour la moyenne, 69,87 et 15,56 pour la variance. *Les contrastes sont encore plus marqués que pour les valeurs associées à la démographie*.

Par ailleurs, la rupture entre le premier wiki et le deuxième en juillet 2004 est marquée négativement en matière d'interventions, avec une baisse des interventions dans les deux populations, passant de 42 à 15 et de 24 à 14. La rupture entre la première liste de discussion et la deuxième qui date de février 2005 se traduit par un effondrement des discussions après un mois de janvier 2005 sans aucune communication.

Sur le graphique de droite *ci-dessus*, on observe donc une corrélation significative entre les contributions des uns et des autres, malgré des irrégularités. Une autre question est de savoir dans quel sens la relation joue.

On remarque ensuite, au-delà d'impressions de similarité sur le premier graphique ci-dessus à gauche, *que les deux populations ont leur dynamique propre en matière d'interventions*; ainsi, la part d'interventions des administrateurs suit une pente décroissante après avoir dominé les échanges de novembre 2003 à mars 2006 (approximativement) ; cette pente n'était pas forcément visible sur les précédents graphiques, en particulier sur la seconde phase :

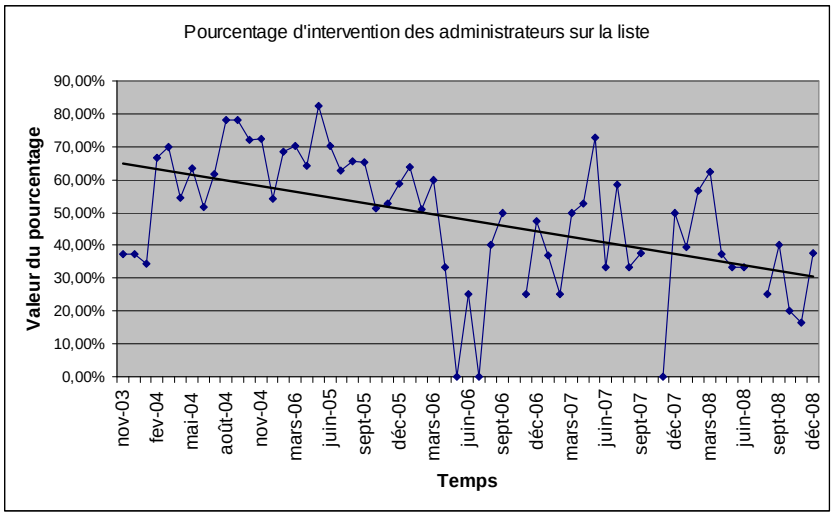


Figure 62 : pourcentage d'interventions des administrateurs sur la liste

On observe un étiage marqué entre mars 2006 et septembre 2006, mais il est difficile de lui attribuer définitivement une signification, même si cela est tentant : le débit trop bas de la liste occasionne un biais qui interdit d'attribuer la même signification aux pourcentages en fonction des périodes.

Cette distribution comporte une signification sujette à discussion. Une hypothèse prioritaire à tester dans la communauté WordPress était celle de la « montée de la bourgeoisie » de KITTUR et al. (2007), mentionnée dans la revue de littérature, correspondant à un renouvellement de la population d'intervenants. Ici, un renouvellement partiel de populations d'intervenants est effectivement observé dans la liste de discussions avec la chute du pourcentage d'interventions des administrateurs, mais nous ne pouvons savoir s'il permet de corroborer significativement l'hypothèse de Kittur, étant donné que la liste de discussion baisse de régime à partir de début 2006, ce qui pourrait être une modalité d'explication alternative au phénomène observé.

En ce qui concerne la relation entre les effectifs mensuels de chaque population et leurs contributions, on observe là encore des relations linéaires différenciées; elle est marquée pour les non-administrateurs, mais nettement moins pour les administrateurs, dont les interventions sont nettement plus dispersées :

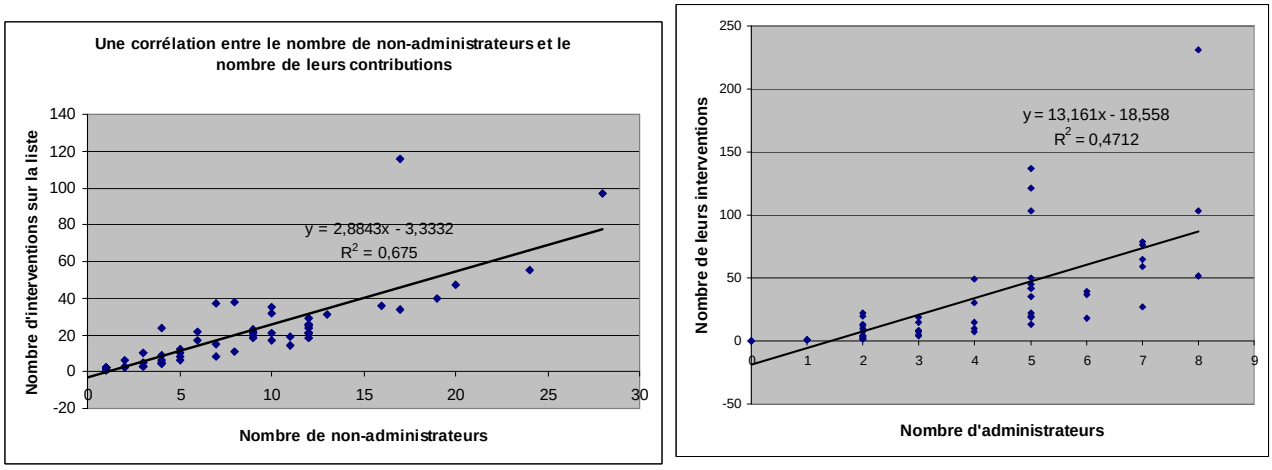


Figure 63 : on observe des corrélations entre le nombre des administrateurs et non-administrateurs et le chiffre de leurs interventions

De même, la baisse du trend des interventions sur les cinq ans ne permet pas forcément de voir la baisse du nombre moyen de messages par intervenants, présentée ci-dessous. L'origine de la baisse de cette moyenne est claire sur le graphique de droite ci-dessous, il s'agit de la baisse de la moyenne du nombre d'interventions des administrateurs, la

moyenne des interventions des non-administrateurs étant d'une remarquable stabilité. C'est là l'utilité de différencier les deux variables :

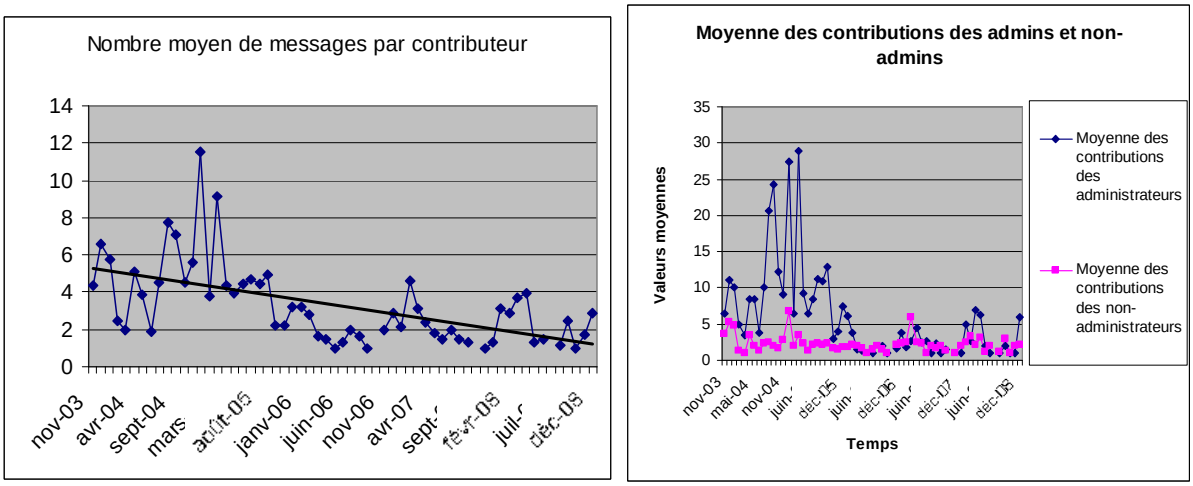


Figure 64 : séries chronologiques du nombre moyen de message par contributeur, différencié à droite entre administrateurs et non-administrateurs

On constate donc encore *deux phases* avec les valeurs moyennes : l'une partant de novembre 2003 et se terminant en mars-avril 2006, où la moyenne des contributions des administrateurs est nettement supérieure à celle des autres, et l'autre, à partir d'avril, où les deux moyennes sont semblables. On observe donc ici de fortes spécificités dans les deux populations

Cependant, les graphiques précédents, s'ils permettent de mieux comprendre la dynamique des interventions et les spécificités des deux sous-populations, ne sont pas suffisants. Ils cachent d'autres caractéristiques structurelles que nous allons commenter à partir des deux graphiques ci-dessous :

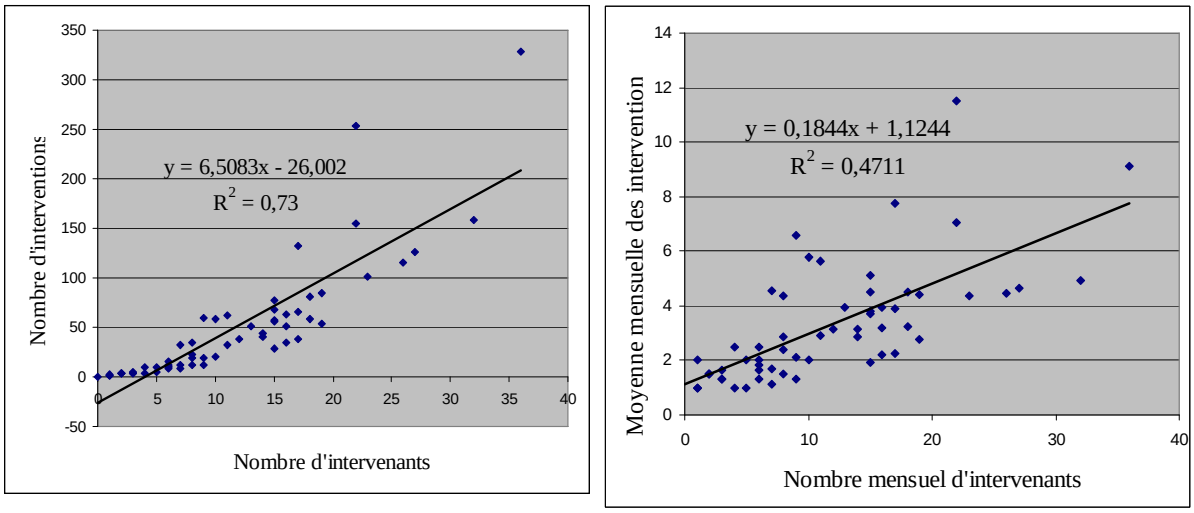


Figure 65 : le nombre d'intervenants fait augmenter le nombre et la moyenne des interventions

Le graphique de gauche ci-dessus, montre que la relation entre le nombre d'intervenants et le nombre d'interventions est linéaire et croissante, voire peut-être exponentielle[801]. Cependant, ce n'est pas la seule relation structurelle observée : sur le graphique de droite ci-dessus, le nombre d'intervenants fait clairement augmenter la moyenne mensuelle du nombre d'interventions, et ce, de plus, de façon linéaire.

Comment interpréter cela ? L'évolution du nombre moyen de messages par contributeurs pourrait obéir à une relation structurelle, qui dépendrait de la taille de la population mensuelle d'intervenants. Une interprétation serait que, « plus il y a de monde, plus on aurait propension à parler », ce qui correspond à un *effet d'entraînement*.

Cet effet d'emballement comporte cependant quelques limites. La première apparaît en constatant l'hétérocédasticité marquée sur les deux graphiques. La seconde est que la relation fonctionne de cette façon avec la population totale, mais pas à l'intérieur de chaque sous-population : sur ce point, nous n'ajoutons pas les graphiques conçus avec les mêmes variables que dans ceux commentés précédemment. *C'est donc l'interaction entre les deux populations qui créé cet effet*.

7.5.2. Les caractéristiques du noyau central des intervenants

Après avoir abordé les interventions dans leur ensemble, nous allons nous concentrer désormais sur les plus gros intervenants pour mieux comprendre l'effet d'entraînement que nous étudions : après identification de cette population, nous allons montrer que des croisements de données corroborent leur rôle décisif dans ce dernier processus.

Voici la distribution des pourcentages des interventions des plus gros intervenants[802], elle montre que l'essentiel des interventions correspondent à celles des quatre plus gros intervenants :

[801] Une relation non linéaire croissante améliore le R2 mais serait trop dépendante des valeurs extrêmes. Nous ne l'avons pas retenue.
[802] L'extraction de cette variable se fait avec la fonction adéquate sous Excel à partir de la population totale.

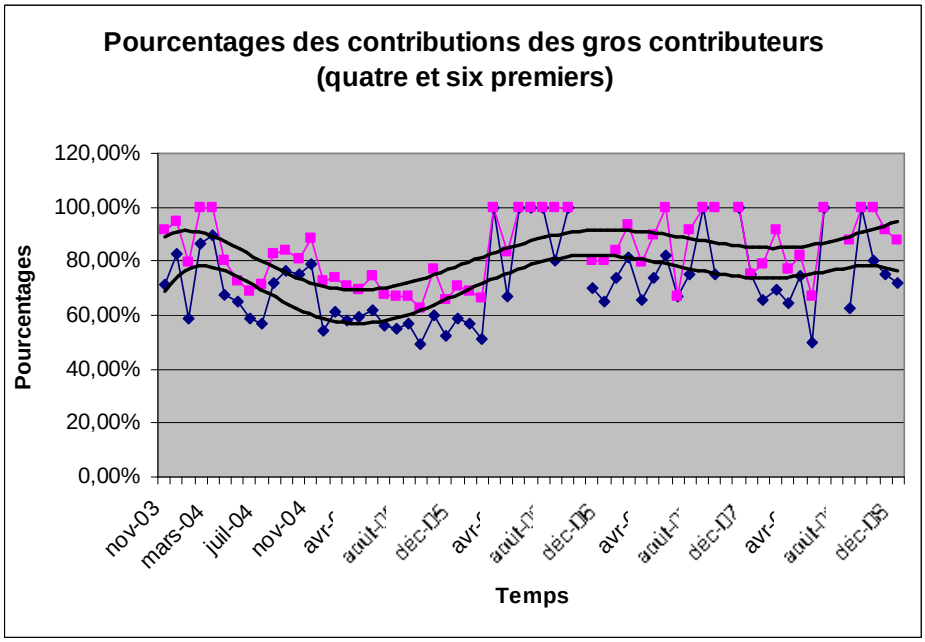


Figure 66 : pourcentage des interventions des quatre et six plus gros intervenants mensuels

La courbe du haut représente le pourcentage de la sommation des contributions des six plus gros contributeurs[803], la courbe du bas le pourcentage de la sommation des contributions des quatre plus gros contributeurs[804]. On remarque qu'il s'agit clairement d'une *régularité structurelle*, on observe toujours un noyau d'intervenants qui effectue *plus de la moitié* des interventions sur la liste de discussions, et ce, même lors des phases de forte activité. Les proportions plus élevées à partir de février 2006 peuvent se comprendre, étant donné que les contributions se sont effondrées.

Le graphique montre l'existence *d'un processus d'entraînement* dans les interventions d'après la chronologie, car la phase de diminution de la part des plus gros intervenants est aussi celle où les interventions dans la communauté, en particulier celles des non-administrateurs est la plus forte. Il reste cependant à en qualifier plus précisément la structuration, ce que tentent les deux graphiques ci-dessous :

[803] Nous ne savons pas de qui il s'agit, à ce stade du raisonnement du moins.
[804] La pertinence des seuils de 4 et 6 a été trouvée par tâtonnements, qui indiquaient qu'il était inutile d'aller plus loin.

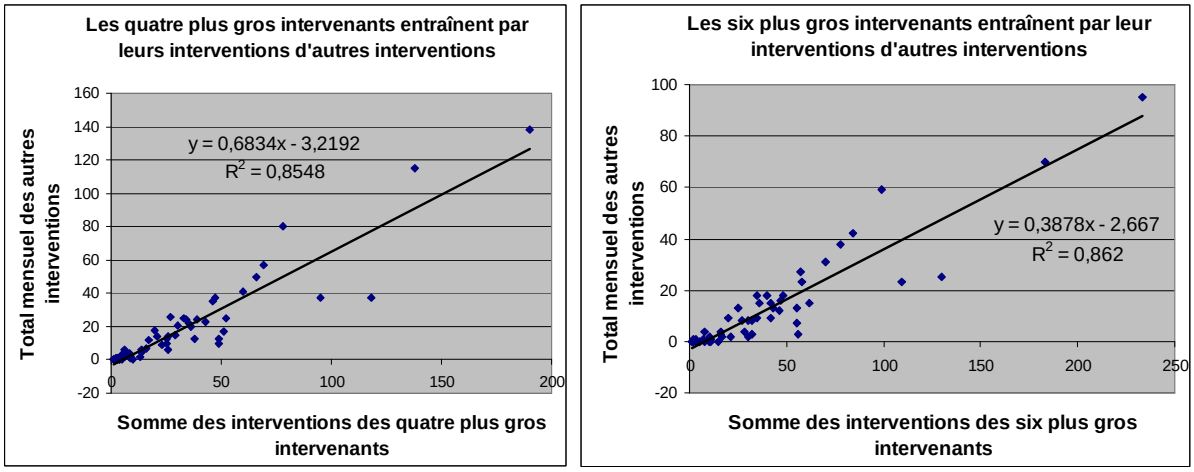


Figure 67 : les plus gros intervenants (quatre et six plus gros) entraînent les autres interventions

Les « autres interventions » ici en ordonnée correspondent au total des interventions *diminué* de celles des sous-populations de gros contributeurs mentionnées en abscisse, bien entendu. Il est ensuite nécessaire d'identifier les plus gros contributeurs. La même technique peut y aider, et montre que le nombre d'interventions des administrateurs est entre quatre et six, *et que ce sont bien eux qui provoquent cet effet d'entraînement* :

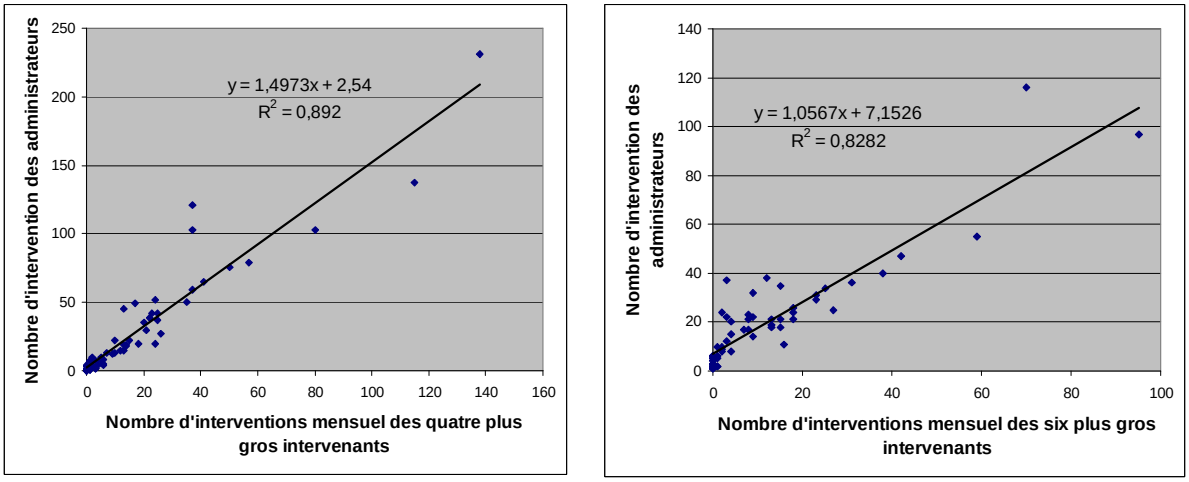


Figure 68 : la corrélation entre le nombre d'interventions des six et quatre plus gros intervenants et le nombre d'interventions des administrateurs montre que ce sont eux qui forment l'essentiel des plus grosses contributions

7.5.3. Les caractéristiques mensuelles de dispersion et de dissymétrie

Il est possible d'analyser encore plus finement le processus de « boule de neige ». Plus les gros intervenants en question interviennent, plus cela provoque une dispersion des valeurs du nombre d'interventions sur la liste :

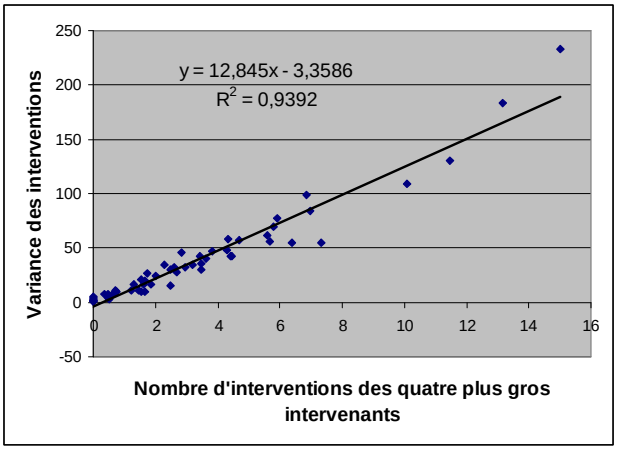


Figure 69 : le nombre d'interventions des quatre plus gros intervenants augmente la variance des interventions[805]

Il est nécessaire en conséquence de présenter cette variable. Voici comment évolue la variance des interventions :

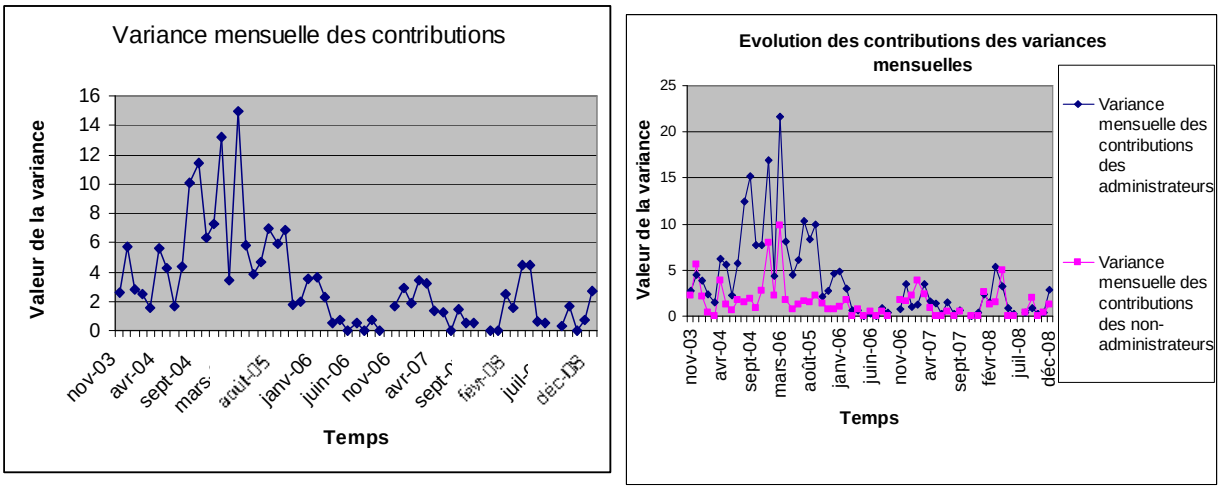


Figure 70 : baisse de la variance mensuelle des contributions et différenciation entre variance des interventions des administrateurs et non-administrateurs

[805] Nous ne présentons pas le graphique avec les six plus gros, cela ne modifie la relation que de façon marginale.

En ce qui concerne la dispersion des interventions, il est possible de faire des remarques semblables à d'autres indicateurs: la spécificité des administrateurs en matière d'interventions est localisée sur une période précise, novembre 2003 à mars 2006. Les approfondissements statistiques sur la liste de discussion confirment nos premières conclusions qui indiquaient une rupture début 2006.

7.5.4. Dispersions et dissymétrie dans les interventions sur la liste

On observe de fortes dissymétries et dispersions mensuelles, se traduisant ici par une corrélation entre écart moyenne-médiane et variance, alors que la première aurait pu augmenter tout en laissant constante la seconde :

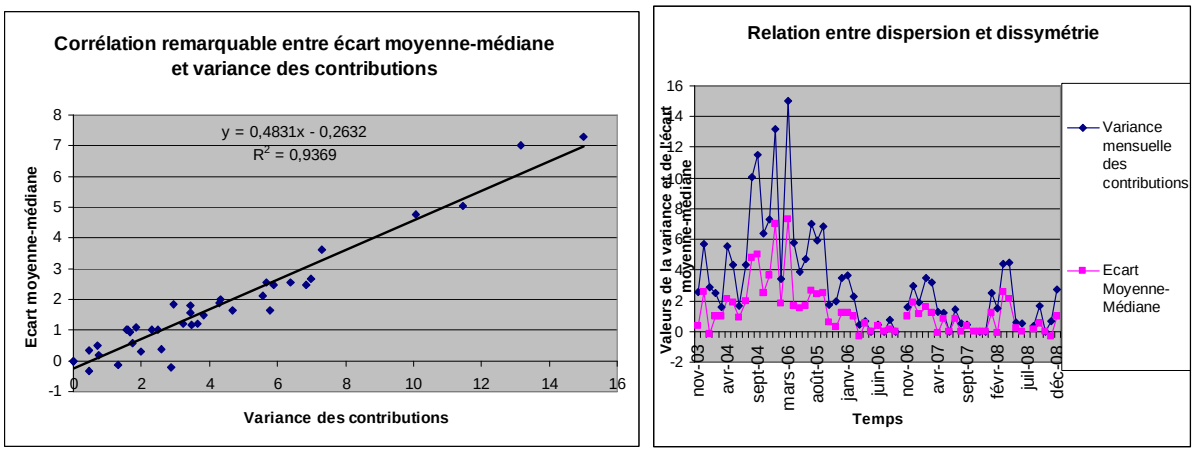
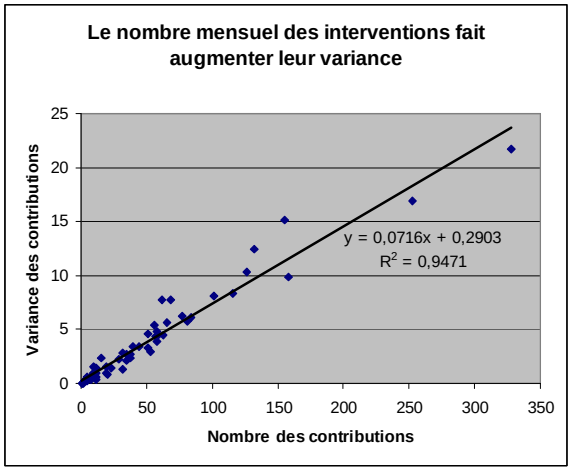


Figure 71 : corrélation entre la variance des contributions et l'écart moyenne-médiane (présentation avec régression et série chronologique des valeurs)

La relation entre les deux indicateurs[806] est forte, avec un R2 de 0,93. Ce type de phénomène est significatif, car cela montre que nous sommes ici à l'opposé d'un monde Gaussien, où la variance peut augmenter alors que l'écart moyenne-variance reste stable. Ici, c'est justement tout le contraire, alors que ce type de phénomène n'était *absolument pas nécessaire*. Ces questions de dissymétrie et dispersions pourraient être une caractéristique significative pour

[806] La variance est un indicateur de dispersion, l'écart moyenne-médiane est un indicateur de dissymétrie.

comprendre la gouvernance de l'organisation massivement distribuée, qui ne serait donc *pas du tout caractérisée par l'uniformité*[807].



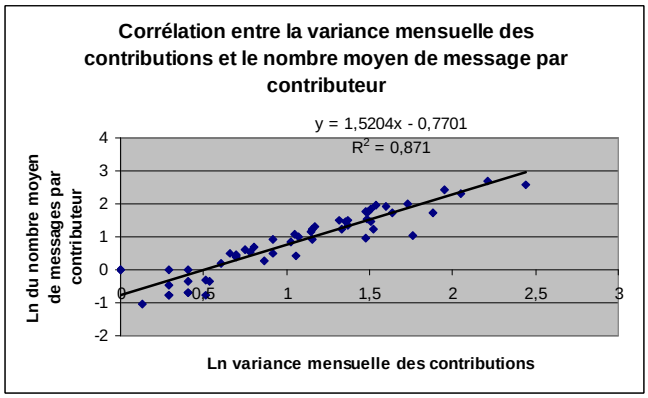


Figure 72 : corrélation entre la variance mensuelle des contributions, le nombre de contributions, et le nombre moyen de messages par contributeur

Un processus d'avalanche crée un effet « boule de neige » : plus le nombre de grains de neige est important, plus la taille d'une petite partie d'entre eux augmente ; de même ici, le nombre de contributions fait augmenter les caractéristiques de dispersion.

7.5.5. Relation entre démographie et dispersion des interventions

En revanche, les caractéristiques démographiques sont moins convaincantes pour comprendre ce type de phénomène d'entraînement, même si on peut observer une relation visible, ce qui laisserait entendre que l'effet d'entraînement est interne ou *endogène aux interventions*, et plus indépendant du *nombre* d'intervenants :

[807] Trouverait-on des régularités semblables dans d'autres communautés, ou ailleurs dans la communauté, par exemple sur la liste de discussion des hackers?

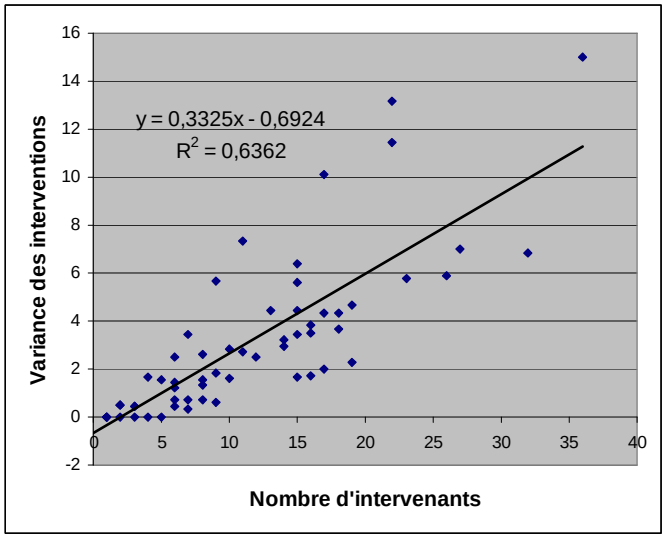


Figure 73 : corrélation entre le nombre d'intervenants et la dispersion des interventions

La démographie peut ainsi avoir un rôle pour comprendre la dispersion des interventions. On constate que l'augmentation du nombre des intervenants par mois augmente la dispersion, mais c'est le nombre de non-administrateurs qui est la variable la plus significative.

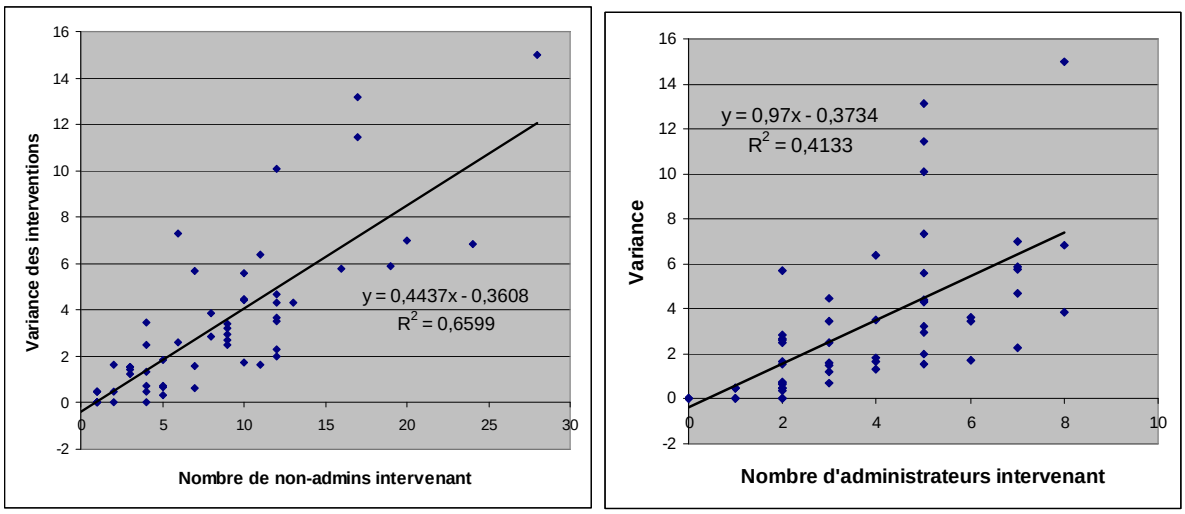


Figure 74 : mise en relation de la variance des interventions et de la démographie des catégories d'intervenants

Le nombre de non-administrateurs intervenant est corrélé avec la variance des interventions. La relation apparaît mieux qu'avec le nombre des administrateurs.

7.6. Comment comprendre les interactions internes aux fils de discussions ?

Nous venons d'examiner la dynamique d'avalanche des interventions. Mais il s'agit d'un processus qu'on peut qualifier de « brut », à partir de variables agrégées, qui fait l'économie du détail plus fin de la dynamique des relations.

Une telle étude plus fine est souhaitable pour comprendre le processus de collaboration pour lui-même, mais il peut compléter aussi la compréhension de la précédente dynamique d'avalanche. *Nous montrons ici en effet, après quelques approches exploratoires, que la meilleure explication de la structuration des contributions de la liste de discussion est celle de l'existence d'un phénomène d'« homophilie » entre catégories d'intervenants*, une catégorie d'intervenants pouvant en attirer une autre à l'intérieur des fils. Nous discutons enfin des explications alternatives à ce phénomène qui ne sont en définitive pas suffisamment convaincantes.

La structuration des fils donne des informations sur la structuration des discussions[808], il était nécessaire de les signaler, même si la régularité la plus significative ne vient pas immédiatement dans cette partie.

Cette partie est la dernière qui se limite à une approche quantitative classique ; nous poursuivrons après dans une dernière partie l'analyse fine de la collaboration par une analyse de réseau qui la complètera.

7.6.1. Statistiques descriptives de l'initiation de fils de discussions et premières approches

Nous commençons par quelques éléments descriptifs rapides.

Le graphique de droite ci-dessous montre que la taille mensuelle moyenne des fils chaque mois est susceptible de *variations significatives*. Cette variable pourrait être caractérisée par des cycles de huit mois dont les sommets et les creux varient entre 5 et 2 courriels par fil en moyenne, à quelques pics et étiages près ; la moyenne globale du nombre de courriels par fil est de 3,4, la moyenne entre 2 et 5 approximativement. Il est cependant difficile d'en dire plus avec les seules données quantitatives, mais il s'agit de variations

[808] Dans un registre de cognition distribuée, on peut citer des travaux comme CONEIN (2003), même si sa problématisation est différente de la nôtre.

significatives de l'intensité des discussions en matière de collaboration, ce qui justifie de s'intéresser au sujet et de retrouver par la suite cette variable[809] :

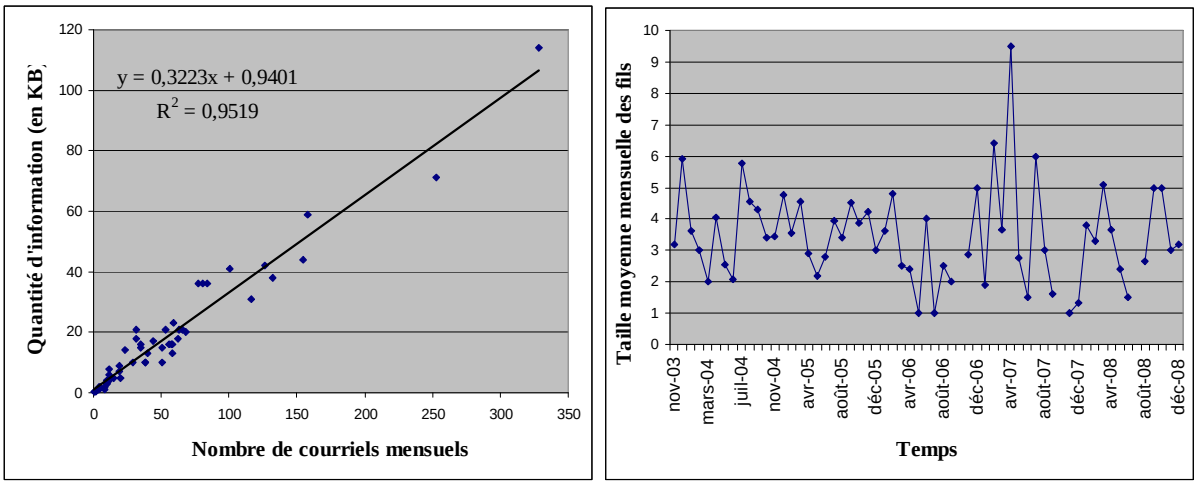


Figure 75 : série chronologique de la taille mensuelle des fils et relation linéaire entre nombre de courriels et quantité d'information

Le graphique de gauche montre en revanche que la relation structurelle entre le nombre de courriels et la quantité d'information qu'ils portent est linéaire, ce qui était attendu à un tel niveau agrégé. Ce n'est pas sur ce point qu'on observera des variations significatives à étudier.

De même, on observe dans les deux graphiques ci-dessous une régularité structurelle entre le nombre total de courriels par mois et le nombre de fils de discussions. La structuration de la liste de discussion obéit à une régularité au niveau agrégé : plus le nombre de discussions augmente, plus le nombre de fils augmente, ce qui traduit *une régularité dans la structuration des interactions*. Ce n'est pas un résultat original, mais il était nécessaire de le signaler.

[809] Cette analyse permet aussi de *compléter celle des rythmes de la communauté travaillés* au chapitre 4, dont la fonction de contextualisation ne permettait alors pas une analyse plus approfondie.

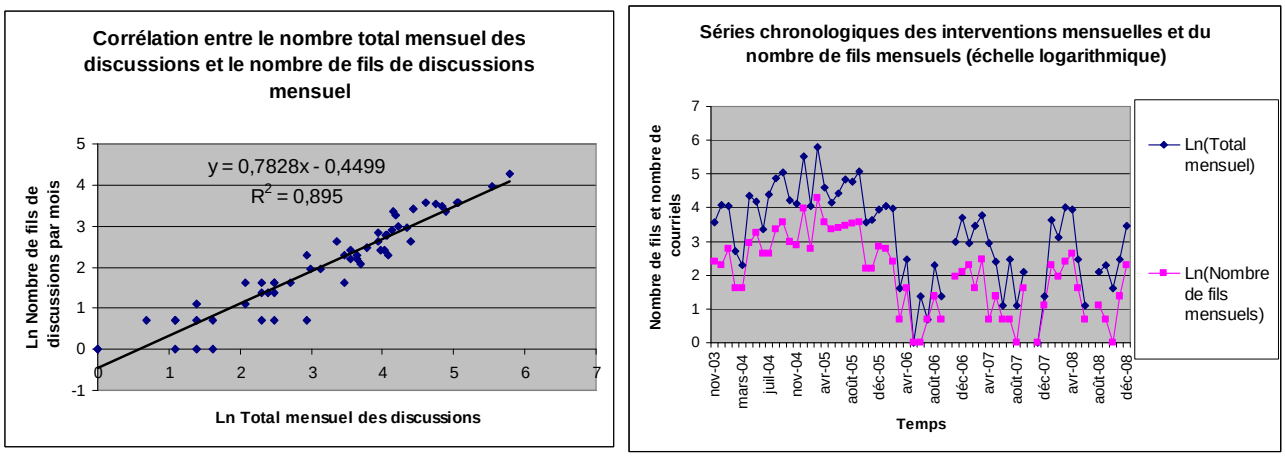


Figure 76 : corrélation entre le nombre total mensuel de discussions et le nombre de fils par mois (relation linéaire et séries chronologiques avec échelle logarithmique)

La série chronologique (graphique de droite ci-dessus) permet de souligner la relative constance d'une telle relation.

En revanche, si on examine l'évolution des fils en fonction de la démographie de la liste, *la relation n'est plus toujours linéaire*, qu'il s'agisse du nombre de fils mensuels ou de leur taille, ce qui montre l'importance de cette dernière variable pour aller au-delà de la relation linéaire précédente :

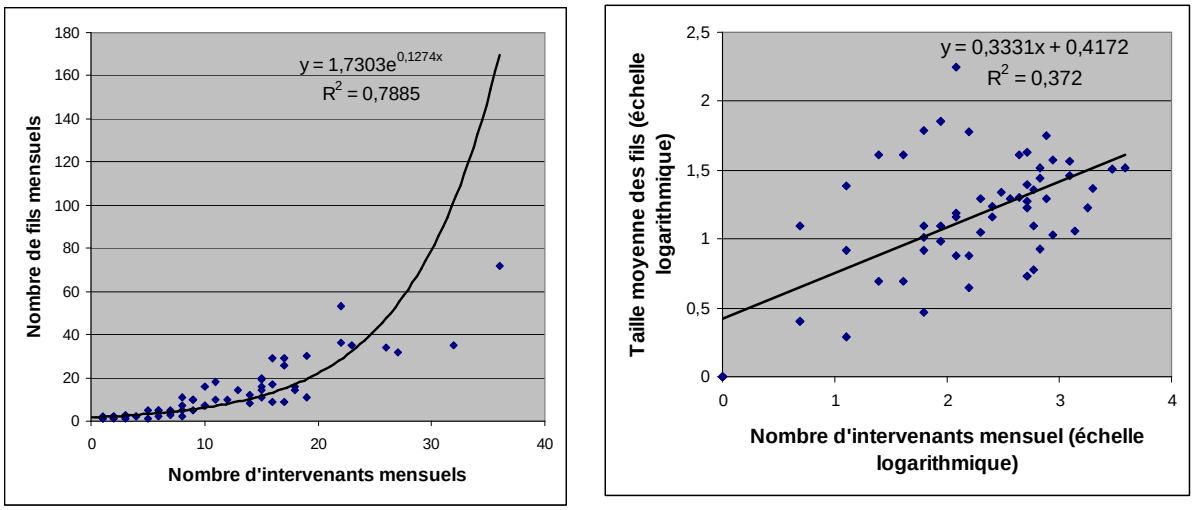
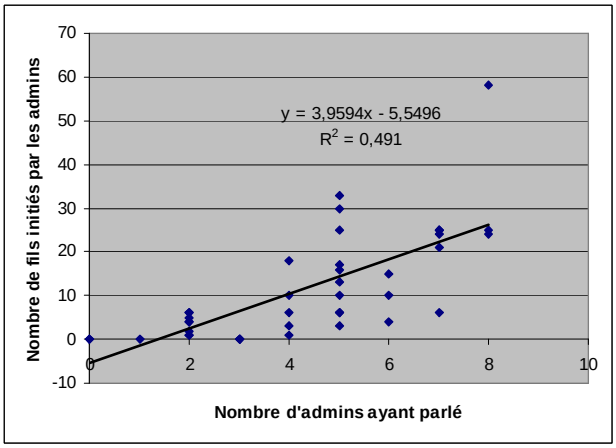


Figure 77 : mise en relation du nombre d'intervenants mensuel, du nombre de fils mensuels et de la taille moyenne des fils

Le graphique de gauche montre que plus le nombre de fil augmente, moins il y a d'intervenants concernés. *Mais postuler une telle relation non-linéaire ainsi est contestable*, à cause du *poids* de certains points éloignés. On se limitera à dire que la relation est croissante.

En revanche, le graphique de droite rapporte le nombre d'intervenants mensuel à la taille moyenne des fils (intensité des discussions à l'intérieur des fils). Il montre plus clairement une *intensification* du nombre d'intervenants *dans* les fils. L'augmentation du nombre de fils (corrélée avec celle des courriels) se traduit donc d'une part par l'intensification du nombre d'intervenants à *l'intérieur* des fils et d'autre part par l'augmentation globale du *nombre* d'intervenants. Il n'était nullement nécessaire qu'il en soit ainsi.

Cependant, si la relation précédente sur le graphique de gauche ci-dessus fonctionne de façon significative au niveau agrégé, elle est nettement moins marquée à l'intérieur de chaque sous-population, ce qui est difficile à interpréter :



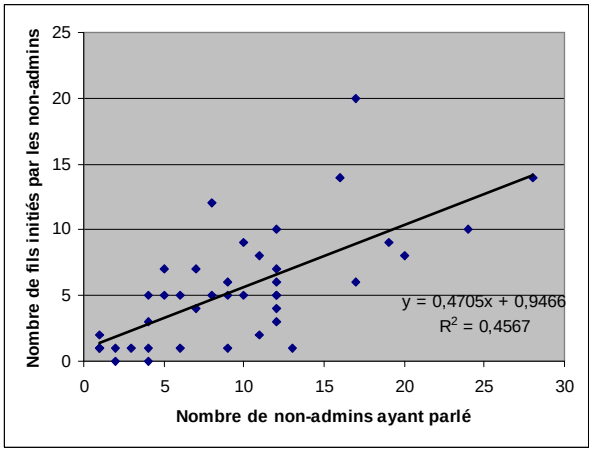


Figure 78 : mise en relation de la démographie des catégories d'intervenants et du nombre de fils de discussions qu'ils ont initié

De plus, si on s'en tient encore aux variables démographiques dont on cherche à approfondir l'effet, il est difficile *ci-dessous* de savoir laquelle des deux populations cause l'augmentation *globale* du nombre de fils, la différence de R2 et les formes de dispersions observées laissant entendre que la population des non-administrateurs a un rôle plus significatif. Cependant, la population des administrateurs a une taille plus limitée, ce qui fait que la qualité des résultats sera nettement inférieure. Nous ne nous prononcerons pas ici.

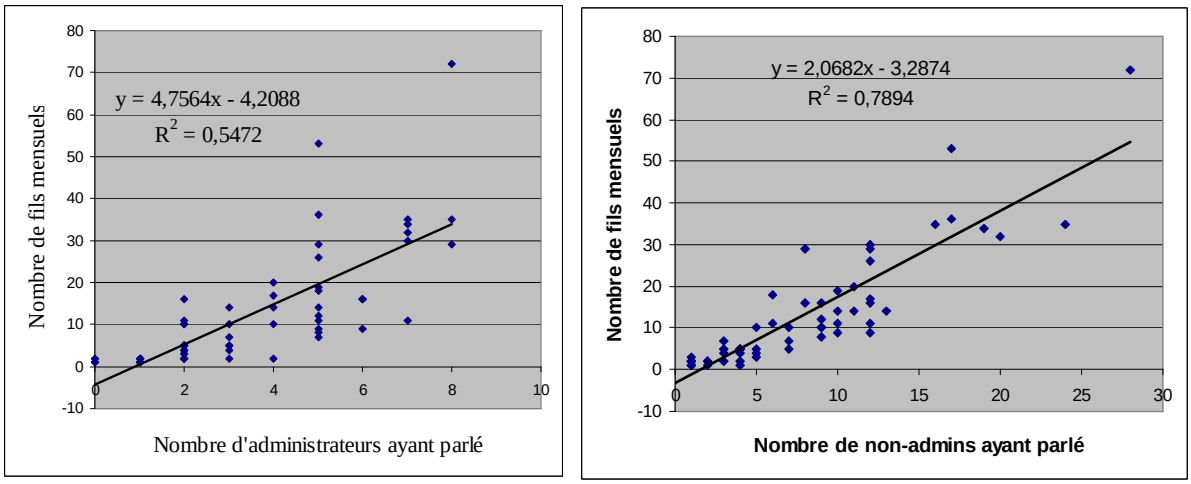


Figure 79 : le nombre de fils mensuels est corrélé au nombre d'administrateurs[810]*, mais aussi au nombre de non-administrateurs*

En revanche, on observe des régularités plus marquées entre le nombre de fils et le nombre d'interventions dans chaque sous-population, ce qui permet d'analyser la relation évoquée précédemment dans la population totale, entre nombre d'interventions et nombre de fils. La relation est nettement plus marquée dans la population des administrateurs que dans celle des non-administrateurs :

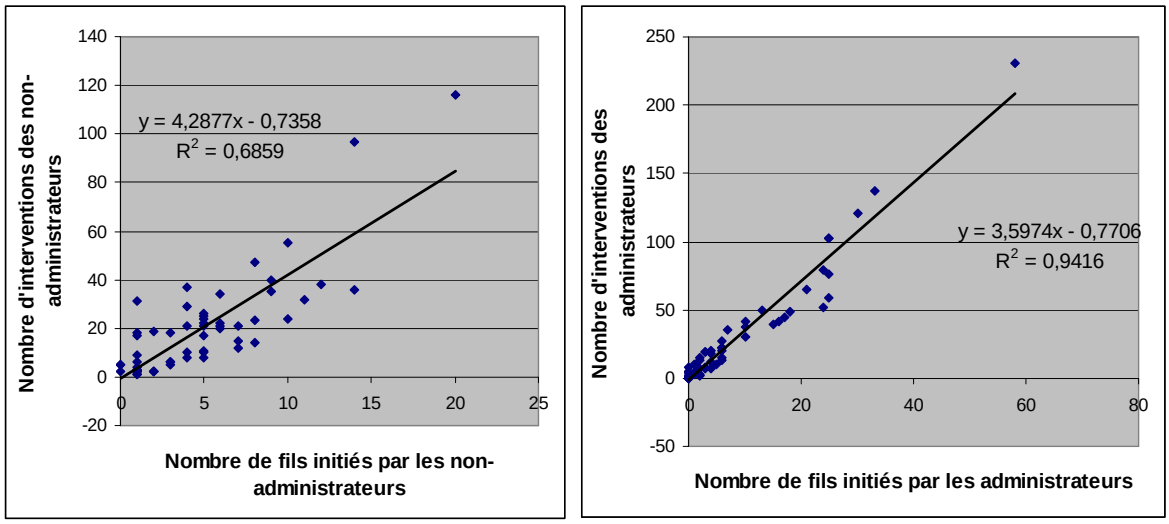
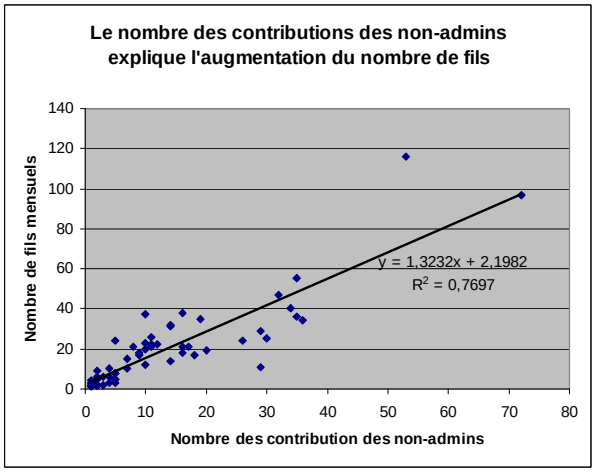


Figure 80 : mise en relation du nombre de fils initiés par chaque catégorie d'intervenants et du nombre de leurs interventions

[810] En ce qui concerne le graphique de gauche, une relation exponentielle non-linéaire donne un R2 de 0,68 mais la relation reste trop dépendante de valeurs extrêmes. Nous nous en sommes tenus à une relation linéaire.

On constate que la variable nombre de fils mensuels est liée de façon plus régulière au nombre d'interventions des administrateurs, comme le montrent les graphiques ci-dessus qui montrent un écart de R2 de plus de 0,25 ; *on peut interpréter cela comme une forme de régularité causée par les administrateurs dans la structuration de la gouvernance sur la liste.* On retrouve cette relation plus claire lorsqu'on lie le nombre d'interventions des deux populations avec cette fois-ci le nombre *total* de fils mensuels :



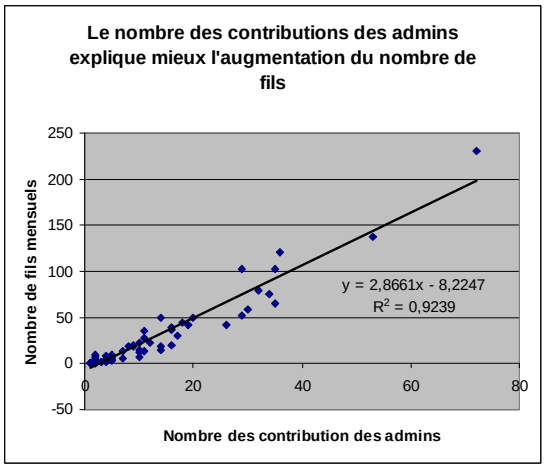


Figure 81 : comparaison des causes de l'augmentation du nombre de fils par mois : le nombre des contributions des administrateurs l'explique mieux

Ces derniers résultats sont conformes à ceux qui précédaient immédiatement, mais l'écart de R2 n'est sans doute pas suffisamment significatif pour qu'on puisse déduire des conclusions définitives.

Il est difficile de savoir quelle est la raison de *l'augmentation de la taille des fils*, qui est une caractéristique de l'intensité des discussions qui varie par cycles, d'un point de vue purement quantitatif. Nous avons travaillé longuement les données sans parvenir à quoi ce se soit de vraiment significativement concluant après avoir testé de nombreuses variables possibles. Outre le précédent graphique qui montre une petite corrélation entre le nombre d'intervenants mensuel et la taille moyenne des fils, nous mentionnons seulement ce qui s'en approche :

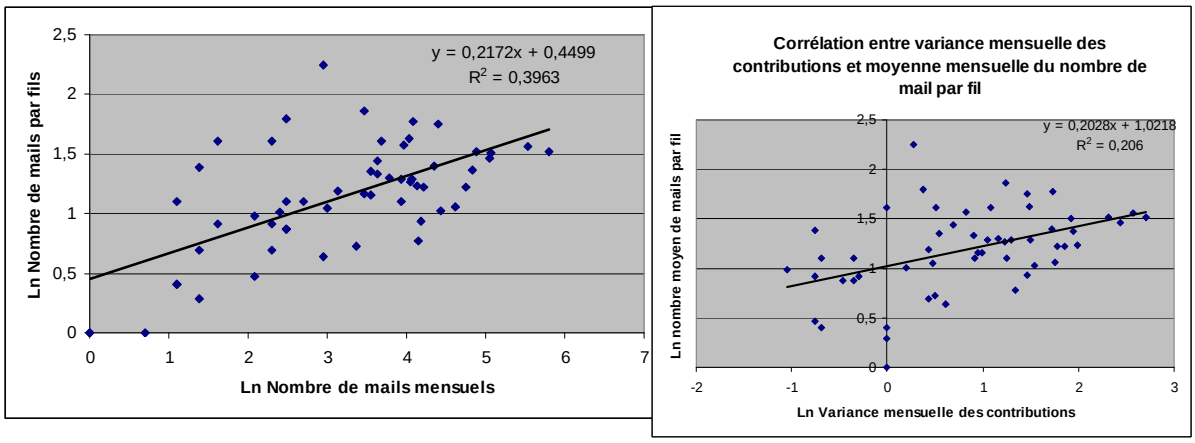


Figure 82 : mise en relation de la variance mensuelle des interventions (échelle logarithmique), du nombre de courriels mensuels et de la taille moyenne mensuelle des fils.

Le seul résultat correspond au premier graphique : le nombre de mails par mois fait augmenter le nombre de mails par fil, à l'échelle logarithmique, mais il est difficile de lui attribuer une signification. Est-ce que cela signifie que les nouveaux courriels ont davantage tendance à se loger à l'intérieur des fils existants (i.e. renforcement de l'intensité des discussions ?). C'est une interprétation qu'on peut retenir mais qui mériterait des approfondissements.

Le dépouillement de la liste de discussion de management de la documentation a permis d'approfondir la question par l'extraction de la variable binaire « initiateur du fil de discussion » (l'initiateur est-il administrateur ou non ?). Elle complète la différenciation précédente entre les deux populations d'intervenants. A partir de là, que peut-on dire des conséquences de *la prise d'initiative* des discussions par un tel ou un tel, spécifiquement par un administrateur ou un non-administrateur ? Nous disposons des chiffres requis différenciant les comportements des « admins » et « non-admins », dont voici une synthèse avant d'aller plus loin :

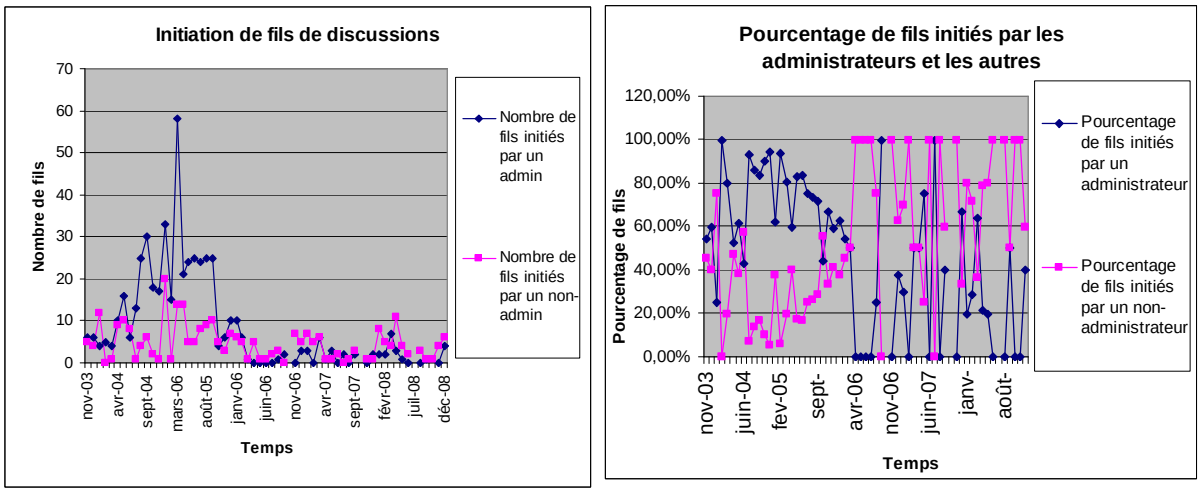


Figure 83 : série chronologique du nombre de fils initiés par les administrateurs et les non-administrateurs, et présentation en pourcentage

On compte en tout 492 fils initiés par les administrateurs, et 277 par les autres. Le ratio sur les cinq années est de 0,56 pour le nombre de fils, alors qu'il est de 0,66 pour le nombre d'interventions. Lorsqu'on regarde les courbes, on remarque une évolution début 2006, entre janvier et avril, les courbes de prise d'initiative des administrateurs passant sous celle des « non-administrateurs », à quelques exceptions près. La comparaison des chiffres sous forme de pourcentages permet de mettre en évidence un phénomène *d'inversion* structurelle à partir approximativement de avril 2006, inversion qui n'est remise en question qu'épisodiquement. Cette rupture renforce encore celles, semblables, observées précédemment début 2006 avec d'autres variables (démographiques, nombre d'interventions…). *La convergence des résultats sur ce point est forte*. Sa portée est cependant ici à *relativiser* étant donné les faibles valeurs sur certains mois. Il faut caractériser d'autres seuils : on en observe un fin 2005 pour les administrateurs, avec une baisse remarquable du nombre de fils initiés, qui passe de 20 à 30 à 0 à 10. Cela ne concerne qu'à la marge les non-administrateurs.

7.6.2. Mise en évidence d'un effet structurel d'homophilie

Raisonner en termes de pourcentage (comme précédemment), et non plus de chiffres bruts permet de mettre en évidence la présence d'homophilie: plus la proportion de fils initiés

par un administrateur augmente, plus la contribution des autres (non-administrateurs) diminue en proportion, et vice-versa. Il n'y a aucune nécessité qu'il en soit forcément ainsi :

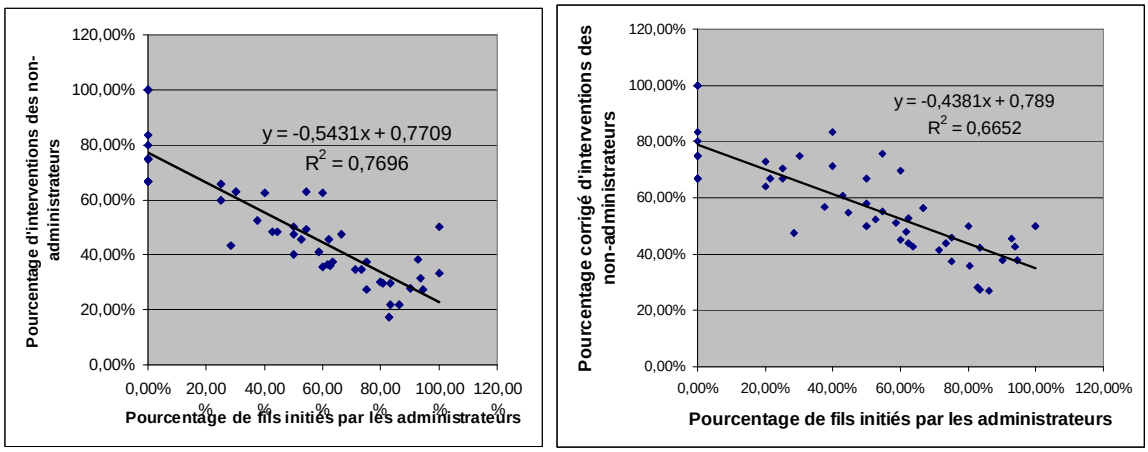


Figure 84 : régressions mettant en évidence l'éviction de la proportion d'interventions des non-administrateurs suite à l'augmentation de la proportion de prise d'initiative des administrateurs

Nous avons ici deux graphiques, l'un avec le pourcentage issu des données brutes du nombre d'interventions des non-administrateurs, l'autre avec des données d'où ont été exclues de la série chronologique en question les données « inverses », i.e. le nombre de contributions des administrateurs correspondant à l'initiation des fils.

En effet, utiliser les données brutes aurait *pu* diminuer excessivement le pourcentage d'interventions des non-administrateurs, étant donné que le phénomène pertinent à étudier correspondait au reste du fil, où se trouve la série chronologique pertinente. Un statisticien dira en effet que le phénomène mis en évidence pourrait se comprendre en partie, étant donné qu'un nombre significatif d'interventions des administrateurs se trouve *déjà* en début de fil. Pour donner un exemple parlant, s'il y a 100% de fils commencés par des administrateurs un mois donné, il serait probable de trouver en conséquence peu de contributions des non-administrateurs sur le reste du mois, même si ce n'est pas toujours forcément le cas. Exclure des données le nombre d'interventions des administrateurs au tout début du fil permet de considérer *uniquement* la série chronologique correspondant au reste du fil. On constate que prendre en compte ce biais *ne modifie qu'à la marge le phénomène observé*, comme le montre le graphique de droite qu'on pourra comparer au graphique de gauche. Le R2 n'a diminué que de 0,1 avec une régression linéaire standard.

Ce « test de robustesse » du modèle est peut-être cependant *trop sévère*, au vu de la taille moyenne des fils (3,4). Cela mériterait d'être discuté.

7.6.3. Interprétation et implications[811]

Une interprétation immédiate serait de penser que les administrateurs se parlent et se reconnaissent entre eux ; c'est un fait stylisé curieux concernant l'organisation massivement distribuée, en particulier sur la structure de régulation et de management qu'est la liste de discussions. L'explication habituellement retenue en analyse de réseaux est celle de *l'attachement préférentiel*.

Un tel phénomène montre les limites de « l'effet d'entraînement » en cascade de l'intervention des admins, puisqu'une augmentation de leur initiative limite les contributions en proportion des autres discutants sur la liste. Au-delà du *seuil de 60%* de fils initiés par les admins, les contributions des admins dominent toujours celles des non-admins.

Nous avons tenté de réfuter cette relation ou de trouver d'autres explications concurrentes, sans succès. *Ces tentatives ont été placées en annexes de ce chapitre.*

Le problème est d'expliquer, ou de mieux comprendre un tel phénomène d'attachement préférentiel, ce qui supposerait une analyse qualitative, mais surtout une analyse de réseaux, ce à quoi nous arrivons.

7.7. Une approche de la structuration de la liste de discussions à partir de l'analyse de réseau (Social Network analysis)[812]

Un des présupposés de cette sous-question de recherche abordant la notion de collaboration massivement distribuée est d'utiliser de façon complémentaire les analyses de type réseau et les statistiques fines. Nous en arrivons à la partie analyse de réseaux qui permet d'aller plus loin que l'analyse quantitative classique en matière de caractérisation des liens entre intervenants.

[811] Au vu de la sévérité du test et pour plus de commodité, nous continuerons donc par la suite avec les données brutes.
[812] Pour l'analyse de réseau et les concepts associés, nous renvoyons à DEGENNE et FORSE (1994).

7.7.1. Le processus de traitement des données

Nous présentons plus loin le résultat visuel de la structure du réseau sur trois ans et demi après passage des données dans trois programmes informatiques et nettoyage final des données. Nous expliquons cependant d'abord brièvement l'ingénierie nécessaire pour parvenir à ce résultat[813].

- La première étape a consisté à concevoir un dispositif d'extraction de données à partir des fichiers sous format « .txt » mensuels récapitulant l'activité de la liste de discussions dans les archives. N'étant pas expert en programmation et encore moins en Python, nous avons sollicité Jean-Philippe Cointet, doctorant au CREA, qui a passé quelques heures à réaliser le programme en Python et nous as expliqué comment le faire tourner[814]. Nous avions cependant conçu un « cahier des charges ». Nous avions demandé un programme qui permette d'une part l'extraction de tous les noms des auteurs de courriels et d'autre part la présentation de ces noms d'une façon qui permette la description de graphes *orientés*. Les fichiers « .txt » permettent ainsi après nettoyage la création d'un fichier « reseau_NomDuFichier » qui met à plat la liste des données (les intervenants sur la liste).

- Une fois l'extraction effectuée, une difficulté consistait à retravailler les données afin d'éliminer les pseudonymes multiples des contributeurs ; par exemple, le chef de projet « MatthewMullenweg » est aussi nommé « MattMullenweg », ou « Matt ». Nous avons simplifié ici en « MATT »[815]. *C'est là que la connaissance ethnographique de la communauté est utile* : il était nécessaire d'éviter les doublons dont le repérage est moins trivial; l'administrateur « Scott Merrill » a pour gravatar « SKIPPY » et nous avons conservé le gravatar. Ce nettoyage des données est *absolument nécessaire* car ne pas l'effectuer peut changer *de façon significative* la forme du réseau, surtout s'il n'apparaît pas très dense à première vue. Pour mener à bien une telle opération, nous avons effectué une première visualisation *(étape suivante)* qui permet de *voir ces doublons* ; ce n'est qu'après que nous avons nettoyé à nouveau à la marge les données dans « reseau_NomDuFichier ».

[813] Ce type de protocole peut être reproduit pour d'autres listes de discussions du même projet, après la thèse, et être utile à d'autres chercheurs.
[814] Le script du programme Python (appelé « parsing ») se trouve en annexes de la thèse.
[815] Ce travail des données se fait avec la commande « remplacer » pour des pertes de temps.

- Enfin, il était nécessaire de convertir le fichier « reseau_NomDuFichier » en format «.net » que le logiciel Pajek (et d'autres) permettent de lire, afin d'obtenir une visualisation et de mener à bien des analyses. C'est le rôle du logiciel « Txt2Pajek ».

7.7.2. Visualisations et remarques exploratoires

Voici donc une visualisation préalable du réseau des listes de discussions « docs » et « wp-docs ». Nous avons adopté la classique présentation « Kawada-Kawai » qui place les intervenants les plus connectés au centre :

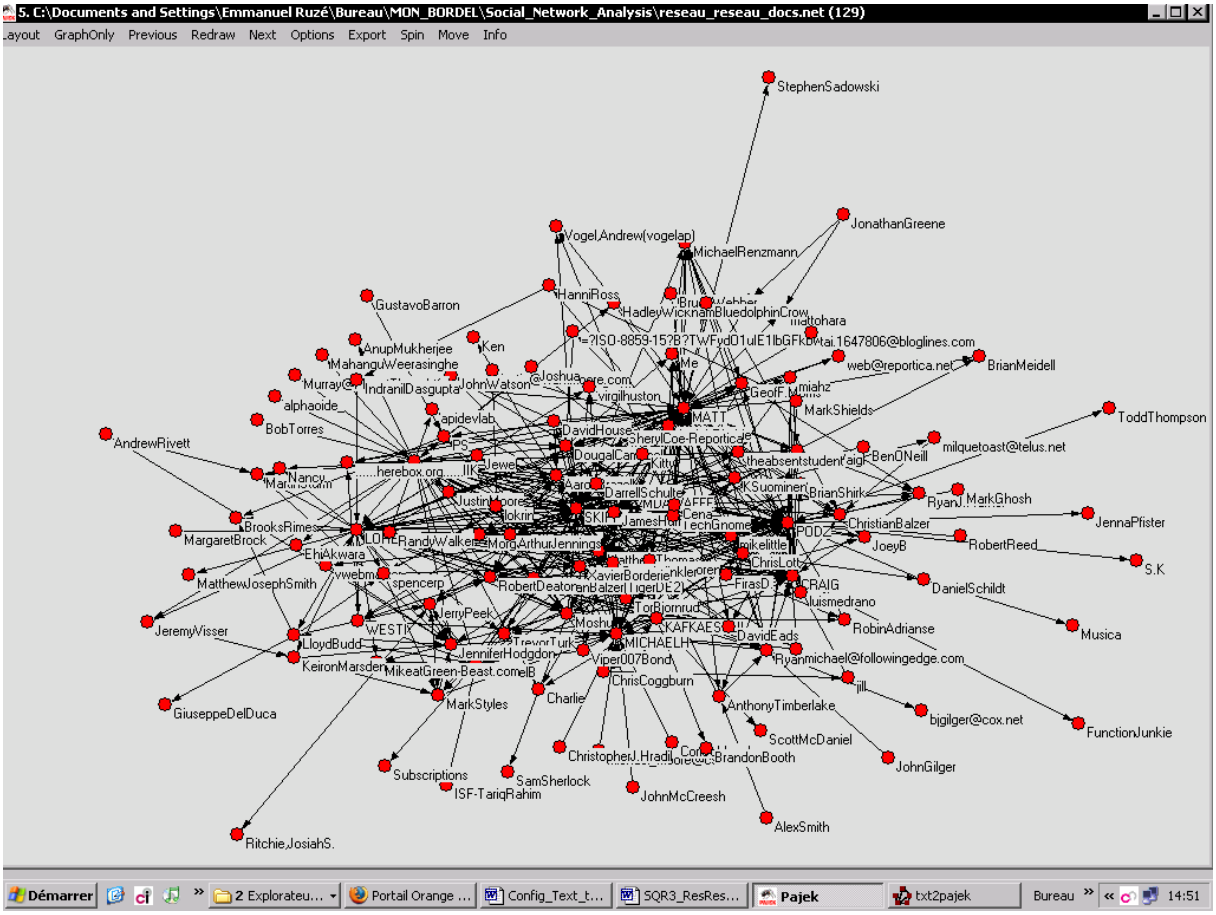


Figure 85 : visualisation du réseau des intervenants avec les plus prolifiques au centre

Chaque point rouge représente un intervenant (« nœud »), chaque flèche un courriel de réponse à un intervenant. Les administrateurs sont écrits en majuscule.

La visualisation est certes insuffisante, mais indispensable pour mémoriser approximativement la centralisation du réseau. Ensuite, un travail avec la souris consistant à déplacer les points rouges correspondant aux nœuds permet de se faire une idée des modes de communication dans la liste. Nous avons dispersé autant que faire se peut les 10 administrateurs autour du centre de l'écran, et réparti de façon homogène les autres intervenants. Cela donne une idée plus claire de ce qui se passe :

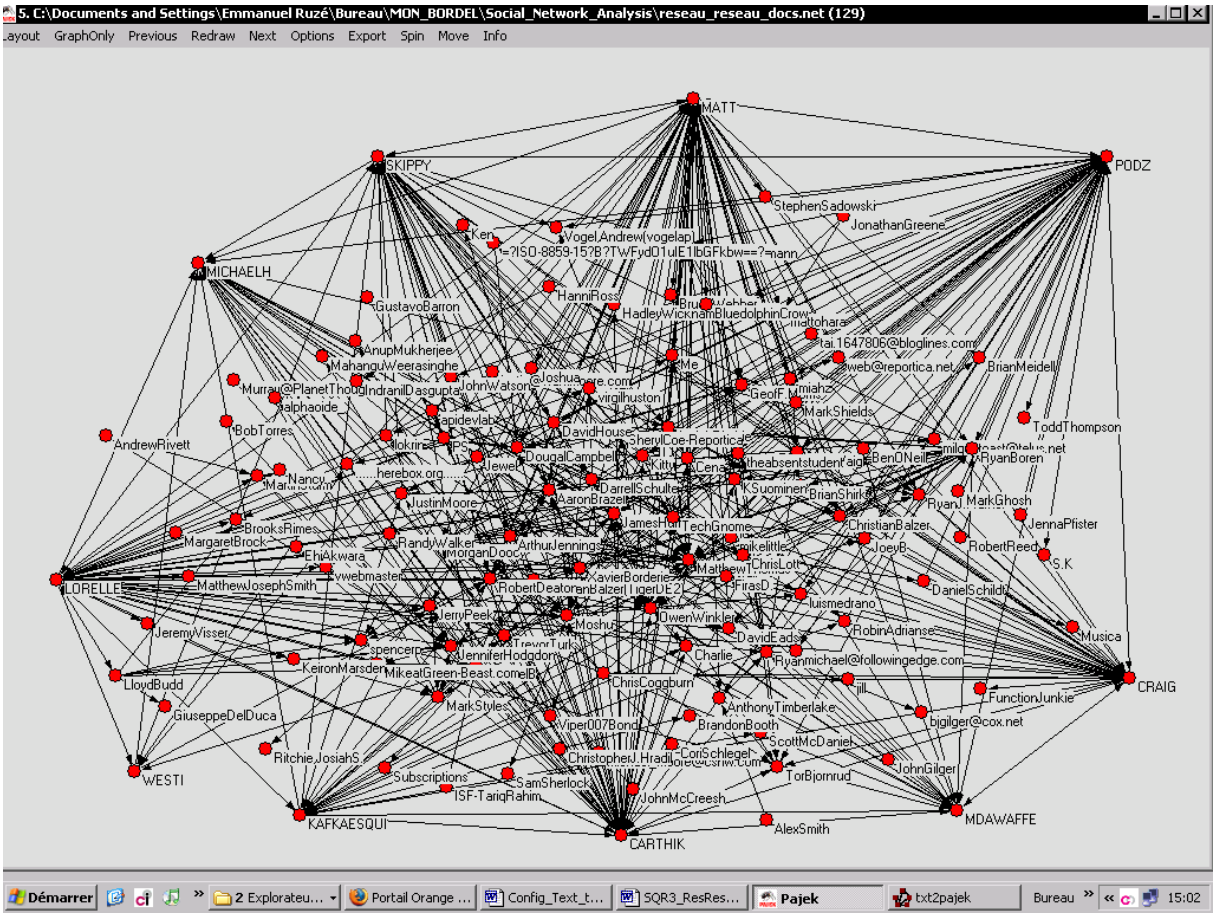


Figure 86 : visualisation du réseau des intervenants avec les administrateurs en périphérie

Nous présentons quelques résultats visuels exploratoires, qui sont pertinents car ils suggèrent des hypothèses à tester :

On observe à vue d'oeil que *la centralisation est en fait relative*, car un certain nombre d'individus sont fortement connectés. Elle ne concerne pas seulement les administrateurs mais aussi certains programmateurs : il reste en effet des zones noires denses au centre du graphe. Cependant, les administrateurs sont à première vue les individus avec le plus fort taux de connexions, une hypothèse qu'il faudra confirmer. Pour prendre une comparaison

géographique parlante, on constate ici que la centralisation dans la communauté ressemble plus à la dispersion des grandes villes allemandes qu'à la centralisation française sur Paris.

On observe, cela est visible sur la *première* visualisation, peu d'interactions entre les contributeurs mineurs, ceux qui interviennent une ou deux fois; le caractère "distribué" de l'organisation est donc en fait limité sur ce point, et nous reviendrons définitivement sur la connectivité par la suite. En revanche, on observe que les interactions *se diversifient assez vite* dès que le nombre d'interventions augmente. Il y aurait quatre catégories d'intervenants, ceux qui sont centraux, interviennent de façon conséquente; une deuxième couronne de non-administrateurs mais informaticiens chevronnés (qu'on ne distinguait pas dans les statistiques qui opposaient deux populations), ceux qui ont un petit noyau d'interlocuteurs diversifiés, mais qui interviennent moins; et enfin une constellation de petits intervenants isolés.

7.7.3. Analyse numérique du réseau

Nous avons calculé à partir des distributions des demi-degrés entrants et sortants la somme de ces deux variables pour l'ensemble de la population pour vérifier si les administrateurs étaient effectivement les intervenants les plus connectés. C'est effectivement le cas sur ce tableau partiel; ils sont en majuscules :

1 "Cena"	76
73 "MICHAELH"	78
51 "MDAWAFFE"	95
36 "OwenWinkler"	133
18 "SKIPPY"	140
56 "LORELLE"	211
4 "CRAIG"	306
13 "CARTHIK"	376
14 "PODZ"	414
2 "MATT"	431

Tableau 4 : somme des demi-degrés entrants et sortants chez les principaux intervenants

Voici des résultats descriptifs obtenus avec la commande « Network Analysis Toolkit » de Network Workbench Tool, selon nous l'outil open-source le plus commode et bien documenté que nous ayons trouvé (sauf pour la visualisation qui donne l'avantage à Pajek):

"Nodes: 129

Isolated nodes: 0

Average total degree: 27.410852713178212

Average in degree: 13.705426356589154

Average out degree: 13.705426356589152

This graph is weakly connected.

There are 1 weakly connected components. (0 isolates)

The largest connected component consists of 129 nodes.

This graph is not strongly connected.

There are 48 strongly connected components.

The largest strongly connected component consists of 82 nodes.

Did not calculate density due to the presence of self-loops and parallel edges.

Many algorithms will not function correctly with this graph."

Ce qu'il faut retenir, c'est que ce graphe est intégralement connecté[816], mais pas pour autant fortement connecté, ce qui confirme les premières remarques visuelles précédente sur les limites du caractère distribué de la collaboration. La taille maximale du sous-réseau fortement connecté consiste en 82 nœuds (individus). En revanche, la distance moyenne entre chaque paire de noeuds n'est pas très importante (le calcul a été effectué avec Pajek ici) :

Number of unreachable pairs: 5648

Average distance among reachable pairs: 2.55366

The most distant vertices: BenONeill (76) and Musica (84). Distance is 6.

Ensuite, l'indice de corrélation en un point («one-point correlations") indique la corrélation entre les arcs entrants (« demi-degré intérieur ») et les arcs sortants (« demi-degré extérieur ») des nœuds. Les nœuds avec nombre de degrés entrants sont dits « hubs » et les nœuds avec nombre de degrés sortants sont dits « autorités ». On peut calculer la corrélation entre les deux variables pour savoir si les « hubs » sont aussi des « autorités ». Si le coefficient de corrélation est significativement différent de 1, cela signifie qu'elle est significative. Ici, le coefficient est de 7.41, à deux décimales près, ce qui est fort significatif et appelle un approfondissement. On peut donc tenter de regarder avec profit les formes de régularités présentes ici :

[816] Tout courriel obtient une réponse, ce qui ne va pas de soi contre l'apparence : un participant peut écrire sur la liste et ne pas obtenir de réponse.

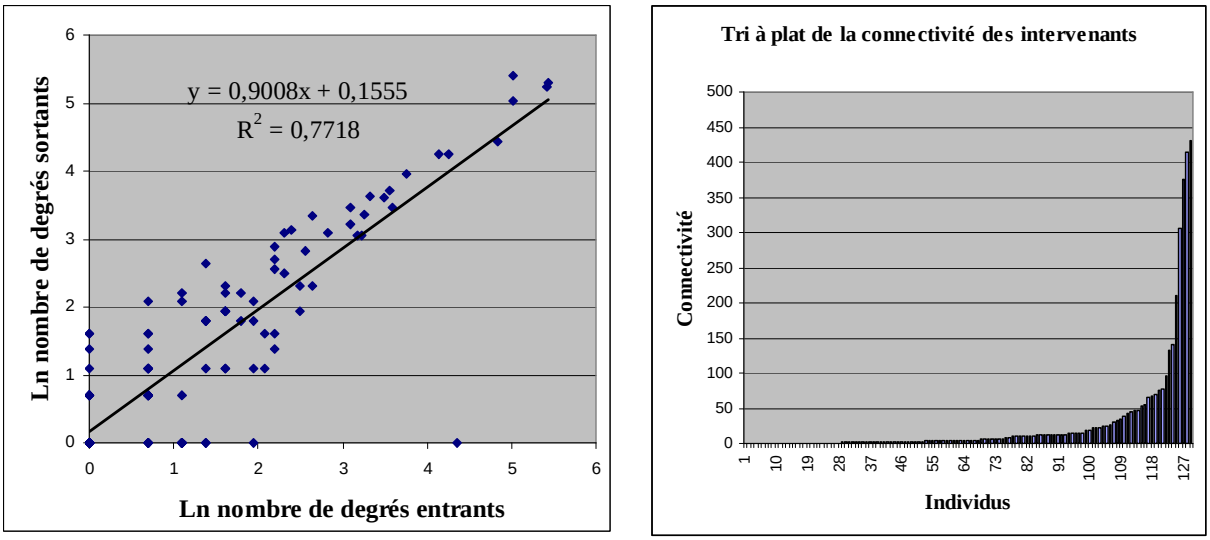


Figure 87 : mise en relation des degrés entrants et sortants et tri à plat des valeurs observées

A gauche, nous constatons que la distribution des demi-degrés intérieurs et extérieurs sont corrélés de façon significative, avec cependant une certaine hétérocédasticité pour les valeurs les plus faibles. A droite, nous avons présenté un tri à plat de la connectivité de chaque intervenant, qui forme une loi de puissance. *On appelle cela une distribution scale-free.* Les distributions des demi-degrés intérieurs et extérieurs suivent aussi des lois de puissance chacun de leur côté (nous ne présentons pas les graphiques). Précisons-le, les deux relations mentionnées sur les graphiques ci-dessus sont indépendantes : on peut avoir l'une sans avoir l'autre, et réciproquement. Avoir les deux est une caractéristique originale ici. L'apparition de telles caractéristiques avec une population réduite n'est pas triviale.

L'interprétation de la première relation signifie que les individus qui reçoivent des messages (« hubs ») sont aussi ceux qui en envoient (« authorities »), et ce, dans une mesure semblable. Cela signifie que l'autorité et l'expertise est distribuée de façon homogène. C'est là une composante à retenir du processus de collaboration, qui peut expliquer sa solidité : il ne dépend ni de certains « hubs », ni de certaines « authorities », si elles venaient à disparaître au point de créer des « trous » dans le réseau.

On peut pousser l'analyse un plus loin en regardant les caractéristiques de la collaboration en examinant justement la réciprocité entre intervenants. Quel est le nombre de nœuds concernés par des relations réciproques (au moins une), i.e. le nombre de dyades ? On peut le découvrir avec les commandes adéquates du logiciel. Dans le réseau, il y a 481 paires, 181 sont des dyades, ce qui fait un ratio de réciprocité de 0,376. On peut aussi examiner la

prévalence de la réciprocité par le ratio d'arcs impliqués dans ces relations (362), qui est de 0,546. La collaboration peut également être comprise en examinant le ratio de triades transitives[817] présentes dans le réseau : 0,269.

Les statistiques descriptives montrent qu'il existe 48 nœuds fortement connectés. Nous avons par ailleurs obtenu avec la même commande du logiciel la liste des demi-degrés intérieurs de tous les intervenants (placée en annexes, car longue). Nous avons choisi un seuil de 5 qui correspond approximativement aux dyades où il y a au moins un administrateur. Cela permet d'obtenir une liste de 77 ensembles de « relations » à 5 arcs ou plus[818], et donc 77 relations de collaborations. A contrario, on peut observer des relations plutôt faibles entre certains administrateurs, ce qui montre les *limites à la marge* de la collaboration dans cette sous-population. Nous dégageons ces relations dans la liste pour les plus gros intervenants où la remarque a une pertinence certaine :

> "There are 2 edges between node: CRAIG and node: LORELLE
> There are 3 edges between node: LORELLE and node: MICHAELH
> There are 3 edges between node: SKIPPY and node: LORELLE
> There are 3 edges between node: SKIPPY and node: CARTHIK"

7.8. Conclusions

La gouvernance d'un système d'information (le wiki nommé Codex) est un processus, *un processus de collaboration*, qui se fait de façon incontournable sur une liste de discussions. Son étude révèle que cet objet dynamique comporte des *propriétés* marquées, au-delà des considérations sur l'existence d'un noyau central de contributeurs (comme chez les informaticiens de l'open-source par exemple). Il s'agit à la fois d'éléments dynamiques mais aussi de formes de stabilité qui permettraient de comprendre la viabilité du processus :

- Une rupture historique qui le scinde en deux parties qui a été approfondie ici : la première correspond à une phase de régulation intense, la deuxième, plus calme, avec des retours cycliques. L'intensité des discussions forme en revanche des cycles indépendants.

---

[817] Le logiciel par exemple dit: "A triad is any triple of nodes (actors) {A, B, C}. In a directed network there are sixteen possible kinds of triads. A triad is said to be transitive if there is a link (tie) from A to B (AB) and a link from B to C (BC), then there is also a link from A to C (AC)."

[818] "There are 6 edges between node: MartinSturm and node: JenniferHodgdon.
There are 6 edges between node: JosiahRitchie and node: Cena".
Cela signifie que la relation est orientée de MartinSturm à JenniferHodgdon, i.e ; que le premier répond six fois au second dans un ou plusieurs fils.

- Des formes de dissymétries marquées liées à des formes de dispersions fortes, qui se traduisent par une instabilité des contributions du noyau central (qui n'est pas « éternel »).
- Des formes d'effets d'entraînement de type « boule de neige », autant en matière démographique qu'en ce qui concerne le nombre d'interventions. Elles ne sont pas incompatibles avec des régularités structurelles comme un taux stable d'encadrement entre administrateurs et non-administrateurs.
- Des formes de spécificités internes à ces deux populations d'intervenants. Entre autres, la structuration des fils de discussions obéit à une logique d'homophilie.
- Nous avons cependant une forme de réseau plutôt lâche avec plusieurs noyaux, la moitié des intervenants étant fortement connectés, avec à côté une constellation de plus petits intervenants. La distribution des interactions entre intervenants est de type « scale-free », avec des formes marquées de réciprocité et de reconnaissance réciproque de l'expertise (les hubs sont aussi des autorités).

Nous avons par ailleurs amplement montré que les administrateurs avaient des comportements spécifiques, au-delà de l'étiquette en question, qui correspond donc à des pratiques collectives *réelles*.

Ce travail permettrait de *nuancer* le modèle d'interaction que propose par exemple Michel Gensollen, pourtant a priori particulièrement pertinent pour les wikis :

> "Les communautés en ligne se caractérisent essentiellement par un lieu virtuel, un corpus structuré, un tableau noir sur lequel chacun peut écrire et que chacun peut aller consulter. Les contributeurs ne cherchent pas à communiquer mais plutôt à modifier de façon pertinente une somme cohérente de données déjà existantes. Les communautés en ligne forment, ainsi, plutôt des cultures que des réseaux sociaux. De même qu'un auteur ne s'adresse à personne en particulier mais contribue à un corpus déjà constitué, de même les participants à une communauté en ligne œuvrent solitairement, souvent dans l'indifférence de qui utilisera leur contribution et sans motivations altruistes ou mercantiles." (GENSOLLEN, 2005)

Il suffit de connaître la nature des sources disponibles, mentionnées auparavant, pour se douter que dans une large mesure, un tel modèle n'est *pas le seul* applicable au Codex : il existerait un réseau de relations sur un autre support dont le rôle est dans une certaine mesure essentiel, et ce au minimum lors des phases initiales du projet. Il se caractérise par des dynamiques propres. Rappelons qu'on classe d'ordinaire les wikis parmi les *outils relationnels*, on entend par là que l'outil permet une mise en relation des individus. Or, cette classification ne va pas de soi, et une première question qu'on peut se poser est : est-ce que cet outil de type « tableau noir » permet une mise en relation des individus ? En fait, la

socialisation se fait sur la liste de discussion, ainsi que sur les onglets de discussions, pas sur le « corpus », ce qui nuance ce modèle.

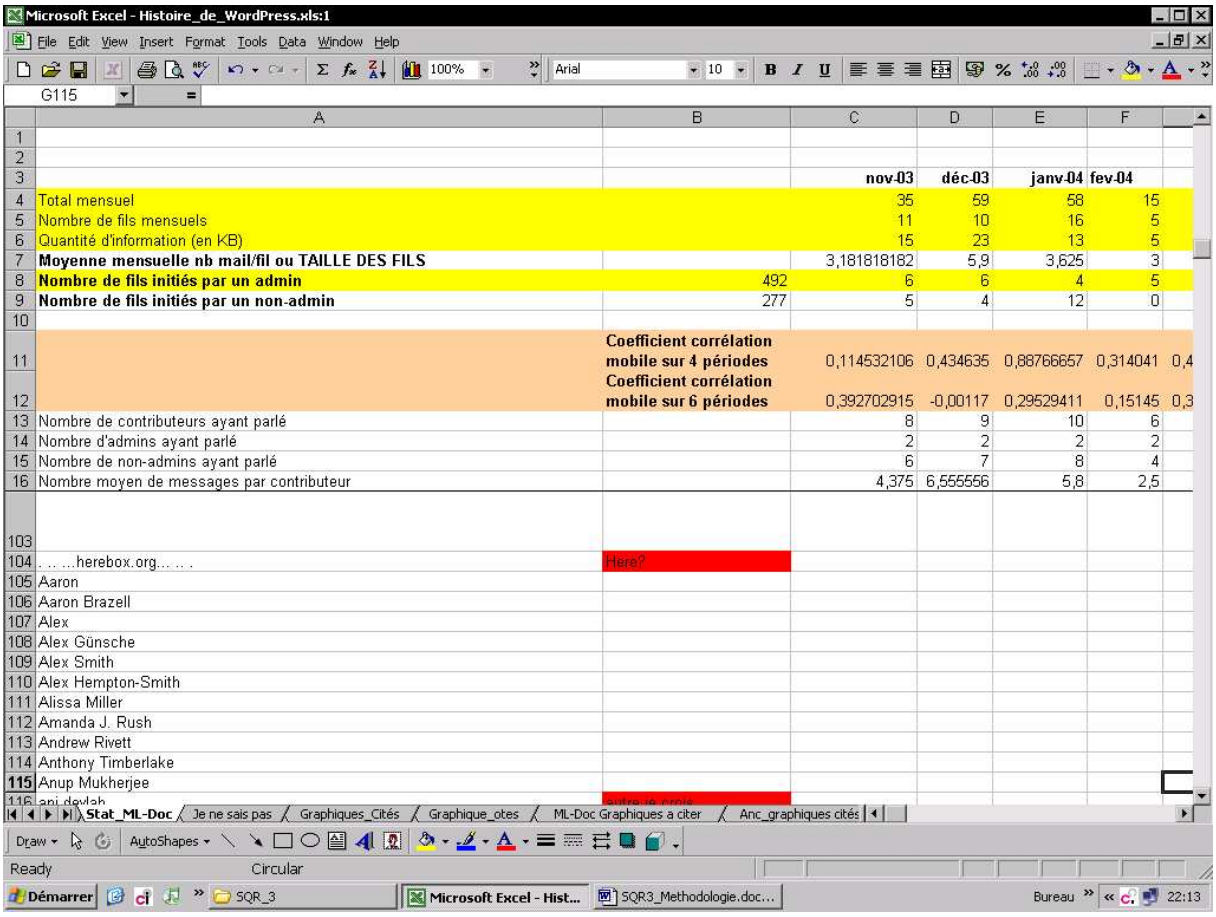
Bien entendu, nous ne prétendons pas épuiser définitivement ici la question, mais nous pensons l'avoir honnêtement défrichée, souvent sans trouver de relations significatives, ce que nous n'avons pas mentionné ici.

D'autres questions n'ont pu être abordées, mais mériteraient de l'être. Ainsi, observe-t-on des différences entre la structuration de cette liste de discussions et celles d'autres listes dans la communauté même, ou dans d'autres communautés qui font usage de documentations ou de formes de support utilisateurs ? Il en est de même pour la forme des réseaux. La régulation par la liste de discussion se différencie-t-elle d'autres formes de régulations, par exemple les onglets de discussion des articles ? Une autre voie de recherche possible serait de travailler une modélisation de la co-évolution individus/thèmes dans la tradition du CREA. Cela pourrait se faire à condition de dépouiller à nouveau les données pour comprendre si un thème est attaché à un individu, comment une telle spécialisation émerge, comment une problématique de gestion se diffuse, surgit à nouveau, etc. Un tel travail de dépouillement supplémentaire avec éventuellement un codage associé (comme le conseillait Jean-Michel Dalle) n'a pas été effectué dans le cadre de cette thèse par manque de temps, mais peut être envisageable par la suite[819].

[819] Ce chapitre de thèse pourrait faire l'objet d'une publication indépendante, en collaboration avec Carla Taramasco et le CREA. Dans ce cadre, des approfondissements en matière de modélisation sont probablement à prévoir. Nous pensons par exemple au calibrage plus précis des lois de puissances, ou à l'approfondissement des caractéristiques du réseau.

7.9. Annexe du chapitre 7

7.9.1. Structuration de la base de données



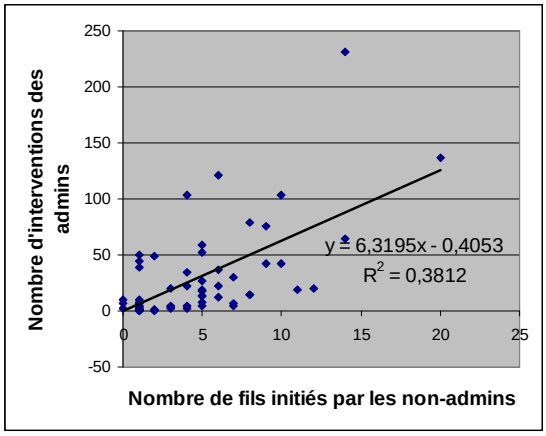
	A	B	C	D	E	F	
3			nov-03	déc-03	janv-04	fev-04	
4	Total mensuel		35	59	58	15	
5	Nombre de fils mensuels		11	10	16	5	
6	Quantité d'information (en KB)		15	23	13	5	
7	**Moyenne mensuelle nb mail/fil ou TAILLE DES FILS**		3,181818182	5,9	3,625	3	
8	**Nombre de fils initiés par un admin**	492	6	6	4	5	
9	**Nombre de fils initiés par un non-admin**	277	5	4	12	0	
10							
11		**Coefficient corrélation mobile sur 4 périodes**	0,114532106	0,434635	0,88766657	0,314041	0,4
12		**Coefficient corrélation mobile sur 6 périodes**	0,392702915	-0,00117	0,29529411	0,15145	0,3
13	Nombre de contributeurs ayant parlé		8	9	10	6	
14	Nombre d'admins ayant parlé		2	2	2	2	
15	Nombre de non-admins ayant parlé		6	7	8	4	
16	Nombre moyen de messages par contributeur		4,375	6,555556	5,8	2,5	
103							
104	. . . . .herebox.org... . . .	Here?					
105	Aaron						
106	Aaron Brazell						
107	Alex						
108	Alex Günsche						
109	Alex Smith						
110	Alex Hempton-Smith						
111	Alissa Miller						
112	Amanda J. Rush						
113	Andrew Rivett						
114	Anthony Timberlake						
115	Anup Mukherjee						

Figure 88 : aperçu d'une partie de la base de données

La structure de la base de données, pour le dire brièvement, est la suivante : quelques variables ont été dépouillées à la main, comme le nombre de courriels mensuels, le nombre de fils, la quantité d'information, la catégorisation des intervenants qui initient les fils. Le plus long est surtout l'identification un à un des intervenants, dont on a dénombré chaque mois les interventions au fur et à mesure. On voit dans la capture d'écran ceux commençant par la lettre « a ». Un tel travail aurait été encore plus long sans l'aide des outils de classement des archives dont nous avons parlé en méthodologie. C'est à partir d'un tel dépouillement fin qu'on peut calculer un grand nombre de variables diverses, visibles par la suite. Les outils utilisés furent Excel, XcelStat, Systat 11, pour l'analyse de réseaux, Txt2Pajek, Pajek et Network Workbench Tool.

7.9.2. Autres explications et tentatives de réfutation du phénomène d'homophilie mis en évidence dans la partie 7.6.

La relation précédente est *spécifique* à la variable nombre des interventions des non-administrateurs (exprimée en pourcentage), pas au *nombre d'individus* non-administrateurs; le nombre *d'intervenants* de chaque catégorie est au contraire corrélé[820] et le nombre de contributions des uns et des autres aussi[821]. La relation n'apparaît pas non plus si on examine uniquement les données brutes, sans passer par le pourcentage[822] :



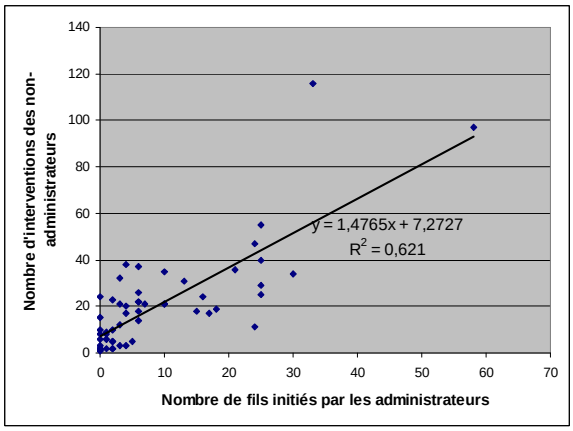
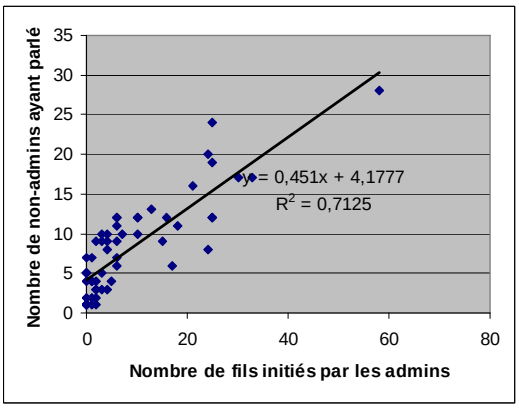


Figure 89 : le nombre de fils initiés par les administrateurs est corrélé imparfaitement avec le nombre d'interventions des non-administrateurs, et vice-versa[823]

Les premiers croisements de données sont donc *trompeurs*, comme ces derniers graphiques, avec uniquement les statistiques brutes, qui croisent nombre de fils initiés par les admins et les contributions des *non-admins*. D'autres croisements de données montrent la spécificité du phénomène, notamment lorsqu'on croise les variables d'initiation de fils et des variables démographiques, exprimées en données brutes ou en pourcentage :

[820] Nous renvoyons à l'étude de la démographique pour la visualisation en question.
[821] Nous renvoyons à l'étude du nombre d'interventions pour la visualisation en question.
[822] Ici, bien entendu, les régressions poseraient des problèmes; le but n'était pas de les valider, mais de constater qu'on ne trouve à coup sûr rien de semblable à la relation décroissante qui précède, ce pour quoi la visualisation suffit.
[823] En revanche, chaque mois, le nombre de fils initiés par l'une et l'autre catégorie n'est pas significativement corrélé (nous ne présentons pas le graphique).



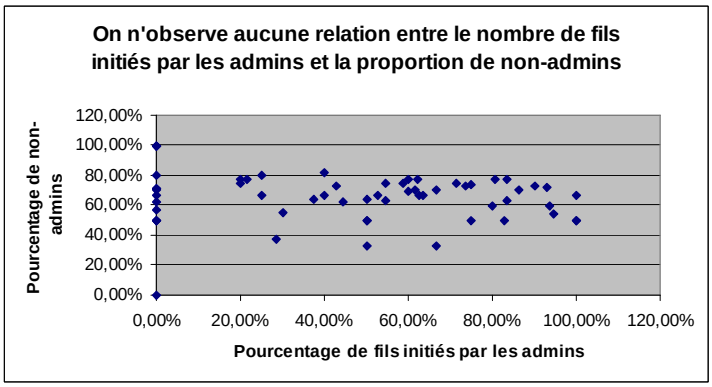
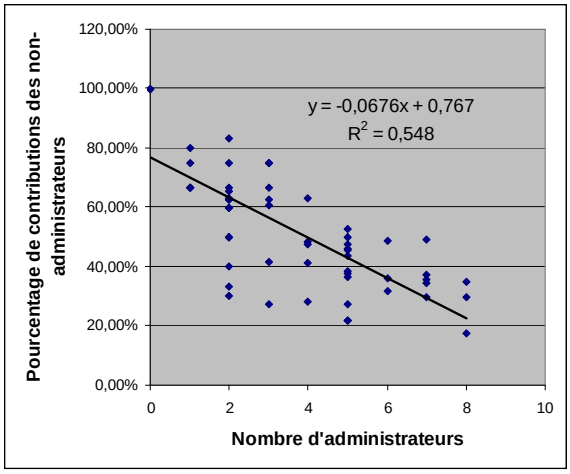


Figure 90 : le nombre de fils initiés par les administrateurs est corrélé de façon positive avec le nombre de non-administrateurs ayant parlé. Il n'y a aucune relation entre le pourcentage de fils initiés par les administrateurs et le pourcentage de non-administrateurs.

Nous n'avons pas ajouté de régression pour le deuxième graphique, le R2 étant insignifiant.

Une autre explication possible serait d'ordre démographique, mais elle comporte certaines limites. Le nombre d'administrateurs fait diminuer en pourcentage de contributions des non-administrateurs, comme le montre le graphique de gauche ci-dessous ; il pourrait être à l'origine de l'effet d'éviction :



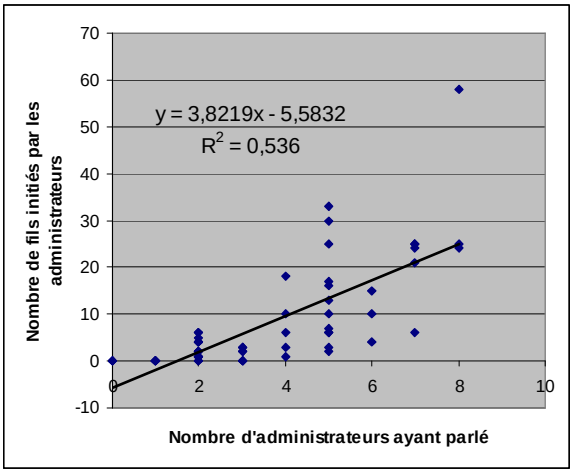
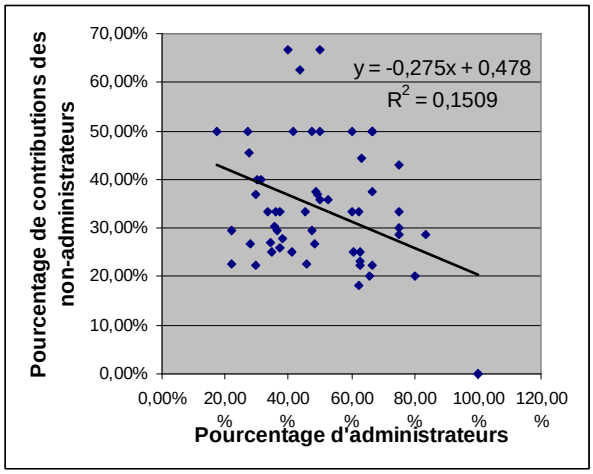


Figure 91 : corrélation négative entre le nombre d'administrateurs et le pourcentage de contributions des non-administrateurs. Corrélation positive entre le nombre d'administrateurs ayant parlé et le nombre de fils initiés par eux.

Cependant, l'explication n'est pas si originale, la variable démographique « nombre d'administrateurs » et « nombre de fils initiés par les administrateurs » étant, dans une

certaine mesure corrélés (graphique de droite ci-dessus). Il est contestable de dire qu'il s'agit d'une explication « alternative ». De plus, si on raisonne en pourcentage pour le nombre d'administrateurs (comme pour l'effet d'éviction), la relation disparaît, comme le montre le graphique *à gauche* ci-dessous :



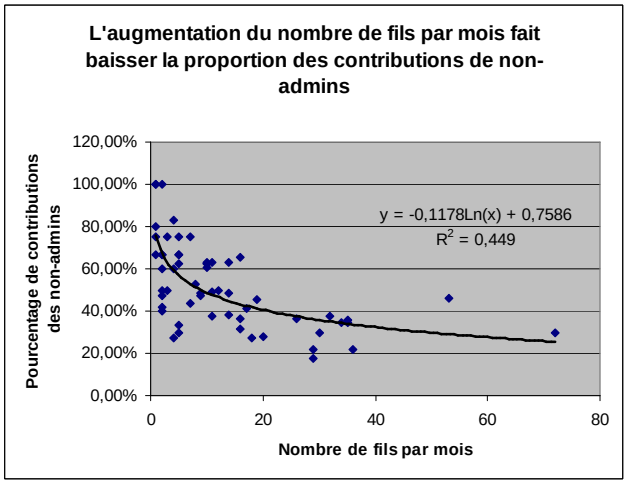


Figure 92 : pas de corrélation entre le pourcentage d'administrateurs et le pourcentage de contributions des non-administrateurs. Relation décroissante entre le nombre de fils par mois et le pourcentage de contributions des non-administrateurs

Enfin, une autre explication possible observée serait que le nombre de fils mensuels ferait diminuer le pourcentage de contributions des non-administrateurs (graphique ci-dessus de droite). Cependant, la relation est moins forte, postuler une relation non-linéaire est sans doute contestable, et *aucune théorie* ne vient permettre de comprendre une telle relation, qui pourrait n'exister que parce qu'il existe un phénomène de dépendance par rapport au chemin, la période d'augmentation du nombre de fils mensuels correspondant *par hasard* à celle où le pourcentage de contributions des non-administrateurs diminue.

En effet, on peut observer dans une certaine mesure un phénomène de « dépendance par rapport au chemin » qui pourrait avoir un rôle pour comprendre le phénomène d'éviction observé. Dans la régression précédente, chaque point représentait des chiffres mensuels tirés de la série chronologique, mais la représentation graphique fait que la chronologie n'apparaît plus. Nous présentons ci-dessous le phénomène *inverse*, i.e. la corrélation entre le pourcentage de fils initiés par les non-administrateurs et le pourcentage de leurs interventions :

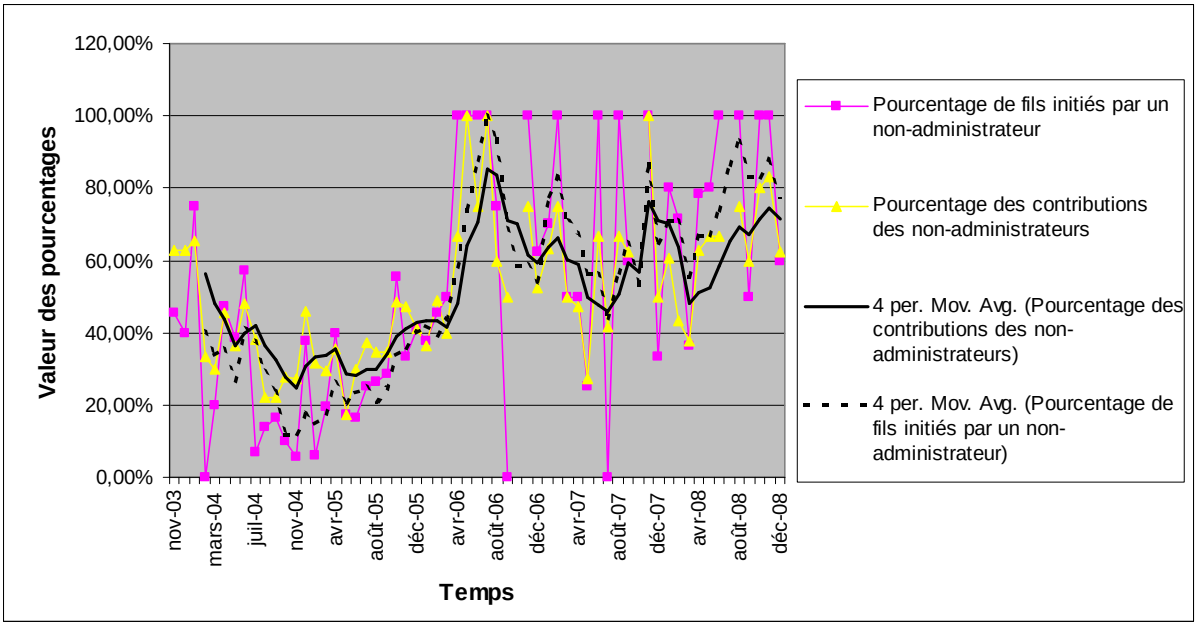


Figure 93 : série chronologique de la valeur des pourcentages des interventions des non-administrateurs et du nombre de fils initiés par eux

Nous avons calculé deux moyennes mobiles pour les deux séries, où on voit par ailleurs bien la corrélation entre les deux séries. On peut mieux comprendre ce phénomène en montrant qu'il se caractérise par une trajectoire de « dépendance par rapport au chemin » (« path-dependency »), sur une première période, et par une inversion de tendance sur une deuxième. Sur la deuxième, les plus faibles valeurs du nombre des interventions peuvent expliquer davantage les oscillations observées.

Chapitre 8 : l'appropriation du Codex dans le reste de la communauté

"A l'âge de l'information, les possibilité d'apprendre encore à lire sont en voie d'extinction" (Martin Heidegger à Hannah Arendt, 22 juin 1972)[824].

8.1. Introduction

Nous nous intéressons ici aux formes efficaces d'appropriation de l'information, ou, pour dire les choses autrement, aux conditions de l'efficacité du SI pour ses utilisateurs. Il s'agit ici en particulier des programmateurs et de l'interface avec le forum. On peut dire qu'il s'agit d'un élément de sa viabilité, ou de la « physiologie » qu'un tel système d'information suppose, c'est-à-dire les relations avec d'autres organes de la communauté WordPress.

Le résultat auquel nous avons abouti par ce travail essentiellement qualitatif, exploratoire et inductif est une compréhension fine du fait que la question de l'appropriation efficace du SI est nuancée en fonction du contexte, en particulier du type de population de contributeurs concerné : en particulier, et pour résumer ici, nous montrons qu'elle est plus difficile et suppose plus de contraintes pour les utilisateurs novices que pour les utilisateurs experts que sont les informaticiens.

Sur le plan méthodologique, nous avons eu l'intuition qu'en cherchant dans les archives des listes de discussions servant à la régulation des autres partie de la communauté, par exemple celles de l'activité d'informatique (programmation, test..) et celles du forum d'aide aux utilisateurs, nous obtiendrions d'une façon inductive des indices sur l'appropriation, la qualité, les usages, etc. de la documentation. C'est une problématique cruciale en système d'information mais aussi en économie (« economics of knowledge management »), qui problématise la question de l'évaluation de l'efficience du savoir traité et produit. Sur les aspects théoriques sous-jacents à cette sous-question de recherche, nous renvoyons à la fin du chapitre 1.

Comme nous l'avions expliqué dans notre chapitre 3 de méthodologie, nous commençons par une explicitation de la méthodologie et du protocole adopté. Nous présenterons ensuite séparément les résultats correspondant au forum d'aide aux utilisateurs puis aux informaticiens de la communauté.

[824] ARENDT Hannah (1999), *La Vie de l'esprit*, PUF, Paris, 571 p. La citation provient du numéro 2 du *Libellio d'AEGIS* de 2006.

8.2. Conception de la sous-question de recherche

8.2.1. L'origine de la sous-question de recherche, une approche historique « opportuniste »

L'origine de cette sous-question de recherche qui n'était pas prévue au commencement de ce travail provient d'une *intuition*, traiter les traces d'usages du wiki dans les autres archives des listes de discussion de la communauté. Cette intuition a émergé parallèlement aussi dans la thèse de Derek Hansen. Notre idée de départ était d'obtenir des éléments révélateurs de l'histoire du wiki, des informations supplémentaires, pour effectuer des croisements des sources. Puis, cette source s'est révélée pertinente pour obtenir des éléments concernant la perception de la qualité, l'efficacité réelle du résultat de la collaboration massivement distribuée sur le wiki, l'appropriation et des usages du « repository of knowledge » associés à cette efficacité. De même que nous avons profité de l'opportunité d'une liste de discussion spécialement dédiée à la documentation, de même nous avons profité de *l'opportunité* de sources privilégiées en travaillant les listes de discussions des autres parties de la documentation. Les premières réactions à leur lecture étaient qu'elles étaient difficilement utilisables, nous pensions tout d'abord nous focaliser sur les seuls passages qu'on pourrait qualifier d'« incidents critiques » évoquant le wiki ; cela a vite semblé déraisonnable car la prise en compte de ces seuls éléments pouvaient faire manquer des informations pertinentes (notamment la régularité des citations et des usages) : leur travail ont montré leur importance quantitative notable, la régularité de l'apparition des mots-clefs, et l'accumulation potentielles de données. Il s'agit donc d'un acte opportuniste, mais aussi d'un réflexe d'historien qui a l'occasion inattendue de rencontrer des « archives » d'une nature pertinente. La méthodologie d'extraction des données a été conçue au fur et à mesure de la lecture exploratoire des données et de la découverte des contraintes et opportunités qu'elles nous imposaient.

8.2.2. Choix des parties de la communauté étudiée et sélection des sources associées.

Le choix des sources adopté se justifie ainsi : elles se rapportent en effet aux deux autres organes d'importance critique dans la communauté (la programmation, l'interface avec les utilisateurs), et nous avons donc choisi de privilégier les sources correspondantes.

Certaines listes ont été cependant *exclues*. Nous avons effectué un travail d'extraction de données à partir des listes de discussions « hackers » (suivie par « wp-hackers »[825]), « forum », « trac », « wp-testers ». Ce sont les principales listes permettant d'obtenir des mentions des mots « wikis », « codex ». Nous avons en effet effectué des « carottages exploratoires » dans les listes de discussions, sur les fichiers au format « .txt » récapitulant l'activité mensuelle, afin de savoir dans un premier temps quelles étaient celles qui étaient pertinentes : certaines, comme celles des sous-projets du logiciel « Spam-stopper », « Wp-install », « CVS et SVN » ne comportent après examen en effet *que peu ou pas de mentions* (moins d'une dizaine au mieux)[826]. Une autre, comme celle des testeurs, retenue, nous a donné des *résultats inattendus* (voir chiffres ci-dessous), et ils complètent ceux de la liste de discussions des hackers, pour montrer que l'usage du wiki parmi les informaticiens est régulier.

Nous rappelons ensuite un arbitrage significatif : nous avons renoncé à analyser systématiquement l'intérieur du forum (chapitre 3)**.** Il était cependant utile de travailler *à la marge* les données « brutes » du forum, en assumant les limites précédentes. Aussi avons-nous examiné les fils de discussions associés aux mot-clefs « Codex », « manual », « documentation », « docs », « wp-docs » dans le forum, classés ainsi sous forme de « tags ». Cependant, on compte nettement plus de fils de discussions qui mentionnent le Codex que ce que suggère la liste de fils de discussions derrière le « tag ». Ce type de croisement des sources (ou triangulation) ne saurait donc avoir un caractère systématique, mais au minimum une *valeur exploratoire et infirmatoire*, ce qui n'est pas négligeable, car de bonnes idées peuvent être ainsi découvertes ou précisées. Nous pouvons apercevoir aussi quelques régularités (les pannes, les demandes de documentations sous un autre format que celui du wiki). La politique de « marquage » nous est inconnue, ce qui occasionne peut-être des biais en matière de conclusions qu'on pourrait inférer.

[825] Il n'y a pas de différence de nature entre les deux listes de discussions des hackers. Seule la seconde est référencée, et la découverte des anciennes archives a été faite par déduction.
[826] Nous n'avons pas eu le temps de trouver le pourquoi de ces absences de citations de façon satisfaisante.

8.2.3. Extraction des données et saturation des références possibles au wiki

Une telle procédure paraît automatisée. Mais c'est déjà une *démarche d'interprétation* cachée, qui conditionne évidemment les interprétations ultérieures ; nous avons désiré ici l'*expliciter clairement*. C'est un point sur lequel LEMERCIER et ZALC (2008) insistent à juste titre.

Le *protocole* adopté, pour en donner une vision générale, a été de filtrer les courriels pertinents en les sélectionnant à partir de plusieurs mots-clefs, ce qui a permis une « saturation » du corpus de courriers électroniques pertinent. En effet, effectuer une extraction à partir d'un nouveau mot-clef (« doc »), pour vérifier si rien n'a été oublié, n'apporte de nouveau fragments de texte pertinents que de façon *marginale*. L'extraction s'effectue à partir des fichiers « .txt » (logiciel WordPad) mensuels et avec la commande CTRL+F. Un second filtrage moins automatique a consisté à rejeter ceux qui étaient hors sujet, ce qui demande une connaissance de la communauté. Cette démarche a permis de constituer des bases de données comportant en tout près de 1400 éléments pertinents (extraits à partir de plusieurs dizaines de milliers de courriels). Enfin, un codage[827] thématique a permis de *catégoriser* les courriels pertinents mais aussi de repérer des passages pertinents (qu'on pourrait appeler, comme Hervé Dumez, « incidents critiques »).

Le problème du chercheur qui explore ces données est d'arriver à extraire le maximum d'informations sur les usages du Codex. Trois mots-clefs ont donc été choisis à cet effet :

-« Codex », le plus pertinent, car c'est le nom officiel du wiki, et toutes les URL citées comportent systématiquement ce terme dans leur extension.

-« Wiki », utile pour examiner les usages du premier wiki, et indispensable pour le début de l'étude de la liste de discussions « hackers ».

-« Doc », un terme racine pour « documentation », ce qui permettait de n'oublier aucun mot pertinent se rapportant à la thématique, et *permet de saturer l'extraction de données*.

Les mots « Codex » et « wiki » ont permis d'extraire approximativement plus de 90% des courriels concernés. Nous avons effectué un travail de vérification de la solidité de l'extraction de nos données en faisant une requête avec la racine « doc » (pour documentation). Le mot clef « doc » a permis de compléter la base de données des mails, mais aussi de se rendre compte que le processus d'extraction « saturait » : en effet, les mails

[827] Encore une fois, codage ne signifie pas forcément "théorie ancrée" (« grounded theory »).

comportant le mot « doc » comportaient le plus souvent l'un ou l'autre des mots-clefs précédents, et cette requête, si elle a peu apporté de matériaux, a permis d'établir clairement qu'il n'était *pas nécessaire d'aller au-delà*. C'est le cas tout particulièrement dans la liste de discussions « forum » : quasiment toutes les entrées comportant ce mot clef étaient déjà connues et comportaient l'un des mots clefs précédents. Dans la liste de discussions « hackers », il y en avait davantage, mais correspondaient le plus souvent à d'autres types de documentations (plugins, « inline documentation », i.e. documentation interne au code, documentation externe, etc.) : ce travail a demandé tout de même un discernement particulier dû à notre connaissance de la communauté. Tous les mails comportant le mot « doc » n'étaient pas pertinents, et *la familiarisation avec l'histoire de la communauté* du chercheur était nécessaire pour effectuer la sélection ; nous n'avons cependant exclu dans un premier temps que les courriels clairement hors de propos, et retenu ceux sur lesquels nous avions un doute, par sécurité[828].

Dans les fils de discussions, l'historique des conversations est conservé, et nous n'avons retenu que les mentions originales des mots clef, et non leurs répétitions dans les textes précédents des courriels (visibles car mise en évidence par des signets « > » plus ou moins empilés). Bien entendu, il peut être implicitement question du wiki dans le reste du fil, qui doit être examiné. Enfin, un fil de discussion peut comporter plusieurs mentions originales du ou des mots-clefs, et nous avons regroupé ces diverses mentions sous un même fil différencié en couleur des autres.

Voici une synthèse[829] donnant une idée pour l'instant globale de la taille des données et du nombre de fragments significatifs ; le total est de 1431 fragments:

---

[828] Nous donnons ici quelques détails sur ces opérations. Nous avons éliminé tout ce qui concernait la documentation extérieure, à laquelle la communauté peut faire référence, et les mentions à la documentation « interne », c'est-à-dire les commentaires des lignes de codes insérées fréquemment dans les fichiers de stockage, et qui n'ont rien à voir avec le wiki (c'est là cependant une thématique récurrente). Nous avons également éliminé ainsi toutes les mentions « doc » correspondant à un étage d'un arborescence de fichiers encadré par des slashs (« /doc »). Pour ce qui concerne les extractions au moyen du mot-clef « wiki », toutes les occurrences n'étaient pas forcément significatives : beaucoup de fragments correspondaient à d'autres wiki des communautés open-source ou à Wikipedia ; de tels fragments auraient été en revanche pertinents pour avoir une vue complète de l'histoire des usages des wikis (sur wikipedia, les articles dédiés à l'informatique sont parmi ceux dont la qualité est la meilleure). Cela sortait du champ exact de notre sujet, *mais rappelle l'importance des usages des wikis dans le milieu des informaticiens*. Nous avons aussi éliminé toutes les références au wiki sous moteur « Trac », dont la nature est différente du wiki sous Mediawiki.

[829] Les données concernant la liste de discussion de documentation, qui est l'objet d'un traitement qualitatif spécifique et d'une étude particulière, ne sont pas mentionnées ici. C'est cependant une source essentielle non seulement en « arrière-plan », car connaître ces archives aide à voir de qui est pertinent « historiquement » dans la communauté, mais aussi pour comprendre des éléments des débuts du forum, la régulation de ces deux tâches

Nom de la liste	Forum	Hackers	Trac	Testeurs
Datation des périodes d'analyse	Juin 05 – juin 07	Juin 04 – juin 07	Février 06 – juin 07	Mars 05 – juin 07
Nombre de mails	4408	18014	14646	4875
Nombre de citations	347	845	95	144
Ratio global	7,87 %	4,7 %	0,6 %	3 %

Tableau 5 : caractéristiques quantitatives du codage des listes de discussions

La nature particulière de la liste « Trac » qui se caractérise par des postages automatiques de mails et peu de textes explique sans doute le ratio plus faible et ce biais interdit d'inférer quoi que ce soit. Cependant, l'écart de pourcentage global suggère que le wiki est cité plus souvent dans la partie « utilisateurs » que chez les informaticiens. Là encore, la prudence est de mise avec les analyses quantitatives.

Quelques mots sur la qualité des données et les biais associés. Nous insistons, il s'agit d'une extraction *systématique* de données, et il ne s'agit pas d'un sondage ; cela autorise donc *une analyse qualitative fine et fiable* et permet de faire des inférences précises sans soulever de problème de représentativité des données de la sorte. Cela ne veut cependant pas dire qu'il n'existe pas de biais liés à l'énonciation. Bien entendu, les hackers ou les « mavens » du forum font usage du wiki, et ces mails ne décrivent *qu'une partie des usages*, parfaitement représentés dans le forum. Cependant, ce biais ne devrait pas forcément affecter *la structure* des usages, qu'on peut retrouver à partir des traces.

humbles se faisait au départ de façon commune. *Nous avons peut-être fait une erreur en codant séparément les deux listes, et il faudrait mentionner le nombre de fragments pertinents.*

8.2.4. Un codage dans une visée d'analyse historique sans faire usage de CAQDAS[830]

Nous disposons donc de sources riches sur les usages du Codex dans les autres parties de la communauté. Le problème est d'en tirer les informations les plus fiables (nous nous devons d'établir des preuves) et les plus riches (les morceaux de textes comportant le plus de signification) ; ailleurs, ceux qui ont moins d'importance doivent faire l'objet de quantification et classification. La quantité d'informations disponible *rend difficile une approche strictement inductive cependant nécessaire* : il n'est pas utile de tout décrire et classer de façon rigoureuse comme pour un codage normal, qui aboutit à un Codebook hiérarchisé. Il vaut mieux agir comme un archéologue, si possible de répondre à une problématique à partir des traces qui sont à notre disposition et extraites au moyen des codes, énoncées à partir de ce que la théorie commande de tester de plus pertinent, mais aussi partir de ce que la lecture et la mémorisation des fragments de dire de façon inductive (on repère des récurrences qu'on peut ensuite effectivement coder, mais sans s'embarrasser de classifications et d'arbres hiérarchiques compliqués).

Nous ferons ce que POINT et FOURBOUL (2006) appelle un « codage tous azimuts », « réseau conceptuel ». Nous pouvons appeler aussi cela « codage thématique ». Il n'aurait pas été satisfaisant de mettre en œuvre un codage axial (repérage de liens entre catégories de sorte qu'on les hiérarchise), les catégories pouvant au contraire être mobilisées avec plus de profil sous des registres thématiques différents en fonction des questions que nous nous posons. Il paraît aussi rétrospectivement impossible de distinguer les codes entre 1/conditions, 2/phénomène, étant donné la nature dynamique de nos fragments qui empêche de mettre en évidence des co-ocurrences. Le codage pour nous n'est qu'un outil de conceptualisation et de catégorisation thématique. Si la conceptualisation n'est pas une condition suffisante pour obtenir des conclusions, la pratique implicite observée en analyse qualitative est de réduire l'analyse au codage… Notre parti-pris, au contraire, c'est avoir une liste unique de catégories « de dernier rang », et de s'interroger à chaque fois sur *la liste des catégories* dont il faut extraire les morceaux de textes, afin d'être sûr de ne rien oublier. Les codes servent à faire des listes « d'items » à partir desquelles on peut se poser des questions, voire tester des hypothèses. Le codage n'est qu'un outil permettant d'extraire l'information de

[830] Le codage n'a pas été effectué par deux chercheurs (« double codage »), car nous n'avions personne de disponible pour mener ce travail. De plus, ce type d'activité comporte des inconvénients, i.e. un codage inadéquat, si l'autre chercheur ne prend pas le temps de bien connaître la communauté. Contrairement aux matériaux d'entretiens, le sens des courriels n'est pas souvent « immédiat », il nécessite un travail d'interprétation ce qui suppose des mois de familiarisation.

façon exhaustive et pertinente. Cette démarche est limitée, et, sur ce point même, elle n'est pas la seule: nous avons également effectué une autre forme de repérage des fragments pertinents en soulignant ceux qui pouvaient constituer des « incidents critiques ». De plus, le raffinement du codage n'est pas notre objectif premier : nous cherchons seulement, lorsque nous testons des hypothèses ou examinons telle ou telle thématique, à être sûr de disposer de l'essentiel des fragments concernés par nos recherches ponctuelles. Notre parti pris est qu'il vaut mieux travailler sur les biais, l'établissement des faits, les méthodologies de test d'hypothèse, et le degré de validité des résultats (comme le font les historiens) que sur les raffinements du codage, ce qui est économe en temps et plus rentable scientifiquement.

8.3. Résultats concernant le forum d'aide aux utilisateurs

Voici un résumé des résultats qui suivent : après quelques éléments sur le résultat du codage et la dimension quantitative des données, nous montrons que l'usage efficace du Codex comporte des *présupposés* et limites nécessaires qui ne sont pas attendus, ce qui est un résultat pertinent pour comprendre les limites d'un tel système d'information. Il s'agit d'une complémentarité nécessaire entre deux SI (forum et Codex), d'un usage qui ne concerne qu'une partie des utilisateurs, plutôt des utilisateurs possédant une expertise minimale et qui concerne davantage les comportements plutôt que les tâches à effectuer; enfin des mécanismes doivent aider (intermédiation) ou dissuader l'usage du SI (valorisation de l'expertise sans usage de liens vers le wiki).

8.3.1. Eléments sur les résultats quantitatifs du codage

8.3.1.1. Statistiques de qualité du codage

Les « ? » signifient que nous ne sommes pas parvenus à coder de façon satisfaisante le fragment en question. Les tirets « - » signifient que le message n'apporte pas d'informations supplémentaires à coup sûr ou fait partie de résidus de messages finalement hors sujet alors qu'il avait été retenu par précaution (second tri).

Nombre de "?"	67
Nombre de cellules	347
Pourcentage de cellules non-codées	19,31%
Nombre de "-"	36
Pourcentage de cellules sans intérêt	10,37%
Pourcentage d'information manquante	29,68%

Tableau 6 : caractéristiques quantitatives du codage de la liste « wp-forum »

Le pourcentage de cellules sans intérêt (nombre de «-») montre qu'un troisième filtrage des données était nécessaire après extraction des mots-clefs.

8.3.1.2. Apports des données quantitatives des fragments extraits et leurs limites

Nous présentons les données avant de présenter les conclusions qu'elles permettent de tirer. Voici tout d'abord un rappel de l'évolution de la liste de discussion du forum sur la période étudiée :

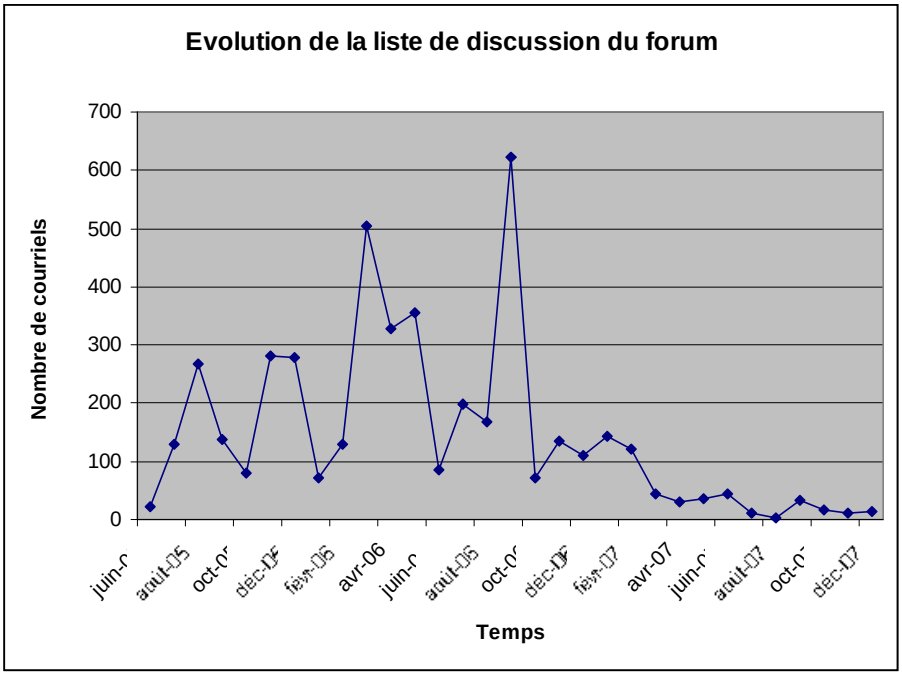


Figure 94 : graphique de l'évolution de la liste de discussion du forum utilisateurs

Elle comporte deux phases nettement différenciées, avant et après novembre 2006, l'une faste avec une forte variance, l'autre, plus calme et sans oscillations. L'évolution des courriels portant sur le Codex comporte la même structure :

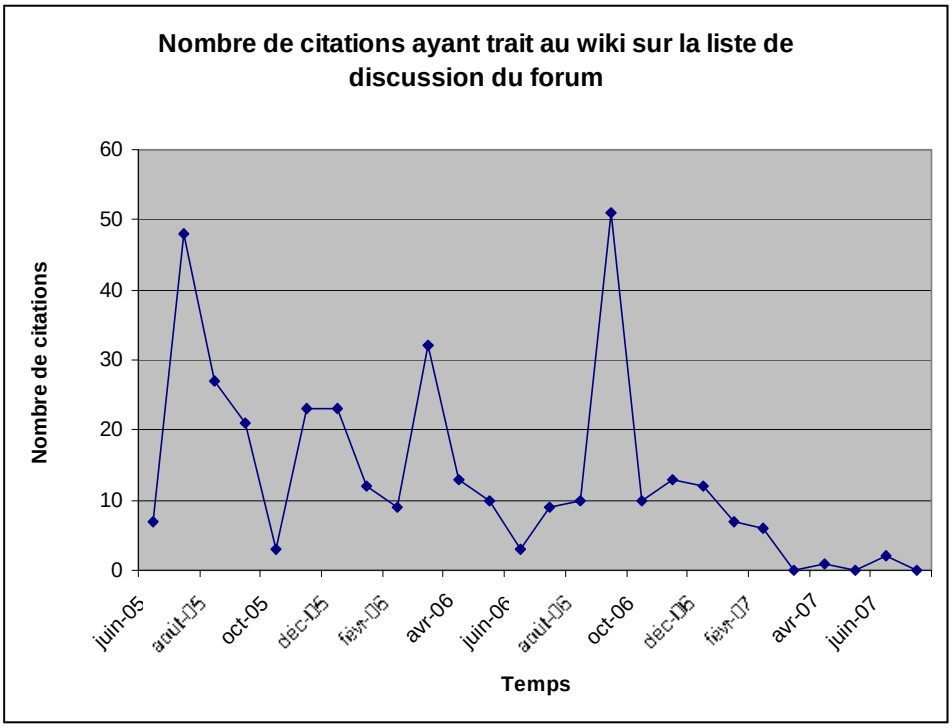


Figure 95 : pourcentage de citations ayant trait au wiki sur la liste de discussions du forum

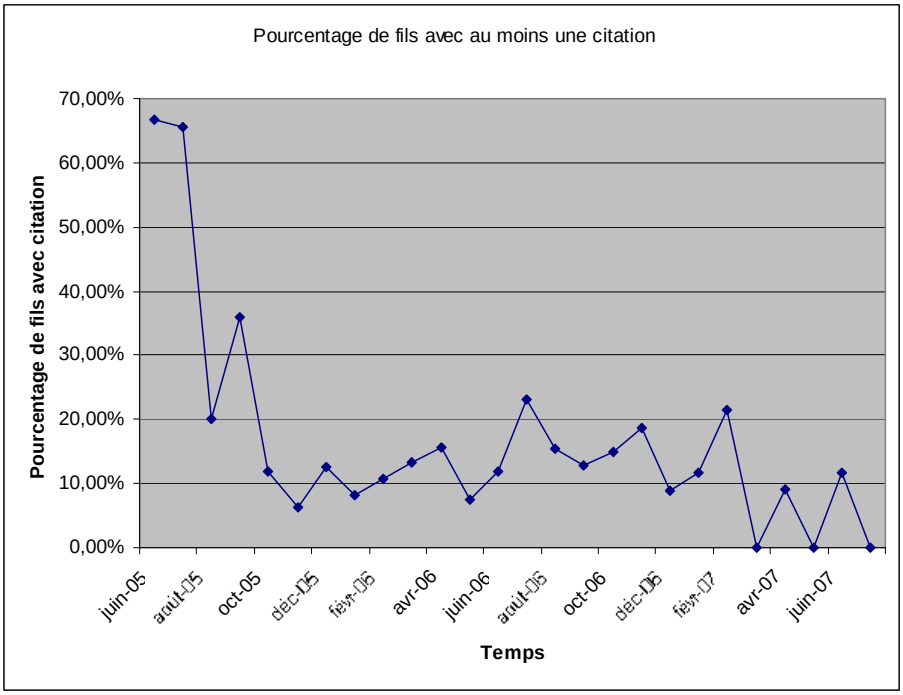


Figure 96 : pourcentage de fils avec au moins une citation ayant trait au wiki

L'interprétation de ces chiffres n'est pas immédiate, on sait les limites des approches quantitatives. *On peut cependant soutenir les éléments suivants* :

- La liste de discussion du forum a connu deux périodes, l'une correspondant à une phase significative de régulation, l'autre à une phase de maturité ; la différence est marquée autour de novembre 2006.
- Le nombre de citations liées au wiki a connu un mouvement semblable, l'un causant donc mécaniquement l'autre, avec la même date de rupture. Etant donné cependant que les chiffres sont faibles après cette date, l'interprétation à partir de ce point des pourcentages est plus sujet à caution.
- La question de la problématique de l'usage du wiki a connu une période « faste » en tout début de l'existence de la liste de discussion, ce que montrent *les pourcentages significatifs* de citations de juin à octobre 2005, (alors que l'autre pic entre août et octobre 2006 correspondrait plutôt à une activité ponctuelle de régulation), ce que confirme l'analyse qualitative des fragments Par la suite, la question des usages devient plus secondaire.

Les données quantitatives ne permettent pas d'en déduire davantage. Nous passons donc à présent aux analyses qualitatives.

8.3.2. Vers une complémentarité entre Codex et forum

8.3.2.1. Un système d'information qui ne fait significativement pas consensus

L'usage du wiki dépend de son acceptation par une partie de la communauté, en particulier de ses contributeurs respectés, qui peuvent approuver ou non un tel usage, autant pour des raisons d'efficacité que pour des raisons d'appropriation problématique. L'opposition la plus marquée à cet égard est celle entre Lorelle Van Fossen, qui le prône, et Podz, pourtant administrateur du wiki, qui en souligne largement les limites[831]. Il s'agit des responsables « officiels » de la documentation et du forum[832]. L'usage suppose aussi un consensus sur la qualité, ce qui est aussi discuté[833]. Plusieurs fragments montrent clairement que l'usage du système d'information ne fait pas consensus :

[831] Par exemple http://comox.textdrive.com/pipermail/wp-forums/2005-July/000052.html, mais beaucoup d'autres passages pourraient être cités.
[832] Source : http://wordpress.org/about/ (consulté le 02/09/09).
[833] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-March/001797.html .

> “I mostly agree with Podz on this. Codex is a huge and valuable resource, but for new users in particular, it's not the easiest thing to use.”[834]

Les sources montrent que tous les utilisateurs n’apprécient pas d’utiliser ou de contribuer à un wiki; c’est le cas dans ce fragment par exemple:

> “[18:11] <tunicwriter> photomatt_wrk: that's a sign of a bad system... not a good way to go about things.
> [18:11] <photomatt_wrk> michel_v: do you have a user page on the codex?
> [18:12] <michel_v> I don't think I have. I'm allergic to mediawiki :)
> [18:12] <photomatt_wrk> it'd be excellent to follow the progress of the document and invite comments
> [18:12] <photomatt_wrk> do you have a preferred method?”[835]

Enfin, des éléments montrent que les utilisateurs ignorent souvent la ressource du Codex. Les utilisateurs sont paresseux, d’après de nombreux fragments de volontaires chevronnés : ils préfèrent poser des questions qu’aller chercher à l’intérieur du wiki (ou aller chercher sur le forum) :

> “I agree about the docs..and users, but only to a point. I think we have a few types of user:
> * The user who is new to the whole thing, maybe had a friend install WP and now they've got an error and they panic. These people do not read, are probably worried, and need clear directions to solve things. While pointing them at Codex maybe useful, they may also just post back with "I don't understand" and we end up helping them. (and that's we do, what the forums are there for). (These people bump)
> * The user who is just lazy. Can't be bothered to search, won't read the sticky that has the answer they need. (These people bump)
> * The user who genuinely has tried but cannot understand for whatever reason, be that css, php, templates or whatever. They have the info, they just need a bit of stepping through it until 'the penny drops' (so
> to speak)
> * The user who has searched, but used the wrong terms. Doesn't really matter why, but if they came at it from a different angle then they got lost it's for us to help them out.
> * The user who genuinely trying something different and posts as much for inspiration as well as assistance.
> And lastly, the users who expect perfect docs, perfect help, for WP to function as they desire, who bleat that OSS shouldn't mean anything less than perfect and wish that we had phpBB so they could no doubt spend their time waiting by reading all the witty sigs and looking at the funky avatars.
> The challenge of course is meeting them all with the same smiling response .... and even with a perfect codex, we'd still need a forum. P. »[836]

---

[834] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-December/000947.html ; “Podz” est l’avatar d’un gros contributeur à la communauté.
[835] Source : http://codex.wordpress.org/IRC_Meetups/2005/July/July20RawLog .
[836] Source: http://comox.textdrive.com/pipermail/wp-forums/2005-June/000017.html .

8.3.2.2. D'une politique ambitieuse, la communauté, face aux difficultés, est passée à des objectifs plus réalistes de complémentarité entre SI

La finalité du Codex est économique, à savoir faire baisser le nombre de fils dans le forum dans un contexte où l'interdépendance entre les deux SI est forte, et inciter à la lecture du Codex. Ce point est bien documenté, mais nous citons un passage de Matt Mullenweg qui a le mérite de montrer qu'il s'agit d'un élément stratégique pour la communauté:

> "Kaf Oseo wrote:
> *> Uh, what? I've never heard of this "vision" of Matt's before.*
> *> Considering just how large a step this really would be, I wouldn't agree it's attainable, not with the current*
> *> technology out there (and most certainly not through the use of a wiki).*
> It's kinda like the dog that chases the car. It's a goal, but you don't really ever want to reach it. ;)
> The mindset is more, for every change that is made to the Codex and WP, will this increase or decrease support requests, to the extent that support is a byproduct of a non-intuitive product?
> I think about you guys a lot, even though there are certainly mistakes that have been made in WP in the past. They usually are fixed up within a version or two. Matt Mullenweg"[837]

La politique officielle du chef de projet était que le système d'information se devait d'être performant au point de rendre inutile un forum d'aide aux utilisateurs :

> "Matt's vision is to create a codex and piece of software so water tight that the forums are obsolete, therefore, the effort isn't in making the forums the most concise source of information, rather, it's a stepping stone to his vision. We that volunteer there must either accept that reality or find another area to volunteer in."[838]

Ou:

> "In fact that should be the goal of the documentation, to be able to answer any question with a link."[839]

Il s'agit donc d'un but affiché… tout en acceptant l'idée qu'il est impossible à atteindre. On peut qualifier cela d'idée régulatrice, plus précisément.

Nous pouvons donc plus précisément évaluer l'efficacité du wiki en fonction du souhait initial de Matt Mullenweg : le remplacement du forum par le wiki. Cette politique extrême mais affirmée a été un *échec*, c'est indiscutable si on la prend au pied de la lettre: le forum est toujours régulièrement nourri de questions, comme le montrèrent les statistiques du chapitre 4 d'histoire de la communauté à propos du forum. Le simple examen des statistiques des fils de discussion sur le forum et le nombre réduit de fils de discussion « WP-Advanced --

[837] Source: http://comox.textdrive.com/pipermail/wp-forums/2006-January/001260.html .
[838] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-January/001257.html .
[839] Source : http://comox.textdrive.com/pipermail/docs/2004-July/000381.html .

Beyond Codex questions » au milieu de la capture d'écran ci-dessous montre que cet objectif était trop ambitieux :

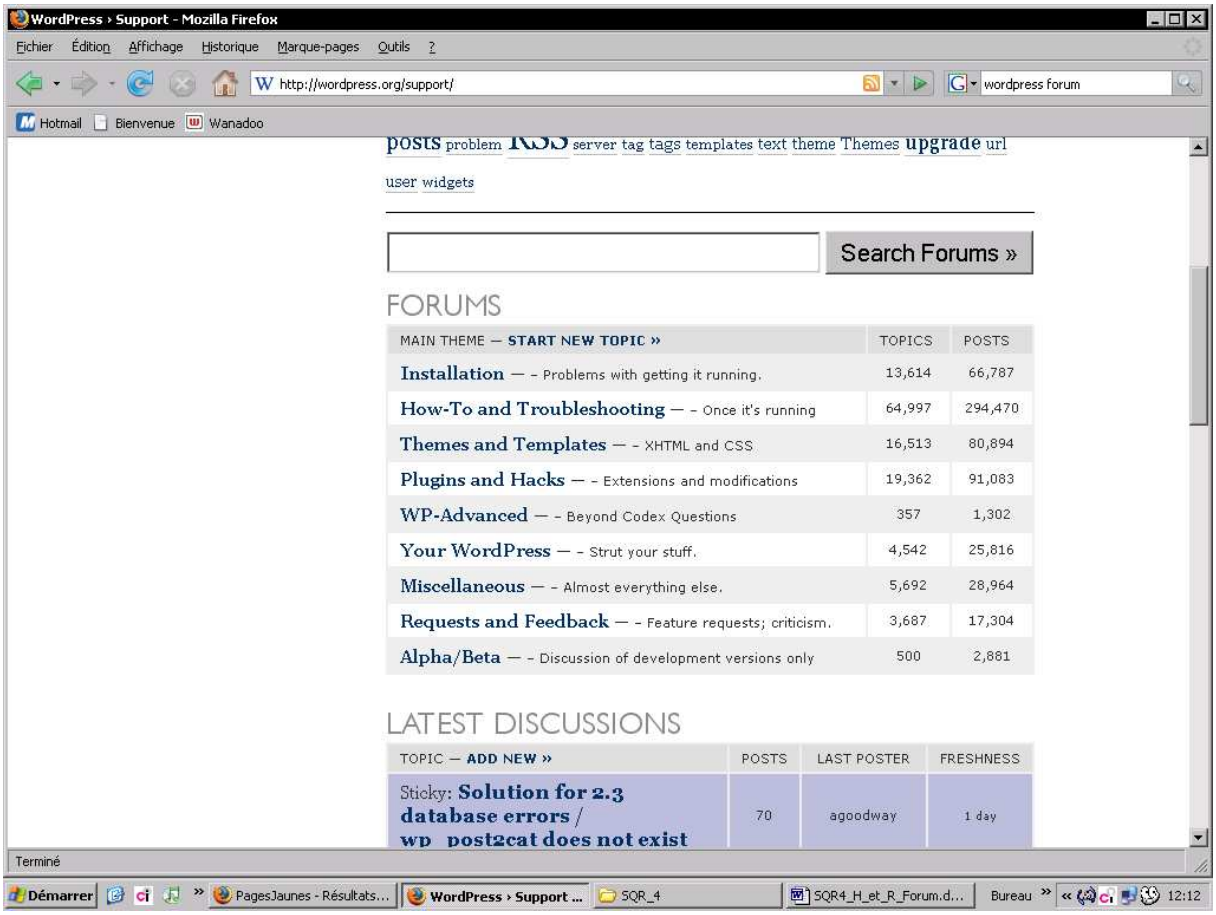


Figure 97 : capture d'écran de la page principale du forum

De tels chiffres concernant le nombre de questions auxquelles le Codex pourrait répondre, visibles dans les chiffres de toutes les catégories autres que celle bien visible au milieu, dite « WP-Advanced – Beyond Codex Questions », laissent perplexe. Il n'est pas aisé de fournir des conclusions définitives. Dans ces fils, le codex est souvent cité, comme le montre une exploration avec google à l'intérieur du forum. Mais est-il cité en permanence comme réponse aux questions par des liens vers lui… parce que les utilisateurs ne le lisent pas (il faut les y orienter, comme on le verra ci-dessous), voire parce que les utilisateurs ne lisent pas le forum… ou s'agit-il de fils de discussion auxquels le Codex aurait pu répondre, mais où il serait insuffisant ? Les deux sans doute, mais dans quelle mesure ? Il faut donc faire attention à l'interprétation qu'on peut en faire ; s'il y a échec à désengorger le forum, ce n'est pas forcément à cause d'une mauvaise qualité de l'information, mais parce que les utilisateurs

sont paresseux et ne vont pas voir le Codex. Il est difficile de le savoir sans creuser davantage le forum. C'est là une limite de notre travail.

On le comprendra, atteindre cet objectif est impossible, et reste un idéal. Même les leaders du Codex comme Lorelle ne sont pas d'accord sur cet objectif :

> "The Codex is a resource and provides a very strong backbone to support the Forum. It is not meant to replace it. Issues that come up repeatedly on the forum may deserve an article on the Codex. Many articles on the Codex link to discussions and solutions on the Forum. When an article substantiates a question on the Forum, link to it. They are very interrelated.
>
> If you really want to help the Codex AND the Forum, help us to spot topics that need to be in the Codex and then write them.
>
> Lorelle »[840]

Cependant, cet échec par rapport à un objectif aussi ambitieux n'est pas la seule façon d'examiner la question de l'efficacité. Nous avons montré dans le chapitre d'histoire de la communauté que la pente de l'activité du forum n'a pas augmenté dans la même mesure que le nombre de téléchargements, qui augmentèrent avec une pente nettement plus forte. Nous pouvons au contraire observer jusqu'à décembre 2005 une *efficience relative*, du moins en ce qui concerne le désengorgement de la fréquentation, mais nous n'en avons que des traces dans les données qualitatives, en particulier celle-ci, qui a du poids étant donné qu'elle est énoncée par Matt Mullenweg[841] :

> « But you guys should still realize that the Codex gets several times the amount of traffic as the forums. (Why do you think we had to move it to Yahoo?) I still strongly believe that time invested there pays off manyfold and if questions aren't being answered in the Codex FAQs, then it should be worked on. Also remember in the redesign that the Codex is going to be what people get when they click on "docs"[842].

Bien entendu, il est dommage de ne pas disposer de données quantitatives de fréquentation au-delà de la période de 2005. Les données qualitatives de la liste de discussion du forum sont donc un complément utile voire indispensable.

La question que soulèvent toutes ces données de provenance et de nature diverses est de caractériser la nature de la complémentarité entre les deux systèmes d'information. C'est sans doute la première question à se poser avant une éventuelle quantification ultérieure.

Une thématique plus pertinente, fine mais attendue, et qui apparaissait déjà dans la citation précédente de Lorelle, était la complémentarité entre savoir explicite, codifié, sur le

---

[840] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-July/000061.html .

[841] Nous ne pouvons disposer des statistiques de fréquentation différenciées en fonction des systèmes d'information examinés. Alexa ne donne rien en la matière. Quelques données apparaissent ponctuellement dans les archives IRC et les listes de discussions.

[842] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-December/000959.html .

wiki, et savoir tacite sur le forum, ce que montrent ces fragments de ces contributeurs critiques :

> "What I actually do not want are codex links all over the place - it's rare that you can home directly in on a codex topic that helps directly and Codex is linked from many places (like every WP blog). What is not linked is all the gold that is in the forums."[843]

Ou:

> "This exercise is not about replacing Codex, nor is it about removing anything from the forums. It is about using knowledge better."[844]

La complémentarité entre les deux était une priorité mentionnée dès le début de la liste de discussion[845] et elle est dans la continuation de la liste de discussion de la documentation qui mentionnait aussi le forum.

La question est contre l'apparence de nature *économique*, car l'effort des participants est une ressource rare et celle-ci doit être répartie entre le Codex et le forum[846]. Une autre question est de savoir si on oriente en priorité les utilisateurs vers le Codex ou vers le forum[847] : la décision de laisser le choix au lieu d'orienter de façon automatique vers le Codex a actuellement été retenue[848], car orienter les utilisateurs de façon automatique sans qu'il le sache comporte des inconvénients[849].

Parmi les codes, ceux évoquant l'interconnexion nécessaire entre Codex et Forum sont largement majoritaires (17 contre 3). Par ailleurs, ceux qui composent des articles ont manifestement tendance à les mentionner sur la liste de discussions (13 fragments). L'interaction entre les deux systèmes d'information est donc forte. L'analyse intra-code des fragments associés à la catégorie « Forum-Codex/Complémentarité » montre cependant qu'un tel tableau positif est à nuancer. Canaliser (« funneling ») les newbies vers le Codex est souvent difficile.

Un examen plus approfondi du forum et Codex pourrait montrer (nous ne pouvons en effet le prouver) qu'il existe un arbitrage entre *connaissance tacite (non structurée) et explicite*. Sur ce point, nous rappelons les propos de celui qui deviendra le responsable du forum, Mark Riley, dit « Podz » :

---

[843] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-December/000943.html .
[844] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-December/000956.html .
[845] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-June/000005.html .
[846] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-July/000053.html .
[847] Matt était pour http://comox.textdrive.com/pipermail/wp-forums/2006-March/001823.html .
[848] Nous avons retrouvé le vote qui laisse libre l'alternative (source : http://comox.textdrive.com/pipermail/wp-forums/2006-September/003459.html); il montre l'utilité d'une complémentarité stricte.
[849] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-March/001825.html .

> “What I actually do not want are codex links all over the place - it's rare that you can home directly in on a codex topic that helps directly and Codex is linked from many places (like every WP blog). What is not linked is all the gold that is in the forums.”[850]

Ces deux concepts décrivent cependant imparfaitement l’état de la connaissance, il y a en effet une part d’explicite dans le forum, mais la rédaction n’est pas standardisée et l’information non formatée ; c’est une réserve de solutions à des problèmes souvent spécifiques rangées de façon anarchique. D’un autre côté, le Codex offre des procédures plus générales, des solutions générales, avec un désir de rédaction correcte forte, mais moins précises et adaptées au cas de chacun.

Par ailleurs, le Codex a fait l’objet de pannes multiples[851], dont les sources ne donnent pas toujours une idée claire et cette insuffisance de nature technique a pu desservir un produit qui était en fait bon. Il faudrait se demander si la grande sensibilité des utilisateurs aux pannes du Codex ne dissuade pas de façon significative de l’utiliser alors que le contenu était satisfaisant, au profit du forum qui n’a pas été affecté par ce type de panne[852]. Il est difficile de cerner la portée de cette limite qu’il nous fallait signaler.

8.3.2.3. La complémentarité entre les deux SI suppose une forme de régulation par une population commune de volontaires.

En cohérence avec ce qui précède, une mesure possible de l’efficience de la documentation, peut se mesurer par le degré d’intégration entre deux sous-communautés, une condition essentielle et nécessaire de la question de la complémentarité que nous évoquions. Une séparation aurait été une *incohérence*[853]. *Nous nous attachons ici à réfuter cette hypothèse*. Cela ne veut cependant pas dire que le Codex est efficient selon d’autres dimensions. La comparaison entre le graphe de la liste de discussion du forum et le graphe de

[850] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-December/000943.html .

[851] Les exemples de rapports de pannes sont multiples. Nous citons ici un passage qui montre avec un certain humour que les pannes affectent directement la perception de l’utilité du Codex : « spencerp wrote: > does someone wanna Copy all the content from it, and start up a NEW Codex, on a NEW domain, and awesome server!? - Already did. It's called Google cache. ;)-Kaf » (source : http://comox.textdrive.com/pipermail/wp-forums/2006-November/003962.html)

[852] Nous ne nous rappelons cependant pas avoir observé de courriel intitulé « forum down ».

[853] Nous rappelons qu’un SI se caractérise non pas seulement par des outils techniques, des processus, des informations, mais aussi par une population d’intervenants.

la liste de discussion du Codex montre que les deux populations se recoupent largement au niveau des contributeurs significatifs de la documentation[854] :

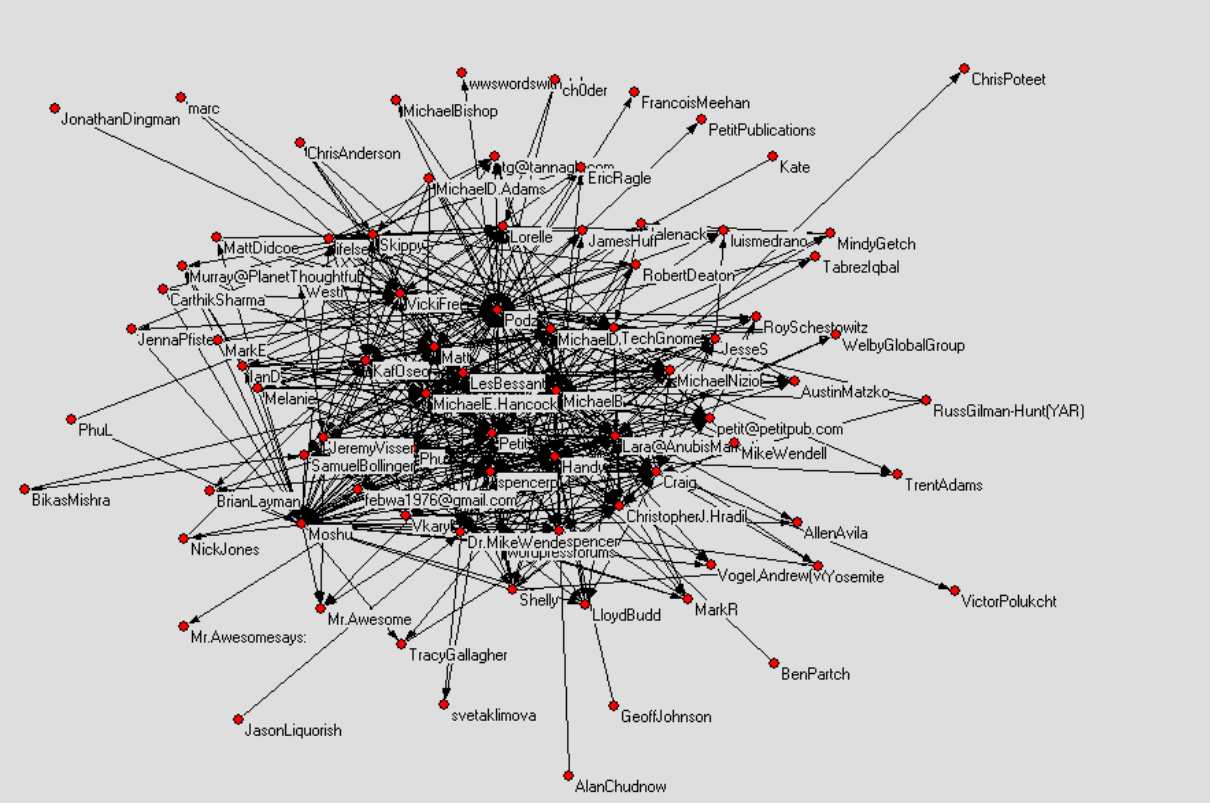


Figure 98 : graphe de la liste de discussion du forum

Nous renvoyons au chapitre précédent pour une visualisation du graphe de la liste de discussion de management de la documentation. La principale remarque qui vient à l'esprit lorsqu'on examine ces deux graphiques, c'est que les participants ayant le plus fort taux de connexion sur les deux listes de discussions se recoupent de façon significative. Nous citons seulement les administrateurs :

- il s'agit de Craig, Lorelle, Skippy, MichaelH, Podz, Matt, Kaf ;
- ce n'est pas vrai pour Carthik, MdaWaffe ;
- Westi, un programmateur, est moins pertinent, car il intervient moins ; Foolswisdom n'est choisi comme administrateur que bien plus tard.

Parmi les non-administrateurs, de nombreux « avatars » sont communs : Moshu, spencerp, Michael D, Techgnome.

[854] Pour la méthodologie de traitement de données, aucun changement n'est à signaler par comparaison avec le chapitre précédent.

Cela peut aussi signifier que l'une des conditions au transfert de connaissance d'un système d'information à l'autre est concrétisée.

En revanche, un passage rétrospectif montre l'inconvénient d'une telle régulation *centralisée* ; comme si les inconvénients de la régulation du Codex avaient un impact sur celle du forum (ce n'est qu'une hypothèse) :

> "An interesting read. Having been heavily involved at one time in both support and documentation, the one thing that comes to mind right now is development. I think that WP would benefit from developing people in a semi-formal way. Mentorship is one way.
>
> For example, I was keen on documentation. Along with Podz and Carthik and others, we got the initial wiki going (after we convinced Matt LOL) and then made the transisition over to MediaWiki and Codex. This is where our lack of system would have helped me...I was not wiki-savvy. Soon we had lots of good folks jumping in to help out, and they were indeed wiki-savvy. However, what happened was that soon these people were basically doing everything, while people like me wound up being relegated to the sidelines.
>
> Please do not construe this as an attack or sour grapes. I could have learned on my own how to do a lot of things, but the point I'm trying to make is that had I been paired up with someone who was more advanced than I, we could have had two "experts" in time. Then we train two, and they train two etc. You get a CSS-effect in your documention team development. The same would be true for support.
>
> Right now *the support and Codex are dominated by a small group*. Again, this is not an attack or judgement, it's simply the reality of a lot of work to be done, and only a core group of folks with the occasional help from a lot of others.
>
> I believe that WP has grown to the point where there needs to be some "middle management" and those folks who would so choose could become involved towards managing support and managing documentation almost as distinct open-source projects of their own.
>
> I hope that I have made some sense in this post. I'll ruminate a bit more and add to this discussion later.
>
> Craig. »[855] .

Enfin, d'après les sources disponibles, un unique passage (« incident critique ») pourrait montrer des oppositions entre documentation et forum, ce qui ne permet que de nuancer qu'à la marge notre présente argumentation :

> "If this double posts, sorry, but my host is being a little funky ATM.... Otherwise, this is what I was trying to send in the first place:
>
> You know what? It's documentation, right? Let the wp-doc people deal with that.... we'll simply post in the forums that the wp-doc people are supposed to working on it, and if they aren't well.....
>
> -tg
>
> Podz wrote:
>
> >*I am livid about this.*

[855] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-October/003788.html. C'est nous qui soulignons le passage le plus significatif montrant l'interconnexion entre les deux groupes.

>It's been zero communication. And I do mean ZERO. Would it have hurt to send a private email mentioning that things could have got busy ? No it would not.

>So, guides:

>- Codex has one for upgrading.

>Permissions ?

>- Codex has that too.

>See you much much later in the week - I found things to do.

>P."[856]

L'hypothèse la plus vraisemblable d'après nos sources est donc que la documentation et le forum ne sont pas deux sous-communautés significativement séparées.

8.3.3. Dans quelle mesure peut-on qualifier la qualité de la documentation pour les utilisateurs et contributeurs du forum ?

8.3.3.1. Un outil comportant des insuffisances significatives mais non rédhibitoires, et perçu comme orienté « expert »

Nous pouvons observer des périodes localisées où l'information sur le Codex entre en crise, ce d'autant plus qu'on observe des coïncidences chronologiques entre les listes de discussion en la matière. Il en est ainsi du long passage placé en premier en annexe. Le texte de cet « incident critique » dur (« flaming ») montre que les causes de la mauvaise qualité de la documentation sont multiples : pas d'implication des hackers pour vérifier les points techniques les plus délicats, manque d'implication des leaders de la communauté, éviction des contributeurs en cas de dépassement d'un certain seuil de complexité ou confusion du wiki (un wiki qui est laid dissuade d'y remédier), refus de s'impliquer lorsqu'il s'agit d'écrire de la documentation sur des points délicats comme la sécurité, inconvénient du « release early » lorsque la documentation ne suit pas, confusion entre les versions du logiciel et incompatibilités qui ne sont pas résolues dans la documentation… Bref, c'est là un tableau général des origines des insuffisances. Le Codex est limité et n'est pas la panacée, et suppose en complément un autre système d'information :

"1. Not everyone uses Codex.

2. From the front page of Codex, the FAQ is not very clearly linked.

3. The format of Codex does not lend itself to people who are worried/confused and searching for information .

[856] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-March/001557.html .

4. This http://codex.wordpress.org/FAQ leads to TEN MORE FAQ's so you are asking someone who could be worried / confused to make decisions about which category their problem falls into. […]

9. If Codex was the answer to all problems, why have we got the forums ?

10. Because sometimes those Codex FAQ answers generate more questions which it does NOT ANSWER. […]

Now if someone wants to discuss this, please do from the point of view that Codex is not a religion nor the answer to all the worlds problems.

And now I'll have a coffee.

P. »[857]

On observe donc, et c'était attendu, des formes d'expressions de mécontentement vis à vis de la qualité du Codex. C'est là l'une des conditions de l'amélioration du SI. Cependant, l'*hypothèse que le niveau de qualité du Codex serait significativement mauvais spécifiquement pour les nouveaux utilisateurs du forum est-elle acceptable ?*

L'une des conclusions qui saute *à première vue* aux yeux est que le décompte des codes indique une approche *plutôt négative* du codex : deux fois plus de *qualification* négative de l'information, mais aussi du Codex en soi. Cela est relativement cohérent. Cependant, toute conclusion quantitative négative s'expose à la critique qui consiste à rappeler que les utilisateurs rapportent les informations qui ne vont pas, et non ce qui fonctionne. Pourtant, on observe de nombreux fragments de codes positifs significatifs qui invalident cette hypothèse, en particulier 19 codes positifs pour la catégorie « Codex ». Enfin, un début de croisement des sources à partir de l'analyse des fragments issus du dépouillement des fils « taggés » « codex », « wiki », « doc », montre un nombre significatif d'opinions positives[858].

Par ailleurs, les connotations associées au Codex le qualifient-il comme expert ou débutant ? 15 morceaux de codes suggèrent une connotation « expert », 7 suggèrent une connotation débutant. Ainsi, en ce qui concerne l'usage pour les utilisateurs novices, la qualification plus experte est une limite à l'efficience du système d'information ; *pour le forum*, elle l'est, elle serait moins gênante pour les hackers. Plusieurs fragments à connotation débutant appartiennent au même fil de discussion, ce qui limite encore le poids de cette catégorie.

Voici trois fragments suggérant une « connotation expert » :

[857] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-December/000946.html .
[858] Par exemple http://wordpress.org/support/topic/35497 .

> “I'd appreciate it if someone could compose a quick forum post about this, and also throw mention of it onto the Codex. I tried, but decided that I couldn't get the proper "tone" for such an announcement on the Codex.”[859]
>
> “I mostly agree with Podz on this. Codex is a huge and valuable resource, but for new users in particular, it's not the easiest thing to use.”[860]
>
> “If we can answer a question we do, and if not we try to point to a page in the Codex or elsewhere that seems to have an answer.”[861]

Le deuxième se catégorise aisément, le premier est un peu moins transparent car il suggère non plus une difficulté de lecture mais une difficulté à intervenir sur le Codex (bien que ce soit un wiki) ; enfin, le troisième suggère que ne pas savoir répondre à une question suppose une délégation à un système qui est plus « expert ».

La dominance est plutôt à considérer le Codex comme un outil difficilement accessible sémantiquement. Toute la question est de savoir s’il y a un lien entre la qualification négative du Codex et sa perception comme outil expert.

Le fait que le Codex soit plutôt considéré comme expert renforce l’idée qu’une forme d’intermédiation par une communauté d’experts est nécessaire, un point que nous avons abordé juste au-dessus.

Enfin, le codex sert-il à réguler les comportements ou les tâches ? Les codes à mobiliser sont tous ceux associés à la « régulation ». Les données disponibles permettent de conclure que le mode de régulation associé au Codex est surtout celui des comportements, ce que montre clairement la dominance de l’un des ensemble de catégories sur l’autre : 23 contre 4. Cette fois, il n’y a pas de biais d’énonciation à notre connaissance, et un raisonnement quantitatif est à notre connaissance acceptable. C’est l’une des conclusions statistiques qu’il est possible de tirer rapidement. Voici un fragment ainsi codé :

> « On the "Forum Rules" page I drafted: http://codex.wordpress.org/User:Skippy/Forum_Rules I said: "The WordPress Support Forums (http://wordpress.org/support/) are made available, free of charge, for the express purpose of helping users with WordPress problems. That's a pretty big umbrella, and a lot of things can reasonably be considered "WordPress problems": PHP, CSS, XHTML, hosting, technical diagnostics, and a whole lot more. It will be impossible to enumerate all permissible subjects; but generally speaking if it has to do with WordPress it's probably acceptable."[862]

Cependant, si un usage domine, il est difficile d’en tirer des conclusions en matière d’efficacité.

---

[859] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-November/000640.html .
[860] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-December/000947.html .
[861] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-March/001999.html .
[862] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-September/000506.html .

L'hypothèse la plus vraisemblable, c'est qu'il y a eu des insuffisances notables et récurrentes du Codex pour les utilisateurs normaux et novices, perçu en cohérence avec cela comme expert plutôt que débutant, mais qu'elles sont à relativiser du fait de l'ampleur de la tâche et de leurs exigences. Qualifier le niveau de qualité est complexe.

8.3.3.2. Efficience du Codex, une centralisation nuancée

L'analyse des archives dédiées aux usages par les utilisateurs montre l'étendue des références au Codex et montre ainsi que le Codex est devenu malgré tout une référence dominante en matière de documentation et a évincé les autres formes malgré ses insuffisances. On peut voir dans ce processus un phénomène de « path dependency », i.e. le standard en matière de documentation reste le Codex, qui avait un avantage en la matière. C'est là une confirmation des résultats que nous avions mentionnés en SQ1.

Cependant, il apparaît que le Codex comme lieu central de l'information a *fait débat*, comme le montre le nombre significatif de fragments codés à ce sujet (6 pour « Codex centralise information », 6 pour « Autre forme de documentation suggérée »). Dans quelle mesure ces données remettent-elles en question le principe de centralisation de la documentation ? Finalement, la documentation concurrente proposée par Podz, la principale partie prenante dans cette affaire, a été fermée car non mise à jour..

8.3.4. Mécanismes favorisant ou limitant l'usage du Codex

8.3.4.1. L'usage du Codex suppose l'émergence de pratiques d'intermédiation et d'orientation des utilisateurs

Nous allons montrer ici que l'usage du Codex n'est pas systématique, « immédiat » et « naturel » pour les utilisateurs, contre l'apparence. On observe que la façon de faire usage de la documentation, d'orienter les utilisateurs vers les bons endroits de la documentation est fondamentale, et suppose une *routinisation des usages*.

Le passage suivant d'un administrateur et responsable respecté du forum souligne d'une part l'idée de connaissance tacite (i.e. non codifiée et structurée) associée au forum, d'autre part que citer le Codex spontanément et systématiquement pour aider les utilisateurs novices ne va en fait pas de soi :

> “What I actually do not want are codex links all over the place - it's rare that you can home directly in on a codex topic that helps directly and Codex is linked from many places (like every WP blog). What is not linked is all the gold that is in the forums. And like TG suggested, slap those FAQ tags where needed. P. »[863]

De plus, l’appropriation du résultat du processus de documentation suppose des pratiques d’intermédiation au niveau du forum et des routines associées. Le résultat de la collaboration massivement distribuée est un système d’information, ce qui suppose que les participants adoptent des pratiques qui permettent d’exploiter le savoir accumulé. Dans le cas qui nous intéresse, les traces relevées sur la liste de discussion suggèrent que ces pratiques ont été adoptées au début de l’existence de la liste de discussion du forum :

> “Although it will be nearly impossible to articulate successfully, I would very much like to also require some measure of "decorum": it's insufficient to just give the right answer. One must be demonstrably supportive (within measure) when providing answers, instead of just spitting out correct facts without context.
>
> (Yes, I've changed my mind a bit on the practice of merely dropping links to the Codex; thanks largely to this list.)”[864]

Ou encore:

> “ASK if people have been to Codex, do NOT assume they have. READ their questions for heavens sake too - there are some answers which illustrate only that the person who pasted the answer is actually keener on raising their stats that they are in seeing a satisfied user. (Which is why public displays of stats are such a bad idea). And ANSWER them - any links to Codex (except for VERY clearly stepped answers) should be supplementary to what you type. They should not be the only thing you write.”[865]

Ensuite, plus généralement, la notion de “manuel” et l’emploi de la réponse RTFM (« read the f[censuré] manual ») sur le forum pose problème, et la notion de page suppose une connaissance du Codex pour orienter vers les bons endroits :

> « http://wordpress.org/support/topic/79820
>
> I think moshu did a great job handling this, however I have to say that were I to get hit with "RTFM" myself, I'd be inclined to react similarly... I don't know where that abbr started, and I don't much care, but it's rude... and the person who used it here, has been seen using it all over the forums as of late. I don't know if he just discovered it or what...
>
> Since we all know what it means, spelling it out would give reason for a thread to be moderated, so why is the abbreviation alright? I have no problems with saying something to the effect of, "It's right here in the codex:" and giving the link. However to me, RTFM sounds like blatant arrogance smacked with a taste of "safe" profanity.
>
> Just my opinion...

[863] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-December/000943.html .
[864] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-July/000087.html .
[865] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-July/000143.html .

~ Lara »[866]

L'outil ne se suffit pas à lui-même, ce qui a demandé des formes d'apprentissage de son usage, d'essais et d'erreurs, et des débats. Cet apprentissage d'une pratique de citation et d'intermédiation qui serait une autre façon d'utiliser quand même le forum est une *limite par rapport à la politique officielle* de désengorgement du forum mentionnée précédemment, car cela augmente la taille des fils. Cette politique de désengorgement, qui suppose d'orienter les utilisateurs vers le Codex, doit se faire avec une certaine efficacité, sinon les utilisateurs mal orientés reviennent :

> "But my question to this is, is the primary goal to funnel people to the codex, or is it to help them solve their problem? I'm finding all too often, someone asks a fairly specific question, one or several members chime in, and then along comes, IMVHO[867], a generic link to the codex. Sometimes even after one link to the codex has been offered up in the thread. If the real goal is to funnel questions to the codex, then perhaps my efforts would be better served working on the codex, rather than working one on one with forum members."[868]

Plus précisément, l'équilibre entre réponse sur le forum et orientation vers le Codex suppose une forme de compromis entre citations de liens, plus aisée, et explications précises, plus coûteuses en temps :

> "I was talking to someone today who volunteers with WordPress about the issue of posting links to help users. He didn't agree. He thought volunteers should HELP and not just post links. I thought there should be a middle ground: help when help is necessary, link when help is easy to read through, and do both when it is appropriate. Anyway, the debate will continue but I thought this post was a delightful response to this issue and I wanted you all to see it:
> http://wordpress.org/support/topic/38172
> Lorelle »[869]

L'orientation des utilisateurs comporte aussi des aspects automatisés, ce qui est une forme de régulation. Automatiser la recherche en priorité sur le Codex, comme cela a été fait, et non plus sur le forum ne va pas de soi. Nous citons ce passage ironique, qui montre que les arbitrages en question par la communauté ont des implications immédiates sur le comportement des utilisateurs, via le paramétrage du système d'information. :

> "Much better. Only problem I encountered was searching for something I KNEW was in the forum, but got sent to a no results page because the default choice for search is docs. That's great from the aspect of "real people search" though! Nice. Thank you VERY much. V »[870]

---

[866] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-July/002913.html .
[867] « In My Very Humble Opinion ».
[868] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-July/000053.html .
[869] *Source : http://comox.textdrive.com/pipermail/wp-forums/2005-July/000062.html* .
[870] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-March/001825.html .

L'orientation automatique vers le Codex est donc inefficace si la documentation n'est pas suffisamment au point. Ce passage a l'intérêt de montrer que le management du forum suppose une connaissance fine des pratiques des utilisateurs.

Bref, l'usage ne va pas de soi, ce qui suppose ici *1/ des intermédiaires 2/ des pratiques correctes d'intermédiation et d'orientation.* Le Codex est une ressource qui ne se suffit pas à elle-même *: elle nécessite des formes d'intermédiation (qui ont supposé un apprentissage collectif). Les gros contributeurs réfléchissent aux pratiques de citation qui se maintiennent :*

> « As has been said before, there is nothing wrong with posting a link to an article on the Codex that will help the person. Describe what the article will help with, and give specific help if you can, but if the article will "hold their hand"...let them have it. ;-) More importantly, with the focus on "Search First", six months or two years from now people will search and find these links to Codex articles as well as the help in the thread and be thrilled to have found the answer they need. We're building a huge resource here."[871]

8.3.4.2. La valorisation de l'expertise et des conseils donnés limiterait les citations, et donc l'usage de la documentation

Un tel effet va dans une certaine mesure à l'encontre de la politique prévue de désengorgement du forum.

La qualité de l'aide en ligne est essentielle pour que la communauté fonctionne bien; cependant, celle-ci n'est pas toujours au rendez-vous lors des citations en question :

> "On more than one occassion Podz has mentioned that there are people who "consistently give poor advice" in the Forums. In general, I believe people try to give the best answer, even if it is just a link to Codex or a link Podz's site. But when it happens, is there a diplomatic way to inform people they've provided the incorrect answer to a question? You'd like to think that anyone giving incorrect answers, would appreciate knowing the CORRECT answer, so that next time they can do a better job of answering the question. If we know people are providing incorrect information, isn't it our job to point that out and educate the forum volunteer? Thanks for any thoughts on this matter.
>
> Michael E. Hancock"[872]

Nous partons de l'hypothèse raisonnable et documentée que dans les communautés open-source, montrer son expertise en répondant aux questions des novices est source de motivation (CONEIN, 2003). Etant donné qu'il est aisé de citer des liens vers des pages du Codex non-pertinentes pour augmenter ses statistiques de participation, la valorisation des

[871] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-July/000052.html .
[872] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-April/002257.html .

réponses « automatique » en postant des liens vers des textes écrits par d'autres pose problème en conséquence. Une question qu'il faudrait se poser, c'est dans quelle mesure ce type de citation du manuel casse le processus de valorisation de l'expertise de ceux qui répondent aux questions sans y recourir. Il est difficile de le savoir, car citer le Codex demande quand même un certain niveau d'expertise étant donné que l'information n'est pas toujours aisée à trouver. Il faudrait des données fines pour répondre à une telle question ; ce passage (un « incident critique ») soulève le problème, mais ne le résout pas, car il est ambiguë (connaissance du Codex requise, mais pas essentielle) :

> « Going back to the qualifications you listed, let me ask some questions and make some observations, ones I'm not sure have been hit:
>
> 1. Have consistently over a period of time answered posts correctly in the forums.[…]
>
> 2. To have a sound knowledge of the Codex
>
> * Do full answers count more than just providing a Codex link?
>
> * What's the ruler used to measure this?
>
> There are folk on the forums right now who dispense good to excellent help, while they rarely to never refer or link to the Codex. If using Codex in this way represents this knowledge, are they not to be considered? »[873]

Par ailleurs, la connaissance tacite valorisée (car fine) se trouve dans le forum, comme nous l'avons vu[874].

Est-ce que la valorisation de l'expertise dans le monde du libre dissuaderait d'utiliser le codex ? Il est clair que le mécanisme de démotivation existe (d'un point de vue économique), mais il est difficile d'en estimer la portée. Cela obligeait cependant à en parler.

[873] Source : http://comox.textdrive.com/pipermail/wp-forums/2005-July/000133.html .
[874] Rappel de la source : http://comox.textdrive.com/pipermail/wp-forums/2005-December/000943.html .

8.4. Résultats pour les informaticiens : efficience de la documentation, qualité perçue, les mécanismes observés

8.4.1. Aspects quantitatifs

8.4.1.1. Statistiques de codage des listes de discussion des informaticiens.

L'arrêt de la codification en juillet 2007 permet de disposer de 868 fragments codés après un premier tri des premiers fragments manifestement étrangers au sujet ou n'apportant à coup sûr aucune information supplémentaire et dont l'énumération est parfaitement inutile.

Les symboles ont la même signification que précédemment. Le nombre de « ? » montre que le codage ne laisse que peu de résidus (moins de 7% au total). Le nombre de tirets « - » montre en revanche l'importance du second tamisage « sémantique ». Leur nombre est probablement dû à la nature technique des listes, au contenu plus pauvre « sémantiquement ». Mais il était nécessaire de les prendre en compte quantitativement, car ils se *réfèrent* au Codex.

	Hackers	Testeurs	Trac	Total
Nombre de fragments	827	143	110	1080
Nombre de fragments codés	656	124	88	868
Nombre de "?"	51	12	12	75
Nombre de "-"	120	7	10	137
Pourcentage de "?"	6,17%	8,39%	10,91%	6,94%

Tableau 7 : caractéristiques quantitatives du codage des listes des informaticiens

Quoique plus nombreuses en nombre de fragments, les données sont nettement plus techniques et difficiles à interpréter que pour la liste de documentation ou de régulation du forum, ce qui est pourrait être une limite de notre travail en matière de fiabilité. Cependant, il n'est pas impossible en soi d'en tirer des conclusions solides, à condition de ne pas faire que du qualitatif, afin de pouvoir dégager des tendances de fond. Le type d'extraction que nous avons effectué n'implique pas une homogénéité de traitement.

8.4.1.2. Une permanence de l'usage du Codex d'après les statistiques des fragments[875]

Voici ce que donnent les fragments codés, après décompte sur Excel et mise sous forme de séries chronologiques :

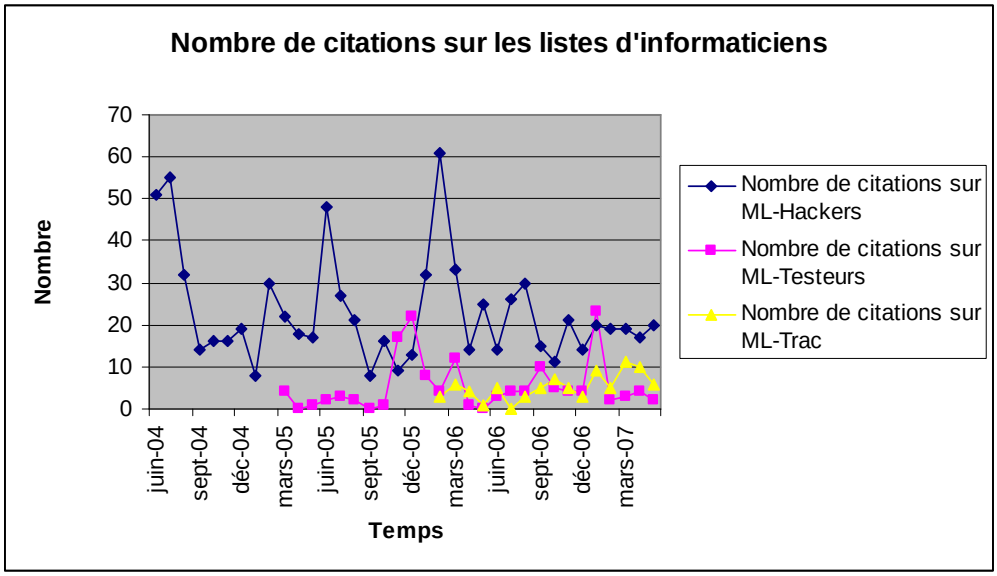


Figure 99 : nombre de citations sur les listes d'informaticiens

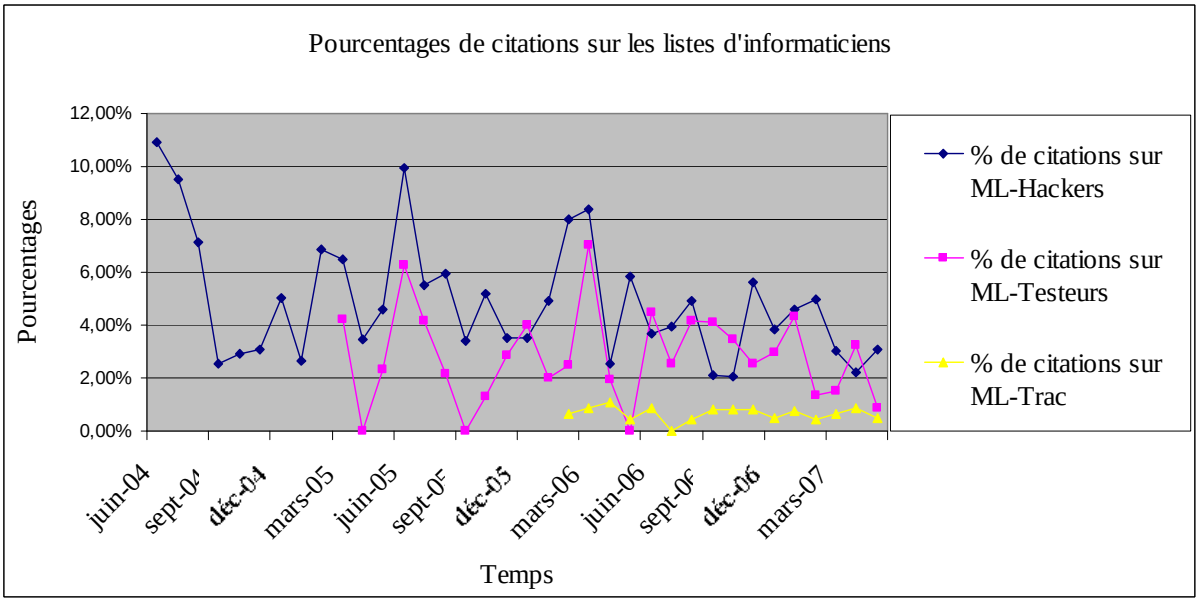


Figure 100 : pourcentage de citations sur les listes d'informaticiens

[875] Nous ne reprenons pas le schéma du débit mensuel de publication sur les trois listes, donné dans le chapitre de l'histoire de la communauté.

Contrairement à ce dont nous disposons pour le cas du forum, les données quantitatives permettent de tirer des conclusions plus claires, étant donné que ces citations correspondent *directement* aux usages. Dans les trois listes de discussions des informaticiens, qu'observe-t-on ? Le Codex est manifestement utilisé par les trois sous-communautés, malgré une baisse légère pour les hackers et les testeurs. Cette approche plus quantitative de la présence mensuelle des fragments étudiés montre en tout cas la *permanence* et la régularité de l'usage du système d'information durant toute la période d'étude dans ces populations de contributeurs : les usages ne tombent pas à zéro. Il est moins utilisé dans « Trac », que ce soit en chiffres bruts ou en proportions; mais cela peut se comprendre aisément étant donné le débit conséquent de la liste et l'absence de vraies « discussions » sur cette liste, qui consiste essentiellement en tickets numérotés d'identification des « bugs ».

On le sait, le Codex comporte une section développeurs. La fonction d'aide à la programmation, de guidage des opérations de codages ou d'amélioration du logiciel est donc observée dans les citations, qui ne sont qu'une partie des usages, le reste étant invisible. Mais sont-elles pour autant efficaces ? Quelles sont les connotations associées aux citations ?

8.4.2. Aspects qualitatifs : dans quelle mesure observe-t-on que le Codex est satisfaisant ?

Le caractère fortement contextualisé de chaque liste de discussion suppose selon nous de poser la question pour chacune d'elles d'abord, au lieu d'adopter une présentation directement comparative, au risque de paraître trop descriptif et de trop juxtaposer. On pourra donc toujours critiquer ce choix, qui nous a fait hésiter, mais ce qui importait était davantage la solidité satisfaisante des conclusions. Nous commençons par les deux listes les moins conséquentes par le nombre de citations pour finir par la plus pertinente.

8.4.2.1. Usages et efficacité sur la liste de discussion « Trac »

Sur la période de février à octobre 2007, l'histoire de l'usage du wiki par cette liste de discussion laisse une impression de sérénité (pas de conflits et d'accrochages sérieux liés à la documentation), malgré l'observation d'insuffisances. La qualification de l'information est largement positive, avec 42 fragments dans cette catégorie, sans compter les fragments dénotant un usage simple, non connoté négativement. Ainsi, cet utilisateur dit que le Codex lui a permis de s'instruire :

> “Comment (by redsweater):
> Good points, Joe. I there should be extreme caution as to not obliterate existing fields when a client doesn't supply a value.
> I didn't really know much about the custom fields until a customer pointed me to this wiki page today:
> http://codex.wordpress.org/Using_Custom_Fields
> So yeah - it seems like it's just sort of a generic repository of keyed values, and plugins can use them for nifty purposes of their choosing.”[876]

Les fragments soulignant la mauvaise qualité de l’information ou mettant en doute le contenu sont de 16, par opposition à un usage majoritairement positif sur l’essentiel des fragments. Parmi eux, seuls 3 fragments ont été codés avec « information fausse » (le code le plus « sévère »). On n’observe par ailleurs aucune forme d’incident significatif concernant le wiki (contrairement à ce qu’on a pu observer ponctuellement sur la liste de discussion du forum, comme on l’a vu).

En revanche, le lien systématique entre insuffisance du Codex et réactivité pour le corriger est difficile à cerner. Les modifications du Codex sont de façon significative souhaitées (13 fragments, ce qui est conséquent sur une population aussi réduite), ce qui montre cependant des insuffisances mais dénote aussi une forme de réactivité (observée sur 5 fragments), et, sur ce point, la qualification de la réactivité, à coup sûr cependant *observée*, est difficile.

On peut penser que les utilisateurs de Trac sont habitués à l’usage du wiki, Trac étant aussi un wiki (spécialisé en informatique), ce qui pourrait permettre de mieux comprendre la sérénité en matière d’usage de wikis, mais c’est difficile à dire précisément. Observer des réactions face aux insuffisances serait également normal pour des « traqueurs » de défauts.

Enfin, l’usage du Codex est orienté : il s’agit surtout de régulation des comportements, souvent des pages « de base » sur la façon de rapporter les bugs, ce qui est attendu pour la liste « trac ». Nous n’avons ainsi pas observé de « surprises » inattendues en matière de contenu cité.

[876] Source : http://comox.textdrive.com/pipermail/wp-trac/2007-October/017676.html .

8.4.2.2. Usages et efficacité sur la liste des testeurs

De même que pour Trac, on note des réactions à la qualité ponctuellement imparfaite du Codex : seuls 10 fragments sur 143 dénotent des insuffisances à des degrés divers. Mais les fragments codés avec une connotation positive dominent largement (28 fragments). On observe tout de même au début de la liste un passage de critique générale:

> "I would like to see the Documentation on the WordPress Codex be improved greatly. I always find myself wandering all over that site only to not find what I am looking for, or finding a small snippit of what I need someplace somewhat related. Latest example was the do_action() method. I wasn't aware that it had to go into some of the theme files. I had a plugin that was suppoed to be clearing out data in the DB when the footer was displayed, yet because the theme I was using didn't use do_action(), it wasn't executing. I found out about do_action() by looking at the default theme. Didn't find the do_action() method at all in the Template Tags page where I feel it should go"[877].

On observe une différence par rapport aux cas précédent : le Codex sert davantage à la régulation des tâches (12 fragments codés), en particulier les opérations de recherches de « bugs ». C'est une spécificité de cette partie de la communauté. Il en est de même de la répartition des rôles (8 fragments codés). Donc, la dimension management de projet du wiki semble plus présente dans les pratiques des informaticiens qui testent le code.

23 fragments dénotent des insuffisances ou des réactions face aux insuffisances du Codex. Cela dénote une forme significative de réactivité chez les testeurs. Cependant, un passage montre que les améliorations ne vont pas de soi, et nécessitent une intermédiation, en particulier par les administrateurs du Codex, qui interviennent comme ici:

> "From: "Sean Hayford O'Leary" <hayfordoleary at gmail.com>
> > Hmm...
> > The Codex needs some updating.
> > Is it fair to simply move the 1.6 page to 2.0 and change the text to be
> > clear that it's complete? I'm not sure if all of the listed feature (
> > http://codex.wordpress.org/Version_1.6) made it in, and if so, which did.
> You can use this as an example: http://codex.wordpress.org/Version_1.5.2
> Also User:Darkfate created http://codex.wordpress.org/Changelog/2.0 so that
> might help also.
> Thanks!
> ~~~~MichaelH »[878]

15 fragments sont des demandes de modifications, 8 dénotent soit une modification effective, soit une promesse de modification. Les conclusions sont d'ordre semblable à la précédente

[877] Source: http://comox.textdrive.com/pipermail/wp-testers/2005-March/000008.html .
[878] Source: http://comox.textdrive.com/pipermail/wp-testers/2006-January/001759.html .

liste étudiée. Cependant, ce qui n'est pas observé dans la liste précédente, il est ici évoqué que les testeurs sont porteurs d'une expertise spécifique, qui se doit d'être explicitée par inclusion dans le Codex, d'après un administrateur de la doc, qui demande des contributions à plusieurs articles:

> "Based on the meetup http://codex.wordpress.org/IRC_Meetups/2006/January/January25RawLog I've create the following document: http://codex.wordpress.org/Testing. That article is to serve as an index to the various testing reponsibilities outlined by
> http://codex.wordpress.org/User:Matt/WordPress_Responsibilities
> I also created the first Testing page, http://codex.wordpress.org/Testing/XML-RPC_and_Atom_API
> If someone would like to edit that document (Testing/XML-RPC and ATOM API) so as to make it a sample for the other Testing documents I will take that 'template' and apply it to the other pages as I create them.
> Also, if you have any better description for any of the sections listed in http://codex.wordpress.org/Testing please feel free to revise them.
> Thanks for any assistance.
> Michael E. Hancock »[879]

Enfin, en matière d'efficience, le Codex est un lieu d'apprentissage marqué, ce qui est la conséquence de cet usage : l'essentiel des fragments codés avec transmission d'une information fiable correspondent à cet usage régulier. Par exemple ce passage qui correspond à un savoir "de base" en matière de conception de logiciel (et qui aurait pu être appris ailleurs):

> "On 7/14/05, Anthony Timberlake <anthonytimberlake at gmail.com> wrote:
> > I am interested in using the beta 1.6 of Wordpress, but I am not familiar of
> > how to use CVS (I have never heard of it). Can anyone instruct me on the
> > steps?
> Be prepared for things to break here and there though :-) It isn't
> even really at alpha stage yet and is still in development, however
> you can follow the instructions below to check out the latest code:
> http://codex.wordpress.org/Using_Subversion
> Cheers,
> Jason Bainbridge »[880]

Contrairement à l'autre liste, des problèmes liés aux usages et plus seulement à la qualité de l'information apparaissent. Certain codes montrent qu'ici encore le Codex est ignoré par une frange non négligeable d'utilisateurs (paresseux ?), par exemple : « However, most still won't

[879] Source : http://comox.textdrive.com/pipermail/wp-testers/2006-January/001986.html .
[880] Source : http://comox.textdrive.com/pipermail/wp-testers/2005-July/000314.html .

know, or might not look in the Codex for such things… »[881] (deux fragments sont codés « Codex ignoré par les utilisateurs »), ce qui montre, de même que les insuffisances ponctuelles du Codex, les limites de cet apprentissage. Sur ce point, le constat est le même que pour le forum. On n'observe pas, contrairement au forum, de problèmes *structurels* d'appropriation, mais seulement quelques amateurs en matière de wiki qui ont eu ponctuellement du mal avec l'outil :

> "Ugh, being that I'm new to the codex/wiki thingy, it took me semi-forever haha... to make up a page, well I basically went with Matt's 2.0_Plugin_Compatibility page route for the layout of mine lol...hahaha.
>
> http://codex.wordpress.org/User:Spencerp/2.1alpha1_Plugin_Compatibility
>
> I have it done now, but, maybe we could keep it up and just keep changing it when need be then, or something? =/ Thanks for your reply David... =) »[882]

8.4.2.3. Usages et efficacité sur la liste des hackers

Le contexte est d'importance pour analyser et interpréter la liste de discussions. Nous commentons ici en particulier un courriel pertinent au tout début du projet de réorganisation de la documentation des informaticiens de janvier 2007. Il permet d'illustrer le fait, à maintes reprises observé, que l'organisation de l'interaction entre documentation et hackers se fasse prioritairement chez les hackers, contrairement aux autres listes où la cela ne s'observe pas, d'où a priori une plus grande facilité des mentions des mécontentements. Nous donnons un exemple d'un tel point d'articulation, à savoir un courriel envoyé aux deux listes et destiné à coordonner la documentation des programmateurs :

> « I am cross-posting this to the Hackers and Docs lists, because I think it may require input from both communities.
>
> Being a programmer who loves clear documentation, I've been looking over the about-to-be-released code for WordPress 2.1, and I have noticed that several things have changed that a plugin or theme developer would want to know (new filters, some functions deprecated, etc.). I have also noticed that the Codex could probably use a bit of reorganization (it looks like several people have discussed this recently in the Docs list archives; I just joined the list).
>
> So what I am wondering is:
>
> a) Are there plans to update the developer section of the Codex for 2.1 changes? I didn't see any in the Docs list archives for Dec/Jan, and I just joined the list, so I am not sure.

[881] Source : http://comox.textdrive.com/pipermail/wp-testers/2007-January/003656.html ; ou aussi : http://comox.textdrive.com/pipermail/wp-testers/2005-December/001588.html .
[882] Source : http://comox.textdrive.com/pipermail/wp-testers/2006-March/002387.html .

b) I am also wondering if it might make sense to start a whole new developer area in the Wiki for 2.1, and in the process, reorganize the existing developer documentation a bit.

c) Once 2.1 is out, will the Wiki need to keep around all the information for 2.0.x as well? For how long? And do we need some documentation on the changes (such as what a developer needs to do to ensure plugin and theme migration, avoid deprecated functions, etc.)?

Thoughts? And yes, I am offering to help, and maybe even to organize projects around this...

--Jennifer"[883]

Cette idée de concentration des discussions se retrouve dans les données quantifiées ci-dessous : la différenciation entre nombre de fragments et nombre de fils de discussions où se trouvent des citations montre que le wiki est l'objet ponctuel de concentrations des discussions sur cette liste (les pics), au-delà des formes de corrélations observées; certains mois, le nombre de fils où il est question du Codex augmente, et le nombre de courriels concernés augmente aussi de façon significative :

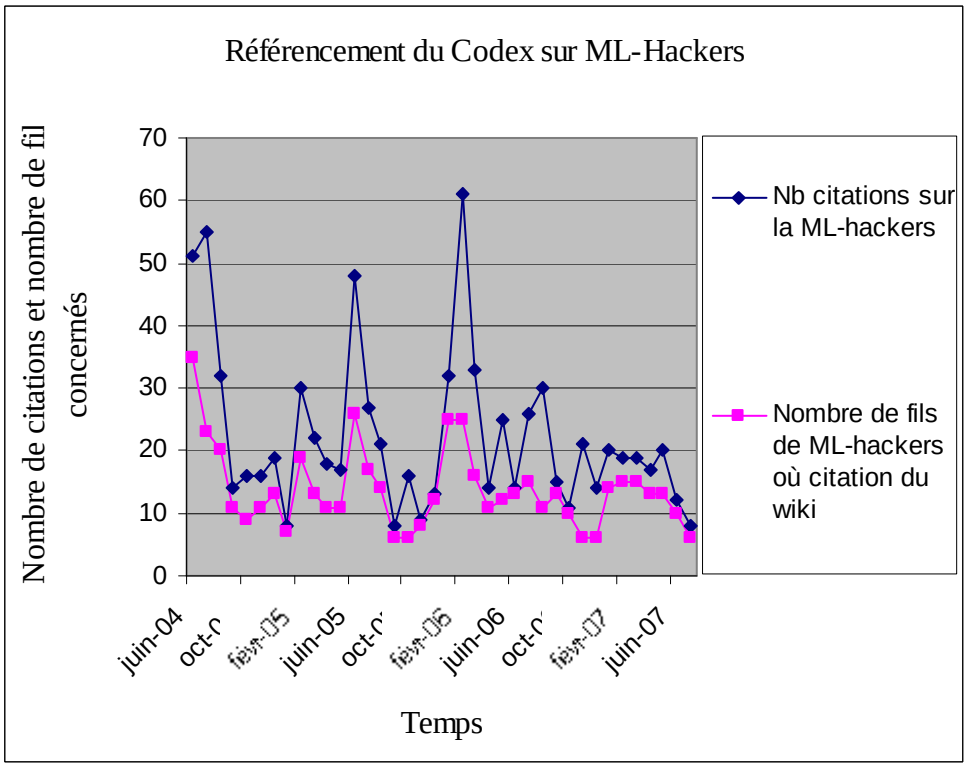


Figure 101 : nombre de fils de discussions et nombre de citations concernées par les mots clefs se référant à la documentation sur « wp-hackers »

[883] Source : http://comox.textdrive.com/pipermail/wp-hackers/2007-January/010362.html . On pourrait aisément trouver d'autres références comme preuve que la question du Codex est davantage gérée sur cette liste : http://comox.textdrive.com/pipermail/wp-hackers/2006-July/007409.html, http://comox.textdrive.com/pipermail/wp-hackers/2006-July/007004.html, http://comox.textdrive.com/pipermail/wp-hackers/2006-May/006240.html, http://comox.textdrive.com/pipermail/wp-hackers/2005-June/001283.html .

Ainsi, l'interaction avec la documentation est plus significative dans cette liste qui est la principale liste de régulation de la communauté, d'où un nombre conséquent de considérations sur la qualité de la documentation[884], et le caractère prioritaire de cette source.

De ce fait, les éléments de mécontentement vis-à-vis du Codex pourraient apparaître de façon structurellement plus marquée dans cette liste de discussion que dans les deux précédentes. Au moins 15 % des fragments sont connotés négativement en ce qui concerne la qualité de l'information, contre 7 % positivement. Les proportions s'inversent par rapport aux deux précédentes. Les proportions sont les mêmes pour le Codex en général. Contrairement aux autres listes, les contrastes peuvent être marqués sur cette liste, entre des fragments explicitement élogieux[885] et des commentaires de répulsion assumés.

D'où viendrait la qualité dans une certaine mesure discutable du Codex d'après les fragments ? Cette liste est plus pertinente que les autres pour obtenir ce type de précisions. Voici un passage qui montre que même si le contenu est satisfaisant, la structure du SI pose parfois problème :

> "(This by the way, is the reaon I hate the wiki) - I searched the wiki, and looked through the main WP site - and couldn't find it. I KNOW it's there - because I saw the page when I was working on my "today in weeks/years past" plugin - but now I can't find it again."[886]

Cependant, il apparaît ici plus clairement que la qualité du Codex est moins un problème de formatage de contenu que de *qualité du contenu lui-même*. Contrairement aux conclusions des première parties de cette thèse, que nous aurions préféré voir confirmées ici, nous devons nuancer pour les programmateurs: il apparaît que les insuffisances relevées ne sont pas dues principalement à un problème d'organisation problématique de contenu (5 fragments codés ainsi[887]), mais à un problème de qualité de contenu lui-même (l'ensemble des autres fragments codés négativement). Ainsi, pour les hackers, plus experts, la qualité du contenu de la documentation se poserait de façon plus significative. Il est aussi fort probable que les hackers connaissent mieux les wikis que les novices, ce qui pourrait faire aussi que sa structuration pose moins problème.

Enfin, on remarque aussi 31 fragments avec d'autres formes de documentation suggérées, ce qui signifie que le Codex n'est pas la seule source d'information, et pas forcément la meilleure.

---

[884] Là encore, on pourrait citer de nombreux passages, par exemple : http://comox.textdrive.com/pipermail/wp-hackers/2006-November/009516.html .
[885] Source : http://comox.textdrive.com/pipermail/wp-hackers/2006-July/007430.html. Ce passage montre d'ailleurs le rôle stratégique de la documentation pour multiplier les utilisateurs d'un logiciel open-source.
[886] Source: http://comox.textdrive.com/pipermail/hackers/2004-June/000351.html .
[887] Par exemple : http://comox.textdrive.com/pipermail/wp-hackers/2006-November/009527.html .

Voici à présent le seul point d'arbitrage observé en matière de qualité, à l'interface entre les deux parties de la communauté : l'efficience de la documentation dépend du degré *satisfaisant* d'intervention des « hackers ». C'est une question qui ne va pas de soi en fait, car cette intervention se devrait d'être *mesurée*, alors qu'intuitivement, le « feedback » des hackers serait systématiquement requis pour permettre la qualité de la documentation. Les hackers n'ont pas forcément envie parce que pas forcément qualifiés[888]. Nous citons deux fragments:

> "Perhaps you would like to reconsider your invitation to have me improve your documentation. I'm code-savvy, you see. That means I am completely divorced from the reality of the casual user experience and have no documentation skills whatsoever.
>
> Paul. »[889]
>
> "I think it's been proven before that coders do not make good documentors. Tons of options will make for a better UE probably, but then, having all those options my cause it to get complicated....more so than it was before."[890]

Cependant, les limites des contributions des hackers sont sujettes à débats. Ailleurs en effet, d'autres passages montrent au contraire qu'ils seraient les mieux qualifiés ; ainsi :

> "Why would someone who *didn't* write the software be expected to write the documentation for it? Who understands the software better, and is in a better position to write the docs? The developers, or someone who tinkers with WordPress? How am I supposed to write the documentation for some_function() when I didn't create the function?"[891]

En complément de ce qui précède, l'équipe de documentation elle-même formule clairement des demandes d'expertises:

> "We are down to only a few articles from the Codex Cleanup Week that are finally ready for review. These are fairly technical so they need some expert attention.
>
> http://codex.wordpress.org/UNIX_Shell_Skills
>
> http://codex.wordpress.org/User:Lorelle/Test_Driving_WordPress
>
> http://codex.wordpress.org/Hardening_WordPress
>
> We are also currently missing an article at http://codex.wordpress.org/index.php?title=Database_Reference&action=edit If you know the person writing the article or have something in mind, we need to have something of value on this page link from the Developer Documentation Table of Contents page. It's been there as a blank document for MONTHS. If we know the intent of such a document, then we can put it on the task list."[892]

---

[888] On peut aussi soupçonner qu'il s'agirait d'un prétexte pour ne rien faire.
[889] Source : http://comox.textdrive.com/pipermail/wp-hackers/2006-May/006184.html ; sur le fait qu'il s'agit de deux tâches de nature différente : http://comox.textdrive.com/pipermail/hackers/2004-June/000019.html .
[890] Source : http://comox.textdrive.com/pipermail/wp-hackers/2006-August/007784.html .
[891] Source : http://comox.textdrive.com/pipermail/wp-hackers/2006-May/006365.html .
[892] Source : http://comox.textdrive.com/pipermail/wp-hackers/2005-July/001803.html .

Cependant, d'une façon globale, *la réactivité des hackers est indiscutable* : 21 fragments montrent des modifications, 28 sont des promesses, alors que 46 sont des demandes de modifications. Ces ratios sont plus conséquents que dans les listes précédentes. On ne peut pas dire que les programmateurs n'interviennent pas sur le Codex, et cela est conforme à la spécificité de cette liste. Des formes de reconnaissance sont clairement observées : "Jennifer, you are doing fantastic work on the Codex!"[893]. On observe aussi des succès en matière de collaboration:

> "Alright, the article has been completely rewritten, according to the suggestions made by all.
> It uses a combination of everyone's ideas, and I think it turned out quite well. Thanks for all the help!
> http://codex.wordpress.org/User:Jalenack/Creating_Tables_with_Plugins. Now, any more suggestions/quibbles with it?"[894]

En revanche, 15 fragments ont été codés « délégation à l'équipe de documentation », ce qui montre les limites des contributions des hackers. En effet, ce passage par exemple montre que l'intervention sur le wiki n'est pas si immédiate et que les gros contributeurs ont un rôle pour cerner les tâches à effectuer :

> "What is currently done (or what is supposed to be done) in the Codex along these lines:
> * The bloginfo function doc page lists the currently preferred inputs (non-preferred are not mentioned). If someone wanted to add a section there listing the older inputs and current replacements, it might be useful. Feel free; if you are not a Codex expert but want to help, make up a list and I can add it to the doc page.
> * Generally, on the function documentation list pages (Template Tags, Function Reference, etc.), functions are marked with which WP version they were created in, and which WP version they were deprecated in. This may not be perfect (the Codex is maintained by volunteers after all), but that is at least the plan.
> * If you visit the individual doc page for a function that has been deprecated, it should say what the replacement function is and what WP version it was deprecated in.
> * There are also version-to-version upgrade guides in the Codex for plugin/theme authors that list deprecated functions and suggest authors update to use new functions.
> If you think the Codex needs additional documentation pages related to of preferred/deprecated behavior, probably the right place to discuss it is on the Doc list. Personally, I think what is listed above is probably sufficient for most purposes.
> Jennifer »[895]

Un tel arbitrage entre contributions des uns et des autres montre donc que l'expertise entre les deux activités est *partagée* entre les deux sous-communautés. Trouver les projets en cours suppose l'intermédiation d'administrateurs, comme ici, avec la citation d'un début d'article

[893] Source : http://comox.textdrive.com/pipermail/wp-hackers/2007-February/010855.html .
[894] Source : http://comox.textdrive.com/pipermail/wp-hackers/2005-June/001482.html .
[895] Source : http://comox.textdrive.com/pipermail/wp-hackers/2007-May/012377.html .

qui ne se trouve pas sur le « main space », mais sur un « user:space » (moins visible sur le wiki) :

> « Here's a start
> http://codex.wordpress.org/User:DavidHouse/Wordpress_Code_Flow
> MichaelH »[896]

Certains contributeurs notables au Codex, comme « MichaelH » justement, orientent dans quasiment toutes leurs interventions les hackers vers les bonnes pages. Ces considérations en matière de délégation confirment d'ailleurs la tendance à la *légitimation* de l'expertise des membres de la documentation de la communauté, un point mentionné antérieurement au chapitre 5, et sur lequel nous n'avions alors pu nous prononcer ; ce chapitre renforce l'argumentation dans le sens de la légitimation.

Enfin, nous mentionnons un autre processus sous-jacent à l'usage du Codex, le Codex est un lieu d'apprentissage, ce qui montre la difficulté de son utilisation lorsque le contenu est plus technique, qui ne serait pas « immédiate » (réponses trouvées immédiatement, ce qui correspond à la différence entre « information » et « connaissance »), et permet aussi de relativiser les mécontentements :

> "Hi,
> If you wish to see a dependencies chart (or function call chart), then you'll probably need to generate one yourself, using a PHP flow analysis tool. It could, however, be rather complex. I suggest that you *learn* the functions and files, as documented in the Codex, starting with the core files and going 'outwards'. This way you will get a feel for what's going on 'under the hood' (WordPress anatomy). You will know which file/function does what. Beyond this, I am not entirely sure what to suggest, but I advise that you do intra-file search using tools like GNUrep to identify code of interest. You could alternatively search the Codex for particular bits of the code. There is some description there about the structure/hierarchy, which is probably the most you can get.
> As code evolves over time, there's nothing that beats experience and once you stop hacking, you'll *forget* variable names, function names, and find them removed, renamed, or relocated.
> I hope this helps a little...
> "@spencerp"[897]

Ce que montre cette réponse complexe, c'est que le Codex est un lieu d'apprentissage et non de réponse immédiate par accès immédiat à l'information, ce qui est pourtant la « philosophie » des requêtes sur le Web ; utiliser un moteur de recherche et un seul article ne suffit pas toujours. Une telle démarche expliquerait pas mal d'échecs! Le Codex n'est pas

[896] Source : http://comox.textdrive.com/pipermail/wp-hackers/2006-August/008067.html .
[897] Source : http://comox.textdrive.com/pipermail/wp-hackers/2007-February/010762.html (c'est nous qui soulignons).

seulement un lieu de référence, mais un lieu d'apprentissage, un manuel qu'il faut lire voire apprendre. La différence est fondamentale avec les utilisateurs novices. Enfin, ce qui en est le corrélat, on trouve davantage de fragments avec une connotation expert qu'une connotation débutant, malgré la présence de 8 fragments codés « RTFM » (« Read the f[censuré] manual ») . Si le codex est à la fois le réceptacle d'un savoir plutôt savant, mais la connotation « manuel de base » qu'on omet de consulter, même parmi les experts, persiste.

Comme dans la liste de discussion « Trac », les fragments codés montre que le wiki sert davantage pour la régulation des comportements que pour la régulation des tâches.

8.5. Comparaisons, discussion des résultats, conclusions, limites et autres orientations de travail

Cette sous-question de recherche sur les usages du wiki dans une communauté rappelle un point crucial : l'apport informationnel d'un wiki dépend du contexte institutionnel, ici les exigences d'un projet open-source. Les utilisateurs sont exigeants (contrairement à Wikipedia, où ce serait plutôt l'inverse, l'indulgence en cas d'insuffisances), et ont des attentes particulières… auxquelles le wiki ne répond que partiellement, et pas seulement pour des questions de contenu : l'accès à l'information ne va pas de soi.

Voici les mécanismes imprévus observés. Il s'agit ici de notre apport au courant émergent de « l'économie du knowledge management », qui étudie des mécanismes et des arbitrages. Il permet de comprendre que les usages du wiki à l'interface entre les différentes parties de la communauté ne sont pas si immédiats que cela :

- Pour le forum, plusieurs *mécanismes* limitatifs sont en jeu:

  - un phénomène de complémentarité « démographique » qui se traduit, au-delà des transferts d'informations et partages de connaissances, par des arbitrages de la population de la communauté entre aide sur le forum ou le wiki parmi les membres du réseau de volontaires en grande partie commun ;

  - la nécessité de gérer la complémentarité entre deux formes de savoirs différents sur le wiki et le forum ;

  - un effet contraire à la valorisation de l'expertise qui peut dissuader d'utiliser le wiki ;

  - la nécessité de réguler la façon de faire usage du Codex, qui ne doit pas être anarchique, systématique, ou irréfléchie.

  - un problème d'incitation à la lecture et à la recherche dans le wiki.

- Pour les informaticiens, où on observe une *permanence structurelle* de l'usage du Codex, on aurait donc comme processus d'appropriation :

- un *degré satisfaisant* de « feedback » des informaticiens et de délégation à l'équipe de documentation ;
- un mécanisme *d'intermédiation* (significatif mais non systématique) i.e. indiquer où se trouve la documentation correcte. L'efficience de la documentation suppose des formes notables d'intermédiation, comme dans le cas du forum.
- un mécanisme nécessaire *d'apprentissage* à cause de l'orientation du Codex vers un niveau significativement expert, d'où davantage de frustrations pour ceux qui sont novices.

Ce chapitre comporte aussi un apport méthodologique en tant que tentative d'analyse des traces dans les listes de discussions. Composer un tableau général de l'appropriation du produit de la collaboration massivement distribuée dans plusieurs contextes d'usages différents (le forum, où l'essentiel des lecteurs sont des amateurs paresseux, la programmation, où les experts sont nombreux) n'est pas trivial, et affirmer clairement que les pratiques sont complexes à partir des seules traces disponibles est un réflexe salutaire d'historien, et montre l'utilité de l'approche historique pour analyser les usages et l'efficience. Nous avons montré en la matière l'utilité de l'articulation quali/quanti, de la prise en compte de la spécificité de chaque archive avant toute comparaison en particulier par des éléments de critique des sources. Nous avons aussi expérimenté les limites du codage, qui sert surtout à classer.

Un autre apport est que par rapport à la maigre littérature antérieure, nous pouvons affirmer que dans notre cas, les hypothèses de HANSEN et al. (2007) n'ont pas été validées, i.e. le wiki sert à rester « dans le sujet », et le wiki sert à éviter les conflits. Même si nous n'avons pas testé cela en priorité avec des codes, cela n'apparaît pas d'après notre connaissance des archives : ce que nous avons pu observer n'a vraiment *rien* à voir. Il n'est d'ailleurs pas étonnant qu'on obtienne un contraste marqué, au vu du caractère réduit du nombre de cas sur la question.

Notre présent travail comporte également des *limites* et suggère d'autres formes de vérifications mais aussi d'autres questionnements

En matière d'efficience de la recherche, il n'est pas sûr que notre méthode de travail, pointilleuse, soit la plus rentable. La procédure que nous avons mise en place par tâtonnement

était particulièrement lourde à la main, qu'il s'agisse du dépouillement des données ou des copiés-collés nécessaires pour tracer les données textuelles. Il y aurait des progrès à effectuer en la matière, mais c'était, au moins dans une certaine mesure, un « sunk cost » nécessaire.

Pour des raisons de déroulement de la thèse, le dépouillement des sources s'est arrêté à mi-juillet 2007 ; contrairement à ce que nous pensions, nous n'avons pas eu le temps d'actualiser jusqu'à décembre 2008 ; pour la liste du forum, cela ne porte guère à conséquence, étant donné que son débit a alors baissé, mais sans doute davantage pour les autres listes. Un dépouillement supplémentaire pourrait apporter des éléments sur les dynamiques observées ou des nuances supplémentaires, avec la même méthodologie. Ce serait un travail de quelques semaines à effectuer par la suite.

Les résultats concernant les usages par les contributeurs au forum mériteraient d'être testés sur les fils de discussion se trouvant sur le forum lui-même : les routines d'utilisation du wiki ont-elles été de fait respectées (pour ne parler que de cela) ? Les résultats concernant les listes de programmation, qui permettent par ailleurs des triangulations des sources, sont en revanche plus solides d'un degré : nous ne voyons pas quels types de tests supplémentaires pourraient être menés sur ce point.

Notre recherche a montré que le mot-clef "doc" est utile, il montre que la documentation du wiki n'est pas la seule à être utilisée; on trouve d'autres manuels, comme le manuel de PhP, ou Wikipedia. A propos de l'usage par les blogueurs d'autres systèmes d'informations de type wiki ou autres ailleurs que dans la communauté, des travaux de recherche supplémentaires sont nécessaires, nous ne pouvons les mentionner ici. Le phénomène apparaît ponctuellement, ce qui est suffisant pour montrer au moins qu'il existe à coup sûr. C'est un aspect pertinent de *l'efficacité* du système d'information Codex par rapport à *d'autres*, que nous n'avons pu traiter. Par ailleurs, peut-on caractériser la spécificité de l'équipe de documentation (et non plus seulement sa qualité) à travers le regard des autres parties de la communauté[898] ? Il est difficile de savoir cependant s'il existe des données suffisantes pour tirer des conclusions définitives solides sur ce sujet. On pourrait sur ce point reprendre tous les fragments et les étudier selon d'autres optiques que celles de l'appropriation et de l'efficacité : plusieurs codages sont toujours possibles.

[898]On parle ainsi de « codex people » (source : http://comox.textdrive.com/pipermail/wp-testers/2007-January/003790.html), ou ailleurs, on parle d'implémentation de la doc, voir de « hack » (source : http://comox.textdrive.com/pipermail/wp-testers/2005-December/001582.html).

8.6. Annexes du chapitre 8

8.6.1. Annexe 8-1 : visualisation du processus de codage

Nous avons travaillé sur Excel pour pouvoir contrôler totalement les procédures de traitement des données. Voici une capture d'écran de ce qui se passe « en cuisine » :

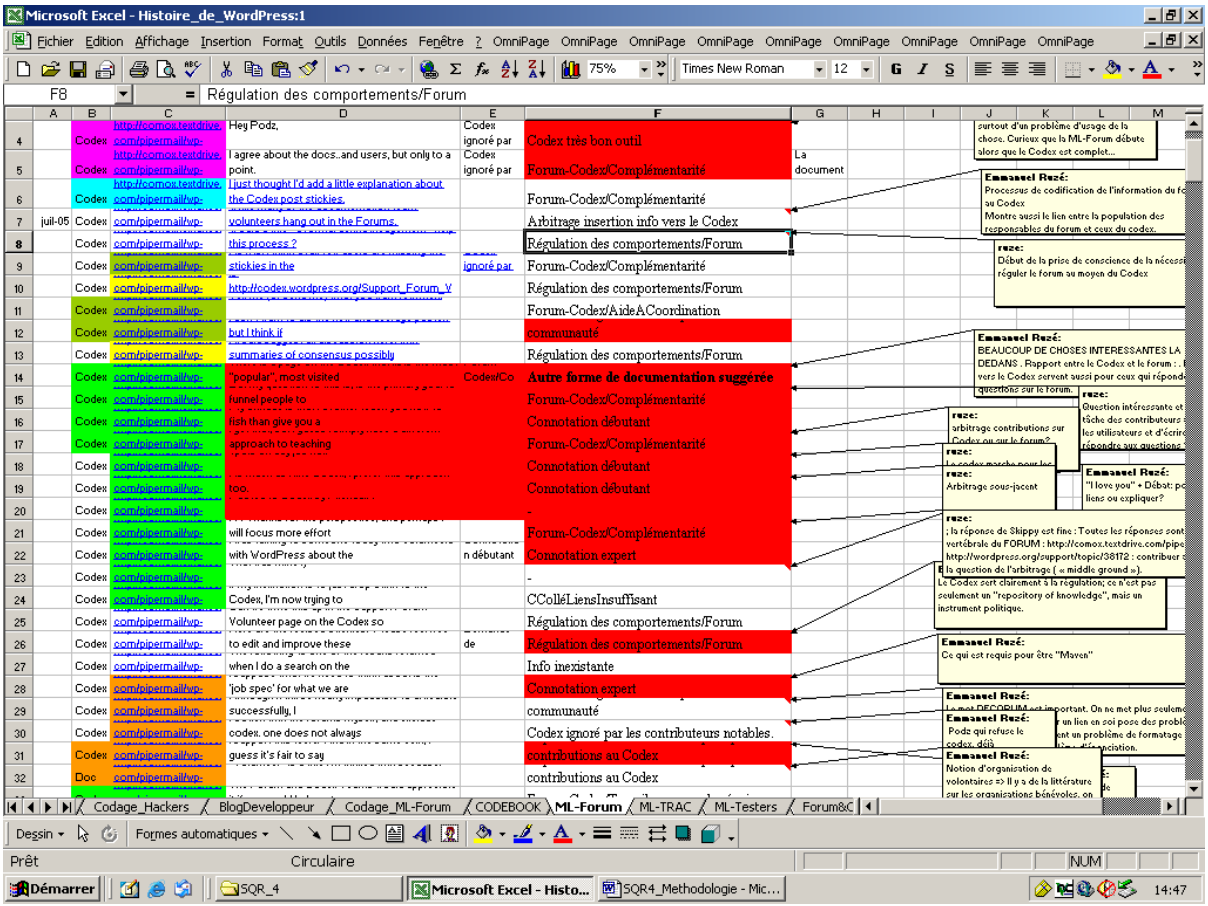


Figure 102 : capture d'écran du fichier Excel où fut effectué le codage des données

Notre extraction de données et notre codage permettent de tracer exactement la procédure d'analyse des données. A gauche, se trouve le mot-clef qui a permis l'extraction d'un fragment, ensuite, en bleu, l'URL permettant de retrouver les données, ensuite, le passage pertinent du texte copié-collé, et enfin les codes associés aux fragments. Les bulles permettent des commentaires et l'identification de textes pertinents, tout aussi utiles, si ce n'est plus, que les codes. Les couleurs sur les cases des URL permettent d'identifier les fils où plusieurs mails comportent le mot-clef. Les cases en rouge indiquent que les données en question ont intuitivement plus d'importance, voire comportent des « incidents critiques ». Les calculs associés aux décomptes des codes se font à partir des fonctions élémentaires

d’Excel, et il est aisé de retrouver les fragments associés à chaque code au moyen des fenêtres de recherche de texte.

8.6.2. Annexe 8-2 : courriels longs importants

Cette annexe correspond à des citations trop longues pour être insérées dans les textes telles qu’elles. Elles correspondent à des courriels particulièrement éloquents évoquant des insuffisances du Codex et donnent une idée de l’intensité des débats.

Courriel 1:

“Caution: inflammatory post follows.
Podz wrote:
>There are many hosts with many setups and as such WordPress documentation cannot cater for all - so it >approaches this by catering for none. It does this for the above reason and also for another - so WP cannot be >blamed. The fault is pushed to the user. Yet no page at wordpress.org carries any advice on permissions. The >readme.html with the package carries no advice on permissions. Codex carries no specific advice on >permissions.
I place the blame for this squarely on Matt's shoulders.
[…]
Not only do I disagree with the requirement for a writable /wp-content/, I disagree strongly with Matt's casual -- almost dimissive – attitude regarding his involvement with the documentation.
I should point out, too, that I saw no effort from Matt to modify the official documentation for his project. He merely criticized it in a trac comment, and left the documentation unmodified. That's leadership for you.
I understand that writing code is more fun than writing docs. But I feel that Matt, Ryan, and any other core developer has an _obligation_ to provide proactive assistance with the finer points of documentation. They know the code better than anyone else, and are _the_ authoritative body for questions or ambiguities. That they choose to adbicate their responsibility explains a lot of the frustrations that have been percolating throughout the community.
Matt's living it up, going to conferences and glad-handing people, while the rest of us are left to struggle with confusing undocumented code, and to try to distill meaning from the madness for the other users who just want to use the damned thing.
> But sites either do not work because of permissions or are insecure because of permissions. So they look to > codex, they look to official pages. And they find nothing because no-one will commit to writing anything > because they fear getting the blame.

It's not just a matter of blame, Podz. It's a matter of competing interests. Every time we try to write thorough documentation that addresses as many conflicting configurations as possible, we get slammed for making "confusing" documentation.

But Matt can sweep in and apply a new theme to the code which _breaks_ a lot of functionality, and renders complex pages completely useless, without so much as an "excuse me, please".

Rather than fix the codex (one mechanism of which would be to install a fresh MediaWiki installation into a new directory, and then manually copy-and-paste pages from old to new -- yes, we'd lose history, but we'd keep the damn docs usable), Matt wants to write a new plugin for WordPress to "solve" the problem. This is perfectly acceptable from Matt's point of view, even though the documentation continues to suffer in the meantime.

But when others come along and volunteer to "solve" the problem of inline function references, Matt poo-poos the whole effort as a waste of time, saying instead that such documentation belongs on a publicly editable wiki. Forgive me for not leaping to participate when the current wiki situation is abominable.

> But they ask in the forums don't they ?

> They expect an answer from those of us there don't they ?

> So devs and people who know all about perms and suchlike are content to let forum helpers get it in the neck

> when things go wrong because they won't write anything.

> Nice.

The devs don't answer because they don't care. The devs care about the fun stuff of writing code. The care about making a revenue stream from all the ancillary services they've developed around WordPress: pingomatic, akismet, blicki, WordPress.com, and whatever other pokers Matt has in the fire.

My recent withdrawl from all things WordPress has helped put some things in perspective. Matt and crew are poor project leaders. Matt's egomaniacal Automattic stuff is hurting the WordPress community. The code continues to grow hodge-podge without a clearly defined vision or roadmap being presented to would-be contributors. Contributions that don't satisfy Matt's undocumented criteria are simply ignored. Trac tickets are closed with terse "wontfix" messages, rather than useful explanations as to why it wont be fixed.

I would very much like to see WordPress thrive and succeed. I would like for the autocratic development model to be relaxed. I would like to read, and discuss, a plan for long-term development and vision. I would like to see specific product release projections so that testing, documentation, and plugins can all be readied _prior_ to a release. I would like to see contributions of all sorts being _encouraged_ rather than dismissed because they don't coincide with one person's particular preferences. I would like to see infrastructure and site issues be dealt with in an open, responsive manner.

Basically, it Matt gets hit by the bus, much of WordPress's successs is screwed. It doesn't need to be that way."[899]

Courriel 2:

"- the elusive search feature: PLEASE make it default to searching the forums instead of the codex!? at least until everyone's confident in the codex's ability to address the specific issue. the results that come up are

[899] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-March/001532.html .

confusing, as the title links often don't even include the words used in the search - so people think their time is wasted, come to the forum, say "I've searched but can't find anything..." and then people get frustrated. When you do the search for something, and then re-search for the forums, you get the best results, period.

- Someone said something a while back about possibly thinking of taking on a "pro" to sort through the codex and make it work. I'm for it - at least to get it started in the right direction. And before you go saying, "Sure, Lara - go ahead and spend someone else's money!" - I'm thinking again as a marketing person. It takes money to make things work. In my first two years of business I reported a loss on my taxes because I put every penny and then some back into the business. That's just how successful companies go. Make it right the first time, and you won't have to do it again - and you'll have happy people to keep coming back. I don't know what plans there are to ever make money off of WP, but if there is any to be made, it'll come back to you three-fold if you do it right, now. Just think: You buy a new food processor, and the manual says "Just put in the blade, add the food, and push the button." - You wind up putting the blade in upside down, forget to put on the cover, toss in a carrot and push the first button you see... the blade pops out and cuts your hand, and you sue the company. You also blog about it, and tell everyone how much that brand of food processor sucks the bag. Bad press is the worst of WP's worries, but it's out there now. Let's nip it in the bud and avoid the downfall.

Okay - I'll stop there for now. ;)

~ Lara[900] »

[900] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-August/003141.html .

Conclusion générale

Les apports de notre thèse pour les gestionnaires des communautés et la recherche en gestion permettent de mieux comprendre *comment gérer le savoir en ligne*, par le déploiement d'un système d'information. L'usage d'un outil wiki est un cas particulier de ces pratiques, mais un tel système d'information comporte des spécificités, notamment notre thèse initiale sur l'émergence de la gouvernance.

Les apports factuels proposés, i.e. la description et compréhension de processus, sont des apports pour les gestionnaires de communauté et pas simplement un dépassement des limites soulignées par la revue de littérature scientifique. Ils permettent de cerner de façon approfondie par la compréhension ou l'explication les mécanismes pertinents et spécifiques pour la gestion des SI sur un support wiki dans le contexte des communautés open-source. Ils fournissent des repères aux « managers » des communautés ou des projets open-source. Ils disent à quoi s'attendre (régularités, problèmes de gestion, solutions émergentes) et pourquoi, mais aussi ce vers quoi tendre leur action ; les solutions adoptées par la communauté WordPress ne sont en effet pas sans intérêt.

Les apports de notre thèse sont de plusieurs ordres : elle propose des résultats factuels, des recommandations managériales, des contributions méthodologiques, ainsi que des pistes de recherches supplémentaires plus ou moins élaborées pour notre cas mais aussi au-delà.

Apports descriptifs, compréhensifs et explicatifs

Nous avons proposé une étude de cas unique et approfondie d'une communauté en ligne qui gère son savoir sur un support wiki. Nous souhaitons mettre en perspective notre travail, en particulier souligner la mesure des limites de l'idéal-type de la « collaboration massivement distribuée » et l'émergence de formes de gouvernance au-delà de cette première forme d'auto-organisation.

La gouvernance est tout d'abord gestion de l'interdépendance entre les organes d'une communauté. Nous avons pu comprendre les modalités dynamiques d'insertion d'un système d'information reposant sur un wiki dans une communauté open-source. Il s'agit d'une trajectoire de contributions et d'organisation aboutissant à une centralisation effective, malgré plusieurs tentatives de remises en question. Cette insertion est un processus dont nous avons souligné les différentiels et les convergences de rythmes. Nous avons également proposé une

mise en intrigue (VEYNE, 1971) de l'histoire de l'équipe de documentation qui permet d'en comprendre les rebondissements événementiels, en particulier par la récurrence de phases d'organisation et par la synchronisation avec la sortie des versions successives du logiciel.

Nous avons montré qu'une telle communauté pouvait résoudre de façon satisfaisante la problématique de formation de bien collectif (« public good ») informationnel alors qu'un tel projet était ardu et contre-intuitif d'un point de vue économique. Nous avons documenté les motivations variées supportant ce processus et les structures des contributions des volontaires qui se caractérisent par un noyau permanent de contributeurs. Cette sous-question sur les motivations et la structure des contributions comporte des limites probables que nous verrons en détail à la fin de cette conclusion ; mais elles ne vont aucunement à l'encontre de notre thèse sur l'émergence de la gouvernance du système d'information.

Nous avons ensuite observé et compris les institutions informelles sous-jacentes au précédent processus que sont les routines. Elles subissent des modifications dans un contexte d'exigence d'efficacité comme celui d'une communauté open-source : cette évolution montre les limites du modèle « pur » de la collaboration massivement distribuée, dont le modèle, Wikipedia, a été amendé sur plusieurs points que nous avons documentés. Les raisons de ce processus d'adaptation des routines ont été comprises par l'analyse de la sélection et l'adaptation des heuristiques par réflexion économique coût/avantage et la mise en avant de principes de justification.

Nous avons expliqué la dynamique sous-jacente à la coordination du processus d'éditions qui a conduit à des interactions centralisées sur une liste de discussions. La régulation d'un tel processus comporte des phases d'activité particulièrement conséquentes se traduisant par la dispersion et la dissymétrie. Elle se caractérise essentiellement par un noyau dur d'administrateurs aux comportements spécifiques et par des phénomènes d'avalanche, de structuration par homophilie et de réseau scale-free avec des formes marquées de réciprocité et de reconnaissance réciproque de l'expertise que suppose le projet de documentation. Tels sont les éléments permettant de comprendre la viabilité de ce réseau. *Il est douteux que des interactions décentralisées sur un wiki eussent pu à elles seules supporter un tel processus de régulation, même si elles ont existé.*

Nous avons montré les limites de l'appropriation et l'usage « spontané » du wiki par la communauté où il s'insère. L'efficacité des usages spontanés du wiki concerne davantage les hackers que les utilisateurs novices, où l'usage problématique du système suppose une forme de régulation informelle et de complémentarité entre SI. Il est globalement impossible de

soutenir que le projet de système d'information ait été un échec pour ces deux catégories d'usagers.

Apport managériaux pour les contributeurs aux communautés, ceux qui sont amenés à les gérer, voire au-delà[901]

Si nous n'avions observé que des formes d'auto-organisation « anarchiques » sans émergence de formes de gouvernance, nous n'aurions sans doute rien à dire ici et aucune recommandation à adresser à des gestionnaires de communauté. Or, a contrario, nous pensons pouvoir dans une certaine mesure en fournir ci-dessous.

Les apports managériaux peuvent, une fois n'est pas coutume, concerner toute personne qui aurait un rôle significatif dans une communauté ayant implémenté un wiki. En effet, ces personnes peuvent être amenées par la suite à prendre des responsabilités de façon informelle et de façon consensuelle avec d'autres qui feraient de même. Nous ne nous adressons donc pas qu'aux chefs de projets. C'est donc en ce sens large que nous parlerons de « gestionnaires » ou de « managers » ci-dessous.

Il n'est pas exclu par ailleurs que ces recommandations aient une portée au-delà des communautés open-source dans des organisations aux caractéristiques plus formelles, mais désirant mettre en place des communautés internes :

> "Frequent changes witnessed presently add another layer of difficulty with almost constant knowledge obsolescence and knowledge update requirements. Corporate knowledge management initiatives are an essential way to overcome these problems, sometimes as a complement to training, other times even as a substitute. *But corporate knowledge management has a great deal to learn from the experiences of online communities, which increasingly organize knowledge in wiki and blog environments, and utilize tags, feeds, aggregators and links to better connect the pieces.*" (SCHNEIDER, 2005. C'est nous qui soulignons.)

Cependant, nous ne nous avancerons pas, ici du moins, sur ce terrain, par prudence.

Un gestionnaire de communauté sera d'abord préoccupé par la trajectoire de légitimation d'un SI et la dimension temporelle de ce déploiement comme outil central de gestion du savoir dans un contexte communautaire contraignant. *Notre recommandation* aux « managers » de communauté est de mettre en place la documentation très tôt dans le projet

[901] Pour une vision concrète et immédiate de l'utilité commerciale de ces questions par des consultants, voir par exemple http://www.wikipatterns.com/ (consulté le 25/09/09). Le lecteur pourra comparer les solutions que nous proposons et celles qui y apparaissent. Même Cap Gemini s'intéresse aux outils « Web 2.0. ». Nous n'avons pas effectué de « benchmark » des recommandations gestionnaires des organisations à but lucratif de ce genre, pour éviter de perturber notre travail inductif par des a priori. Il concerne d'abord les communautés open-source, qui ne sont pas forcément la cible des consultants mais ne sont pas moins pertinentes.

de conception communautaire de logiciel, au moins en même temps que les premières versions du logiciel, et de mettre en place une régulation de cette documentation de la même façon par rapport aux flux (linéaire ?) des éditions et de création d'articles. Notre cas montre que la chronologie et la temporalité du déploiement du SI est décisive[902]. Sur un plan plus microanalytique, la narration que nous avons construite montre l'importance de l'enchaînement des phases de contributions et de réorganisation davantage qu'un poids de la conflictualité[903]. Enfin, « wiki wiki » signifie à l'origine en hawaïen « vite vite »… alors que les rythmes observés peuvent être effectivement dynamiques et oscillatoires, mais aussi *lents*. Ce contraste pourrait être à l'inverse accepté sans crainte ou interventions inutiles par un gestionnaire de communauté.

Un gestionnaire de communauté sera préoccupé par l'émergence d'un nombre satisfaisant de contributions informationnelles et par le processus qui y mène. *Notre recommandation* « hayékienne » aux managers est de veiller au maintien de ce processus délicat sans cependant intervenir par une politique d'incitations, mais en veillant en amont à ce que le problème de problème de bien public de second ordre soit résolu pour qu'il ne gêne pas l'émergence des contributions. En ce qui concerne le caractère émergent « spontané » des motivations à la contribution, il faudrait faire le contraire de nos autres recommandations de cette conclusion : il ne faut sans doute pas y « toucher » par une gouvernance mal venue ou par une logique marchande trop précoce. Ce point est délicat et amène celui qui suit.

Un gestionnaire de communauté sera préoccupé par l'émergence de « bonnes pratiques » en matière d'organisation du système d'information. *Notre recommandation* aux managers est de veiller à un équilibre satisfaisant mais incontournable entre routines issues de l'open-source et de l'open-media. Cet équilibre penche ici curieusement vers le premier type d'organisation en particulier en ce qui touche la structuration du contenu (contrôle de l'ajout de contenu, de l'architecture, limitation du principe « edit mercilessly », etc.) mais moins en ce qui touche la gouvernance de l'équipe spécialisée en matière de documentation (supervision et modes de communication hybride, etc.). Le processus de sélection des routines montre les limites significatives du modèle pur de la « collaboration massivement distribuée », et l'évolution des routines serait un processus normal dans ces communautés.

[902] Le cercle vertueux et les caractéristiques des dynamiques ont été résumés en conclusion du chapitre 4.
[903] Nous pensons utile de signaler que Matthias Den Besten, membre du jury de pré-soutenance et spécialiste de Wikipedia, a affirmé que cette narration littéraire est selon lui du passage le plus instructif de la thèse, alors qu'il est modélisateur en informatique et économie. Une narration très concrète peut être donc utile aussi aux managers.

Notre cas ne saurait permettre de dire en revanche où se situe cet équilibre pour d'autres communautés de taille différente.

Un gestionnaire de communauté sera préoccupé par la logique dynamique de la collaboration entre contributeurs, qui supporte les processus précédents. *Notre recommandation* pour permettre la viabilité de la collaboration et de la coordination de l'activité est de centraliser *au minimum partiellement* cette activité sur une liste de discussion pour des raisons d'économie cognitive. Ce processus suppose un nombre suffisant d'administrateurs dont l'autorité et la confiance distribuée se caractérisent par des interactions réciproques fortes. Ce processus ne va cependant pas jusqu'à nécessiter des flux importants comme dans le cas de la conception de logiciel et une connectivité forte de l'ensemble des intervenants. Un gestionnaire de communauté devra aussi profiter de l'effet d'entraînement qui conduit à un nombre tout à fait significatif d'interventions à synchroniser avec des phases d'organisation du projet. Un tel processus n'est cependant pas sans inconvénient d'après notre cas. Les contraintes managériales ne sont pas triviales et sur certains points, *nous n'avons pas de solution à proposer* ; en particulier, les textes étudiés montrent qu'il n'y a pas d'émergence d'un « middle management » dans le SI, et que cela aurait posé des difficultés (chapitre 6), malgré le succès de la communauté. Sur ce point précis, le mécanisme de gouvernance qui permettrait une telle émergence reste inconnu.

Un gestionnaire de communauté sera préoccupé par les formes d'appropriation du savoir que fournit le SI. *Notre recommandation* aux managers est de rappeler que l'appropriation du savoir par les utilisateurs n'est pas « naturelle » et « immédiate », mais peut être un échec partiel pour des raisons que nous avons exposées, au-delà même de la qualité parfois décevante du contenu du wiki. Elle suppose des formes d'intermédiations significatives par une *communauté* voire des formes élaborées de « feed-back »: un système d'information n'est pas qu'un outil et un contenu.

Apports méthodologiques et théoriques

Nos contributions en matière méthodologique concernent en particulier l'approche historique en management des systèmes d'information et l'étude des communautés virtuelles.

Nous avons montré l'intérêt d'une approche historique selon plusieurs perspectives méthodologiques, qualitatives (plusieurs type de codages et d'usage des données textuelles), quantitatives (extraction et reconstitution par décomptage « manuel », étude de séries

temporelles, économétrie simple, analyse de réseau) et avec des données comportant des problèmes différents, soit un excès de données demandant sélection ou « carrotages » (chapitre 8), soit au contraire des données lacunaires et/ou indirectes imposant une reconstitution à partir de traces (chapitre 6). Les données traitées comportent en effet de fortes spécificités qui peuvent se retrouver ailleurs lorsqu'on étudie le « déploiement du numérique ». Pour les sciences de gestion, l'apport est indirect et concerne les études longitudinales, en particulier en montrant le caractère fructueux d'approches plurielles. De plus, rien ne dit que ces approches ne sont pas pertinentes aussi pour des wikis ou communautés « internes » à des organisations, même si nous n'avons pas réfléchi à cette question. Une tentative d'adaptation dans cette voie pourrait être envisageable.

Sur le plan théorique, nous aimerions insister sur un point important ; notre thèse propose un rapprochement entre les sciences de gestion et les sciences cognitives[904]. Nous avons en particulier proposé ici des formes d'approfondissement local des relations entre économie cognitive et approches historiques en gestion : cette approche théorique en économie s'est révélée féconde autant sur des matériaux qualitatifs (chapitre 6 par exemple) que quantitatifs (chapitre 7). Cela peut sembler incongru, mais nous rappelons que certains chercheurs ont déjà évoqué ce type de rapprochements de façon significative dans leurs travaux (LORENZ, 2001). L'apport de cette thèse concerne autant les économistes hétérodoxes que les chercheurs en gestion, d'ailleurs proches.

Limites méthodologiques, théoriques et recherches futures envisageables

Une thèse proposant une étude de cas est souvent incomplète (DUMEZ, 1999). Parmi les limites de notre travail, *deux en particulier* sont a priori pertinentes pour comprendre la gouvernance de communautés faisant usage de wikis. Nous n'avons pu les travailler de façon aussi approfondie que nous aurions souhaité. Nous les avons dégagées comme des hypothèses de travail sérieuses au point qu'elles permettent de mieux comprendre notre cas, ce qui est aussi un apport en soi. Nous allons donc dire où nous en sommes sur ce point.

Nous commençons cependant par évoquer brièvement d'autres pistes de recherche qui importent moins à notre connaissance. Tout d'abord, une autre approche sociologique aurait

[904] Notre travail a été à notre connaissance la seule collaboration entre le CRG et le CREA depuis fort longtemps. Il existait à notre connaissance une collaboration entre Christophe Midler et Laurent Thévenot, il y a plus de 15 ans. Des approfondissements en collaboration avec le CREA seraient souhaitables.

permis par exemple de mieux comprendre les motivations des contributeurs, par des questionnaires électroniques, et de compléter le chapitre 5. Il serait souhaitable de compléter notre approche historique « pure » à partir d'archives par des questionnaires pour aborder davantage les motivations et peut-être les problèmes d'appropriation. Ensuite, un projet open-source est, sans une certaine mesure, un projet : or, cette thématique est un programme de recherche en soi en sciences de gestion, et un dialogue avec des spécialistes de management de projet, que nous n'avons pas mené jusqu'à présent, pourrait être fructueux, bien que l'outil wiki serve davantage à la régulation des comportements qu'à la régulation des tâches (chapitre 8). Enfin, même si nous avons évoqué la question de l'évolution des routines, la question de l'apprentissage organisationnel aurait pu sans doute être davantage approfondie.

Voici à présent la première piste de recherche significative. Sur ce point, rappelons que nous avons assumé dans cette thèse une approche historique, ce qui permet de mettre en évidence aussi les limites de cette approche. La présente piste relève plutôt de la sociologie et est apparue comme pertinente à la marge durant notre travail[905].

Le rôle de la profession de rédacteur de manuels techniques peut être un élément structurant dans les communautés open-source faisant usage de wiki. Cette question de la professionnalisation de la rédaction de documentation a eu un rôle « en amont » pour mieux comprendre certaines caractéristiques observées, comme le désir de qualité en matière de contributions (chapitre 5), cette motivation pouvant être sur ce point *décisive*, ou peut-être l'émergence de forme de contrôle du processus d'édition (chapitre 6). Il est cependant difficile d'en dire davantage à partir du cas WordPress et au-delà pour le moment. Notre travail sur un mode d'implémentation d'un nouvel outil (à l'époque) prolonge la réflexion de BUREAU (2007) sur le rôle de la profession comme principe structurant pour comprendre la dynamique d'une technologie. Cela permettrait en fait de nuancer la rupture que constituerait selon lui les wikis[906] : les catégories d'utilisateurs ne seraient pas si « brouillées » que cela étant donné que le caractère « professionnel » de certaines contributions décisives *persiste*. Nous ne pouvons apporter de réponse définitive sur ce point à partir de nos archives: la discussion avec Sylvain Bureau mériterait de se poursuivre par une étude du rôle de cette profession pour mieux comprendre la dynamique de tels systèmes d'information, les wikis

[905] D'autres chercheurs comme Sylvain Bureau et Jean-Baptiste Suquet ont approfondi entre autres dans leur thèse la thématique de la professionnalisation. Nous ne sommes pas spécialiste de cette question, et l'inspiration vient d'eux, indirectement et inconsciemment. Nous les en remercions au passage.

[906] Ce point sur la professionnalisation est une reprise d'un dialogue de fin 2006 avec Sylvain Bureau qui nous citait dans la conclusion de cette précédente thèse du CRG. Nous sommes peut-être responsable des présentes limites de ses propos, sa thèse ne portant pas sur cette technologie mentionnée en ouverture de son travail.

étant des outils qui peuvent mettre à mal le rôle des professionnels de la documentation. Il y a là un *paradoxe pertinent* à étudier pour les sciences de gestion: le fait que les SI reposant sur des wikis *se servent* de la professionnalisation de rédacteur de manuels techniques pour être viables tout en *la contestant* en étant ouvert à *tous* pour une bonne part d'entre eux, fait que le phénomène n'est pas trivial. Il est certainement à mettre en relation avec les formes de contributions et d'interventions en forme de loi de puissance observées, les professionnels formant une partie significative du cœur des contributeurs dont l'articulation avec les autres contributeurs plus mineurs et non-experts pourrait être problématique car moins légitime. C'est là peut-être une limite, cette fois « sociologique », au modèle « pur » de la « collaboration massivement distribuée ».

Un tel questionnement sur ce paradoxe n'aurait pu en effet être mené dans une thèse portant *sur un seul cas*, même si ce dernier permet de documenter un paradoxe pertinent. Il faudrait en étudier plusieurs, comme l'a d'ailleurs fait Sylvain Bureau, et aller sans doute au-delà de quatre cas. Une *méthodologie mixte* par entretiens et étude des archives (en extrayant des archives la racine « profession » par exemple) des communautés serait recommandée[907]. Il y a là une piste de recherche à mener en gestion et un sujet d'article envisageable.

Voici à présent la deuxième piste de recherche. Dans le cadre de notre étude de cas, il est une question de gouvernance que nous pensions traiter par une approche en économie cognitive, mais sur laquelle nous n'avons pu aboutir, en particulier faute de temps. Il s'agit de la question du rapport entre contributions et pouvoir dans la communauté. Le modèle sous-jacent qui permettrait de comprendre la dynamique conjointe des contributions et discussions serait celui du « signal coûteux » (« costly signalling »). Voici comment un économiste envisage ce mécanisme:

> "We propose an explanation of cooperation among unrelated members of a social group, in which providing group benefits evolves because it constitutes an honest signal of the member's quality as a mate, coalition partner or competitor, and therefore results in advantageous alliances for those signaling in this manner." (SMITH et al., 2000)

Plus un individu donne, plus il se signale aux autres comme crédible. En sciences cognitives, DESSALLES (2006, 2007) a étendu ce type d'approche d'économie cognitive à la question de la communication et du langage en général, et pensait montrer en quoi Wikipedia confirmait ces théories. L'idée du modèle de Dessalles est qu'il est intéressant de donner et

[907] La contrepartie serait alors de ne pas étudier les cas aussi en profondeur que dans la présente thèse.

d'être celui qui donne pour acquérir du pouvoir dans un groupe, pour montrer qu'on est un « allié crédible ». Son inspiration est aussi biologique.

Nous pensions traiter ce point au chapitre 5. Nous n'avons pu mener à bien cette question, même si nous avons pu mettre en évidence des éléments en ce sens ci-dessous. Le premier, i.e. le graphique ci-dessous, montre qu'il existe un lien entre contributions sur le wiki et discussions sur la liste de discussion parmi ceux qui interviennent sur l'importante liste de régulation « hackers ». Le second est un fragment que nous avons déjà cité, qui montre qu'une population de contributeurs paraît monopoliser ce travail.

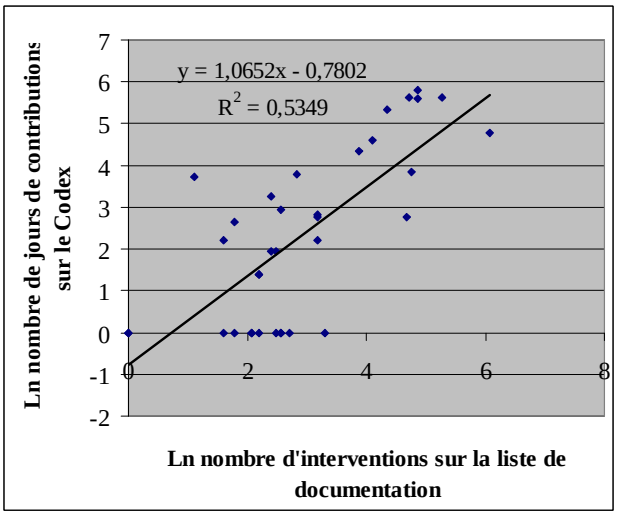


Figure 103 : mise en relation du nombre d'intervention sur la liste de discussion et du nombre de jours de contributions sur le Codex

> "I was keen on documentation. Along with Podz and Carthik and others, we got the initial wiki going […]. Soon we had lots of good folks jumping in to help out, and they were indeed wiki savvy. However, what happened was that soon these people were basically doing everything, while people like me wound up being relegated to the sidelines. I could have learned on my own how to do a lot of things, but the point I'm trying to make is that had I been paired up with someone who was more advanced than I, we could have had two "experts" in time. Then we train two, and they train two etc. You get a CSS-effect in your documention team development […]. Right now the support and Codex are dominated by a small group. Again, […] it's simply the reality of a lot of work to be done, and only a core group of folks with the occasional help from a lot of others."[908]

[908] Source : http://comox.textdrive.com/pipermail/wp-forums/2006-October/003788.html

Ce type de modèle en économie cognitive décrit un mécanisme de synthèse qui fait le lien entre le processus politique de la discussion collaborative (chapitre 7) et les contributions au wiki (chapitre 5). Nous soulignons aussi que ce mécanisme émergent permettant la gouvernance d'un wiki aurait des *effets contraires*, comme dans le cas précédent de la professionnalisation : d'une part il permet de comprendre mieux les motivations à contribuer[909], d'autre part, il créerait un *effet d'éviction des contributions ou de la collaboration*, que le fragment précédent permet de suggérer. Ce mécanisme est délicat et ses effets sont incertains. Il peut aussi être responsable de l'absence *par éviction* de « middle management » dans la communauté, un problème managérial que nous avons évoqué ci-dessus et au chapitre 6.

Une recherche en gestion sur ce type d'émergence supposerait un travail fin sur *ce processus*, et donc une recherche de données pertinentes en ce sens. Pour arriver à tester la pertinence de ce type de modèle, nous aurions dû disposer d'une base de données « utilisateurs » *complète* incluant tous les types d'interventions dans notre communauté. Or, ces variables sont séparées sur Excel pour le moment, à l'exception de ce que nous avons pu effectuer localement à partir des contributeurs au Codex qui interviennent aussi sur la liste de discussions des hackers. Un travail supplémentaire d'ingénierie aurait été trop long. Il aurait fallu aussi une enquête qualitative plus complète corroborant davantage le fragment ci-dessus. L'approche théorique devrait aussi spécifier le modèle théorique et en particulier le processus qu'il décrirait, comme le demandait DUMEZ (2005), le processus « politique » restant encore trop général pour un chercheur en gestion.

Ces deux pistes de recherche sont importantes pour soutenir notre thèse sur l'émergence de la gouvernance d'un système d'information reposant sur un SI ; si elles sont vérifiées, elles permettraient de rendre encore plus solides nos présents propos. Or, notre cas suggère au minimum de *fortes présomptions* en ce sens. Elles permettent en tout cas de nuancer en connaissance de cause les conclusions du chapitre 5 sur les motivations et les contributions.

D'autres pistes de travail sont enfin envisageables au-delà de la communauté WordPress.

[909] Quelques fragments évoquant au delà de la méritocratie le lien entre pouvoir et contributions on été relevés. Ils décrivent ce processus par les termes "Volonterr based organization" (sic ! Source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/001031.html) ou "Do-ocraty" (source : http://comox.textdrive.com/pipermail/wp-docs/2005-September/001008.html).

Il serait profitable de généraliser les conclusions de la revue de littératures pour une publication exhaustive traitant de la littérature sur les wikis. Par ailleurs, notre étude de cas a permis de souligner l'importance de nombreuses questions, en particulier celles de la coordination des contributeurs et de l'appropriation d'un SI, par comparaison avec les limites du traitement de cette question dans la littérature. Cela permettrait d'interroger à nouveau la littérature, en un mouvement de « retour de spirale ».

Une piste attendue après une étude de cas unique en profondeur qui profite de matériaux opportun est de généraliser nos conclusions de chacun de nos cinq chapitres de résultats à d'autres communautés open-source faisant usage de wikis. Des travaux ultérieurs sur une population représentative de wikis dans les communautés open-source serait selon nous absolument prioritaires, ce que montre en particulier ROTH (2008). Cela serait fructueux autant sur des projets moins connus, plus petits, ou non référencés, qui sont a priori les plus nombreux, mais aussi sur les wikis de projets d'une importance *prioritaire* pour le mouvement open-source, comme le wiki d'Ubuntu par exemple présenté au chapitre 1. Nous avions commencé à créer une base de données en début de thèse répertoriant des centaines de projets employant des wikis. Elle est à présent périmée et à réactualiser. Cela pourrait faire l'objet d'une thèse à nouveau, ou d'un mémoire d'habilitation, par exemple. Si on s'intéresse par exemple aux aspects dynamiques de l'organisation des communautés, des recherches restent à coup sûr à mener sur ce point : or, la démarche que nous avons adoptée, par exemple au chapitre 4 avec le dépouillement du nombre de courriels est *reproductible* et d'autres chercheurs savent à présent comment la mener. Nous pensons par exemple au groupe de chercheurs du récent projet ANR « CCCP Posodie »[910], qui aborde la question des wikis et dans lequel travaillent des chercheurs de l'ENST. Enfin, sur le plan technique, des collaborations équilibrées avec des informaticiens seraient intéressantes afin de pouvoir étudier des wikis qui n'appartiennent pas à la fondation Wikimedia. C'est l'un des problèmes techniques auxquels nous nous sommes heurté.

Une comparaison de nos résultats avec les résultats d'autres enquêtes ou études de cas, présentes ou à venir portant sur les wikis « corporates », très nombreux et auxquels nous avons fait allusion dans notre revue de littérature, serait également un travail pertinent, mais plus délicat à mener étant donné que le contexte institutionnel n'est pas le même que dans les communautés open-source.

[910] Il n'y a encore rien sur Internet à propos de ce projet.

Annexes générales

Annexe 1 : liens vers les principales archives en ligne sélectionnées

Le wiki de la communauté appelé Codex :
http://codex.wordpress.org/Main_Page

Les archives des listes de discussions :
http://lists.automattic.com/mailman/listinfo

Les archives des anciennes listes de discussions non répertoriées dans l'entrepôt officiel :
http://comox.textdrive.com/pipermail/hackers/
http://comox.textdrive.com/pipermail/docs/
http://comox.textdrive.com/pipermail/cvs/

Gmane, système « miroir » d'archivages de listes de discussions :
http://www.gmane.org/

L'archivage de l'évolution de la communauté par WebArchives :
http://web.archive.org/web/*/http://wordpress.org/

Les archives des « minutes » de dialogue sur canal de « chat » IRC :
http://codex.wordpress.org/IRC_Meetups

Le blog de WordPress dédié au développement technique du projet:
http://wordpress.org/development/ .

Le forum :
http://wordpress.org/support/

Annexe 2 : autres liens menant vers la communauté WordPress

Sites d'entreprise : http://automattic.com/ , http://wordpress.com/ .

La page principale du projet : http://wordpress.org/ .

La page d'origine sur Sourceforge: https://sourceforge.net/projects/cafelog/ .

Le wiki Trac, destiné en particulier au développeurs de « plugins » : http://dev.wp-plugins.org/ mais aussi : http://dev.wp-plugins.org/wiki/ .

Répertoire des archives des mailing-lists :http://codex.wordpress.org/Mailing_Lists .

Le CVS (Content Version System): https://sourceforge.net/cvs/?group_id=51422 .

Une page « communautaire » : http://planet.wordpress.org/ .

Annexe 3 : glossaire de vocabulaire informatique

Nous n'avons pas cité tous les termes informatiques ou les abréviations mentionnés dans notre travail de recherche, en particulier dans les archives. Nous indiquons seulement ceux que le lecteur pourra trouver dans cette thèse, en particulier dans les fragments d'archives. Mais nous avons tenté de simplifier le propos en la matière, en limitant leur emploi. Nous supposons que le lecteur sait ce qu'est un courriel, un forum, un wiki, un blog par exemple. Nous n'avons pas en revanche défini des termes techniques usuels en sciences sociales. Bien entendu, les définitions sont *élémentaires et succinctes*, et nous n'avons pas précisé à chaque fois les sources d'informations étant donné que ce guide n'a pas d'intérêt pour la recherche et sert juste d'aide à la compréhension.

Administrateur ou « admin » : le statut d'administrateur sur un wiki donne au minimum des possibilités d'interventions techniques supplémentaires à celui qui en dispose. Il s'agit pour l'essentiel de bloquer sur plusieurs points les interventions des utilisateurs au besoin, protéger des pages, intervenir sur les historiques d'éditions, contrôler les éditions des utilisateurs (patrouiller), importer du contenu, changer l'emplacement des pages. Ces possibilités ne disent rien en revanche du comportement effectif des administrateurs, en particulier en matière de management du contenu et des utilisateurs. Ces prérogatives peuvent varier d'un wiki à l'autre.

Apache : principal logiciel de serveur open-source.

API : « Application Programming Interface », bibliothèque de fonctions logicielles.

ATM : "at the moment".

Beta versions : versions intermédiaires imparfaites d'un logiciel publiées entre les versions officielles numérotées voire baptisées.

bb2 : logiciel de forum open-source ayant servi entre autres d'inspiration pour créer WordPress.

Browser: navigateur Internet.

Bug : défaut de programmation.

Bureaucrats : dans le cas de WordPress, synonyme d'administrateur.

ChaMots (« CamelCase ») : pratique qui consiste à accoller deux mots. C'est une routine des systèmes wikis pour créer automatiquement des liens « derrière » le nouveau terme. Notre lecteur aura sans doute deviné l'origine visuelle du mot « ChaMots ».

« Content Management system » (CMS): les logiciels de gestion de contenu sont l'une des grandes catégories de logiciels proposées par Framasoft (source : http://www.framasoft.net/, consulté le 10/06/09).

Creative Commons (CC): désigne un ensemble de licences libres permettant la diffusion et la diffusion d'œuvre culturelles sur Internet.

Cascading Style Sheets (« CSS ») : feuilles de style en cascade, c'est-à-dire sous forme de présentation séparée reproductible.

Custom fields: dans un CMS, champ ou catégorie à définir puis à remplir.

Concurrent Version Systems (CVS) : logiciel qui permet de traiter les changements dans les entrepôts publics de code source.

DB : « database ».

Debian : l'une des suites Linux.

DocBook : logiciel de production livresque et documentaire (http://www.docbook.org/).

Dotclear : moteur de blog francophone renommé.

Draft : brouillon, ici première ébauches de documentation sur les pages utilisateurs.

Drupal : logiciel de gestion de contenu (http://drupalfr.org/).

Dummies : novices.

Dumps : entrepôts des historiques des wikis sous format XML.

End-user : utilisateur final (d'un logiciel).

Extreme programming : programmation extrême, méthode agile de programmation adaptée aux petites équipes, et qui pousse à l'extrême des principes d'organisation simples.

FAQ : « Frequently Asked Questions », questions fréquemment posées dans une communauté.

Featured article : pratique de Wikipedia qui consiste à mettre en évidence sur la page principale les articles dont la qualité est reconnue.

Filezilla : logiciel de transfert de fichiers d'un ordinateur à un site web.

Flaming : pratique usuelle en particulier dans les communautés open-source qui consiste à s'exprimer vis-à-vis d'un participant de façon dure, voire insultante, de façon individuelle mais aussi collective. Il s'agit d'une forme de sanction.

Fonctions (« functions ») : composants élémentaires du langage PhP.

Forum : site et logiciel classique sur Internet où une communauté pose des questions (« posts ») et publie des réponses.

Framasoft : portail renommé de présentation de logiciels libres.

Free Software : logiciel dont la licence permet ad libidum l'usage, l'étude, la diffusion, et la modification.

Geeks : personne parfois bizarre passionnée par un domaine précis, en particulier l'informatique.

GNU : système d'exploitation libre.

GPL : General Public Licence.

Gravatar : avatar universel, c'est-à-dire identifiant numérique valable pour plusieurs localisations du Web. Le lecteur en trouvera beaucoup dans cette thèse, parfois étranges (« MDAWaffe »).

Greymatter : entreprise de logiciels. http://www.greymatter.com/ .

Groupware : logiciel permettant le partage de documents à distance, ou logiciel de travail collaboratif.

Guidelines : « lignes directrices », c'est à dire un ensemble de principes permettant d'orienter le comportement des participants, sans pour autant qu'il y ait sanction en cas de transgression.

Hacker : expert en informatique en particulier dans le domaine du logiciel libre. Ce terme n'implique aucunement le fait de pratiquer des actions illégales.

HTML : langage de programmation « Hypertext Markup Langage » servant à créer des sites Web.

IMVHO : « In My Very Humble Opinion ».

IRC : logiciel de « chat » servant d'outil de coordination et dont les interactions verbales sont parfois archivées. Ce n'est pas qu'un outil de divertissement.

IT : « information technologies ».

Javascript : langage de programmation servant aux pages web interactives.

KDE : l'une des suites Linux.

Linux : important système d'exploitation open-source.

Literate programming : écriture du programme « à la main » sans intervention d'autres programmes.

Log : retranscription numérique, texte intégral des interactions entre participants.

Login : mot de passe.

LOL : « laughing out loud », expression signalant qu'on rit de façon marquée dans une conversation sur support numérique

Lurkers : participants à un projet ou à un canal de communication, qui ne font rien, ne parlent pas.

MediaWiki : logiciel (« moteur ») wiki servant à Wikipedia et aux projets de la Wikimedia foundation.

Main space : espace d'un wiki où se trouve le contenu informationnel, par opposition au « user-space ».

Mozilla : fondation renommée ayant fait le choix de l'innovation sur Internet en proposant des logiciels libres, comme Firefox.

« My 2 cents » (ou « 2bobs ») : cette expression est très employée dans la communauté : elle signifie qu'une personne apporte une contribution qu'elle juge minime (comme à la quête par exemple), qu'il s'agisse d'idées sur la liste de discussion, ou de contributions directes au Codex.

MySQL : logiciel de base de données open-source.

Newbie : utilisateur novice.

Nightlies : Il s'agit des versions « nocturnes », i.e. « non visibles à la lumière ». Il arrive aux informaticiens de parfois rester travailler par addiction à leur code la nuit.

Open-Office : suite de logiciels libres servant aux fonctionnalités usuelles bureautiques (traitement de texte, tableur…).

OSS : open source software.

PhP : langage de programmation servant à concevoir des sites web dynamiques.

phpBB : logiciel open-source de création de forum (http://www.phpbb.com/)

Plugin : logiciel servant de fonctionnalité additionnelle à un autre logiciel.

Post : « billet », i.e. ensemble de textes postés sur un site relationnel permettant un dialogue par affirmation et réponse.

QA : question and answer.

Release : « publication », en général d'une nouvelle version d'un logiciel.

Release early, release often : principe usuel dans les communautés open-source qui consiste à publier tôt et fréquemment les nouvelles versions afin d'obtenir des réponses des utilisateurs et d'améliorer le produit.

Repository of knowledge : "entrepôt de savoir".

RTFM : « Read The F[censuré] Manual », expression abrégée potentiellement désagréable où on conseille à un utilisateur qui questionne a priori inutilement la communauté d'aller lire le manuel.

Sandbox : signifie « bac à sable », et fait allusion à une page du wiki où chacun peut s'entraîner à écrire (pour apprendre la syntaxe wiki par exemple) sans danger pour les autres contributeurs, un peu comme un enfant écrit sur le sable qui s'efface facilement.

Scholarpedia : encyclopédie collaborative fondée sur le principe du wiki et des contributions d'experts identifiés.

Skin : « peau », c'est-à-dire présentation graphique d'une application ou d'un logiciel.

Sourceforge : site de centralisation de nombre de projets open-source.

SSII : Société de Service en Informatique et Infogérance.

Stub : « souche », un article à l'état de brouillon accessible à tout le monde sur le Codex et que tout le monde peut compléter, et un « draft », un brouillon soumis à une autorisation, quoique aucun mécanisme technique n'empêche de le faire. La liste d'articles de cette catégorie est générée automatiquement sur une page spéciale.

Subversion (SVN) : logiciel de suivi des modifications du code du logiciel.

Systadmins : synonyme d'administrateur.

Tag : mot catégorie servant à classer les textes du Web.

Talk : désigne les onglets de discussion du logiciel MediaWiki.

Template : « pochoir », ici patron de mise en page pour le blog WordPress.

Thème : decorum numérique d'arrière-plan d'un logiciel de gestion de contenu susceptible d'être reproduit d'une page web à l'autre, comme le papier peint peut être identique d'une pièce à l'autre.

Thread : fil (de discussion), en particulier dans les archives de listes de discussions.

Trac : Moteur de wiki spécialement dédié au management de projet logiciel.

Upgrade: « mise à jour », en général d'un logiciel.

URL : suite de termes dans la barre d'adresse d'un moteur de recherche servant à localiser les sites web.

User:talk : onglet du wiki se trouvant à côté de l'onglet de page personnelle d'un utilisateur et servant à communiquer entre utilisateurs, comme une messagerie publique.

User space « espace utilisateurs », ensemble des pages d'un wiki servant à l'expression personnelle des utilisateurs, par opposition au « main space ».

Village Pump : lieu virtuel de discussion des aspects techniques de Wikipedia.

Walled gardens : « jardins clos », c'est-à-dire espace de production informationnelle isolée dans un wiki. La philosophie wiki recommande au contraire le travail public et fortement inter relié.

Web 2.0. : désigne une philosophie et une pratique du Web où les utilisateurs sont producteurs du contenu et interagissent entre eux, appropriation qui s'opposerait à un Web où le contenu serait statique.

Web standards : standards du Web, une tentative de l'association W3C entre autres de normaliser l'informatique de l'infrastructure technique de la *Toile*, et non pas du réseau Internet. http://www.webstandards.org/ .

Widget : composant élémentaire d'interface graphique

WP : WordPress !

WP-MU : WordPress Multi-Users.

Workflow : « flux de travail » entre autres dans une organisation.

XML : « langage extensible de balisage »

Annexe 4 : script du programme d'extraction (« parsing ») en langage python écrit par Jean-Philippe Cointet

```
import sys
import string

def input(prompt='', converter=eval):
  while 1:
    sys.stdout.write(prompt)
    sys.stdout.flush()
    line = sys.stdin.readline().rstrip('\n\r')
    try:
      return converter(line)
    except (KeyboardInterrupt, SystemExit):
      raise
    except Exception, e:
      print str(e)

def convertStr(s):
        """Convert string to either int or float."""
        ret = -1
        try:
                ret = int(s)
        except ValueError:
                #Try float.
                try:
                        ret = float(s)
                except ValueError:
                        pass
        return ret

if __name__ == '__main__':
        #fichier = input("Enter file name to be converted: ", str)
        fichier = "reseau_hackers.txt"
        sortie = "event_"+fichier
        reseau = "reseau_"+fichier
        #print fichier
```

```
	#print "Result: %d" % fichier
	#print "Result: %d" % input("Enter integer:", int)
#print "Result: %r" % input("Enter expression:")
	fh = open(fichier)
	fg = open(sortie,"w")
	fe = open(reseau,"w")
	numero =0
	clause = 1
	compteur = -1
	words = 2000
	auteur  = 5000
	premier = 0
	clause = 0
	date = 0
	compteur=0
	auteur =  {}
	nom = ''
	for line in fh.readlines():
		#print premier
		if clause == 1 and date == 1:
			#print "date"
			decompose = string.split(line)
			for i in decompose[2:]:
				#print i

				fg.write(i)
				fg.write(" ")
			fg.write("\n")
			date = 0
			clause = 0
			compteur = 1

		if (line[0:5]) == 'From:' and premier == 1 and clause == 0:
				print "nom"
				clause = 1
				premier = 0
				date = 1
				decompose = string.split(line)
				for i in decompose[4:]:
					print i
					print '\t'
					nom = nom + i
				nom = nom[1:-1]
				print nom
				fg.write(nom)
				fg.write("\n")
				compteur =0
		if (line[0:4]) == 'From' and premier == 0 and clause == 0:
				print "debut"
				premier = 1
```

```
                        clause = 0
                        compteur =0

            if line[0:7]=='In-Repl':
                    print 'Inside in-rep'
                    decompose = string.split(line)
                    for i in decompose[1:]:
                            reponse = i
                            #print reponse
                    if auteur.has_key(reponse):
                                    fe.write(nom+'  '+auteur[reponse]+'\n')
            if line[0:10]=='Message-ID':
                    decompose = string.split(line)
                    for i in decompose[1:]:
                            mesid = i
                    auteur[mesid]=nom
                    nom =''
            fg.write("\n")
    fh.close()
    fg.close()
    fe.close()
```

Annexe 5 : autres indicateurs d'analyse de réseau

Number of vertices (n): 129

	Arcs	Edges
Total number of lines	1768	0
Number of loops	67	0
Number of multiple lines	1084	0

Info/Networks/Indices :

The number of vertices: n=129
n ln n =626.91580
Total adjacency index: m=1768
The Zagreb group index 1: M1=735496
The Zagreb group index 2: M2=77650283
The Randic connectivity index: Xr=32.06468
The Platt index: F=731960

Index of relinking: P=22.77778

Info/Partitions/50
---
4. All Degree partition of N12 (129)

Dimension: 129
The lowest value: 1
The highest value: 431

The highest clusters values:

Rank	Vertex	Cluster	Id
1	2	431	MATT
2	14	414	PODZ
3	13	376	CARTHIK
4	4	306	CRAIG
5	56	211	LORELLE
6	18	140	SKIPPY
7	36	133	OwenWinkler
8	51	95	MDAWAFFE
9	73	78	MICHAELH
10	1	76	Cena
11	16	70	RyanBoren
12	12	68	MatthewThomas
13	46	66	MorganDoocy
14	85	55	MichaelB
15	49	54	KAFKAESQUI
16	53	47	JamesHuff
17	27	46	ChrisWaigl
18	121	45	JenniferHodgdon
19	5	42	JosiahRitchie
20	59	39	AaronBrazell
21	45	34	RyanDuff
22	20	32	Moshu
23	70	30	RobertDeaton
24	3	27	MichaelRenzmann
25	96	24	TechGnome
26	8	24	DougalCampbell
27	7	22	GeofF.Morris
28	48	22	ArthurJennings
29	9	22	ChristianBalzer
30	31	19	TorBjornrud
31	26	18	JustinMoore
32	29	15	lokrin
33	89	15	WESTI
34	19	15	Kitty
35	88	14	XavierBorderie
36	77	14	DavidEads
37	63	13	virgilhuston
38	122	13	MartinSturm

39	116	13	JerryPeek
40	61	12	FirasD.
41	57	12	KSuominen
42	119	12	MarkStyles
43	94	12	DavidHouse
44	117	12	LloydBudd
45	33	12	JohnWatson
46	11	11	Me
47	69	11	ChristianBalzer(TigerDE2)
48	86	10	DarrellSchulte
49	81	10	AnthonyTimberlake
50	37	10	miahz

Frequency distribution of cluster numbers:

Cluster	Freq	Freq%	CumFreq	CumFreq%	Representative
1	29	22.4806	29	22.4806	BobTorres
2	16	12.4031	45	34.8837	milquetoast@telus.net
3	7	5.4264	52	40.3101	BrianShirk
4	10	7.7519	62	48.0620	BrianMeidell
5	6	4.6512	68	52.7132	StephenSadowski
6	3	2.3256	71	55.0388	theabsentstudent
7	4	3.1008	75	58.1395	HanniRoss
8	2	1.5504	77	59.6899	mikelittle
10	5	3.8760	82	63.5659	RyanJ.Markel
11	2	1.5504	84	65.1163	Me
12	6	4.6512	90	69.7674	JohnWatson
13	3	2.3256	93	72.0930	virgilhuston
14	2	1.5504	95	73.6434	DavidEads
15	3	2.3256	98	75.9690	Kitty
18	1	0.7752	99	76.7442	JustinMoore
19	1	0.7752	100	77.5194	TorBjornrud
22	3	2.3256	103	79.8450	GeofF.Morris
24	2	1.5504	105	81.3953	DougalCampbell
27	1	0.7752	106	82.1705	MichaelRenzmann
30	1	0.7752	107	82.9457	RobertDeaton
32	1	0.7752	108	83.7209	Moshu
34	1	0.7752	109	84.4961	RyanDuff
39	1	0.7752	110	85.2713	AaronBrazell
42	1	0.7752	111	86.0465	JosiahRitchie
45	1	0.7752	112	86.8217	JenniferHodgdon
46	1	0.7752	113	87.5969	ChrisWaigl
47	1	0.7752	114	88.3721	JamesHuff
54	1	0.7752	115	89.1473	KAFKAESQUI
55	1	0.7752	116	89.9225	MichaelB
66	1	0.7752	117	90.6977	MorganDoocy
68	1	0.7752	118	91.4729	MatthewThomas
70	1	0.7752	119	92.2481	RyanBoren
76	1	0.7752	120	93.0233	Cena

78	1	0.7752	121	93.7984	MICHAELH
95	1	0.7752	122	94.5736	MDAWAFFE
133	1	0.7752	123	95.3488	OwenWinkler
140	1	0.7752	124	96.1240	SKIPPY
211	1	0.7752	125	96.8992	LORELLE
306	1	0.7752	126	97.6744	CRAIG
376	1	0.7752	127	98.4496	CARTHIK
414	1	0.7752	128	99.2248	PODZ
431	1	0.7752	129	100.0000	MATT
Sum	129	100.0000			

Annexe 6 : description précise des interactions entre intervenants dans le réseau de la liste de discussions de la documentation

- Les relations en gras sont celles où il y a plus de 5 arcs de l'un vers l'autre.
- Les relations en italique sont celles qui sont inhabituellement basses entre administrateurs.

Edges: 1768
There are: 67 self-loops.

There are: 1084 parallel edges.
They are as follows:
There are 4 edges between node: OwenWinkler and node: MorganDoocy
There are 5 edges between node: MATT and node: MorganDoocy
There are 3 edges between node: RyanDuff and node: ArthurJennings
There are 2 edges between node: SherylCoe-Reportica and node: LORELLE
There are 3 edges between node: MDAWAFFE and node: LORELLE
There are 4 edges between node: PODZ and node: MichaelB
There are 4 edges between node: ChrisWaigl and node: CRAIG
There are 4 edges between node: JohnWatson and node: PODZ
There are 2 edges between node: CARTHIK and node: AaronBrazell
There are 3 edges between node: RyanBoren and node: LORELLE
There are 2 edges between node: CARTHIK and node: TechGnome
There are 2 edges between node: LORELLE and node: OwenWinkler
There are 3 edges between node: MichaelRenzmann and node: MATT
There are 2 edges between node: TrevorTurk and node: PODZ
There are 5 edges between node: ChrisWaigl and node: MATT
There are 3 edges between node: JenniferHodgdon and node: CARTHIK
There are 2 edges between node: virgilhuston and node: LORELLE
There are 2 edges between node: IndranilDasgupta and node: MATT
There are 2 edges between node: Me and node: MATT
There are 3 edges between node: JustinMoore and node: CRAIG
There are 2 edges between node: LORELLE and node: KAFKAESQUI
There are 2 edges between node: MATT and node: RyanDuff
There are 2 edges between node: ChristianBalzer(TigerDE2) and node: LORELLE
There are 6 edges between node: MartinSturm and node: JenniferHodgdon
There are 4 edges between node: MDAWAFFE and node: JamesHuff
There are 2 edges between node: ChrisWaigl and node: CARTHIK
There are 2 edges between node: MATT and node: Me

There are 3 edges between node: MichaelRenzmann and node: RyanDuff
There are 3 edges between node: RyanDuff and node: PODZ
There are 2 edges between node: KAFKAESQUI and node: MichaelB
There are 3 edges between node: MorganDoocy and node: ChrisWaigl
There are 14 edges between node: MATT and node: CARTHIK
There are 2 edges between node: FirasD. and node: CRAIG
There are 2 edges between node: .....herebox.org..... and node: MATT
There are 3 edges between node: PODZ and node: JamesHuff
There are 5 edges between node: MATT and node: MatthewThomas
There are 5 edges between node: CRAIG and node: SKIPPY
There are 4 edges between node: PODZ and node: AaronBrazell
There are 2 edges between node: RyanDuff and node: ChrisWaigl
There are 6 edges between node: CRAIG and node: JosiahRitchie
There are 2 edges between node: MichaelRenzmann and node: Cena
There are 7 edges between node: LORELLE and node: MichaelB
There are 2 edges between node: Jewel and node: LORELLE
There are 3 edges between node: JohnWatson and node: CRAIG
There are 6 edges between node: LORELLE and node: PODZ
There are 6 edges between node: MATT and node: SKIPPY
There are 2 edges between node: RobertDeaton and node: SKIPPY
There are 7 edges between node: JustinMoore and node: PODZ
There are 6 edges between node: MATT and node: Cena
There are 4 edges between node: MDAWAFFE and node: PODZ
There are 2 edges between node: RobertDeaton and node: DavidEads
There are 5 edges between node: MATT and node: MATT
There are 13 edges between node: CARTHIK and node: MDAWAFFE
There are 2 edges between node: CRAIG and node: miahz
There are 11 edges between node: MDAWAFFE and node: CARTHIK
There are 6 edges between node: CRAIG and node: OwenWinkler
There are 2 edges between node: ChrisWaigl and node: ChrisWaigl
There are 2 edges between node: DavidEads and node: MATT
There are 3 edges between node: CRAIG and node: TorBjornrud
There are 3 edges between node: MorganDoocy and node: MorganDoocy
There are 3 edges between node: TechGnome and node: PODZ
There are 3 edges between node: AaronBrazell and node: PODZ
There are 2 edges between node: StephenSadowski and node: MATT
There are 2 edges between node: MichaelRenzmann and node: ChristianBalzer
There are 2 edges between node: DougalCampbell and node: CRAIG
There are 4 edges between node: MATT and node: RyanBoren
There are 6 edges between node: CARTHIK and node: RyanBoren
There are 5 edges between node: JenniferHodgdon and node: LORELLE
There are 4 edges between node: PODZ and node: MorganDoocy
There are 6 edges between node: CARTHIK and node: KAFKAESQUI
There are 6 edges between node: MorganDoocy and node: PODZ
There are 2 edges between node: CRAIG and node: CRAIG
There are 2 edges between node: CRAIG and node: JohnWatson
There are 5 edges between node: CRAIG and node: ChrisWaigl
There are 19 edges between node: OwenWinkler and node: PODZ
There are 7 edges between node: OwenWinkler and node: MDAWAFFE
There are 2 edges between node: AaronBrazell and node: AaronBrazell
There are 7 edges between node: OwenWinkler and node: CARTHIK
There are 2 edges between node: KAFKAESQUI and node: CARTHIK
There are 3 edges between node: RyanBoren and node: CARTHIK
There are 2 edges between node: RobertDeaton and node: PODZ
There are 3 edges between node: TechGnome and node: MichaelB
There are 14 edges between node: SKIPPY and node: MICHAELH
There are 12 edges between node: RyanBoren and node: PODZ
There are 2 edges between node: lokrin and node: MATT
There are 8 edges between node: CARTHIK and node: MICHAELH
There are 4 edges between node: KAFKAESQUI and node: PODZ

There are 2 edges between node: KAFKAESQUI and node: KAFKAESQUI
There are 12 edges between node: PODZ and node: PODZ
There are 2 edges between node: MarkStyles and node: MATT
There are 4 edges between node: SKIPPY and node: CRAIG
There are 4 edges between node: AaronBrazell and node: LORELLE
There are 13 edges between node: CARTHIK and node: OwenWinkler
There are 41 edges between node: CRAIG and node: MATT
There are 3 edges between node: JamesHuff and node: JamesHuff
There are 2 edges between node: Moshu and node: LORELLE
There are 2 edges between node: PODZ and node: KAFKAESQUI
There are 4 edges between node: LORELLE and node: virgilhuston
There are 2 edges between node: CRAIG and node: LORELLE
There are 4 edges between node: Moshu and node: CARTHIK
There are 2 edges between node: PODZ and node: RyanDuff
There are 5 edges between node: OwenWinkler and node: SKIPPY
There are 2 edges between node: AnthonyTimberlake and node: MATT
There are 2 edges between node: LORELLE and node: web@reportica.net
There are 2 edges between node: CRAIG and node: MichaelRenzmann
There are 4 edges between node: PODZ and node: ArthurJennings
There are 2 edges between node: CARTHIK and node: TorBjornrud
There are 2 edges between node: LORELLE and node: WESTI
There are 8 edges between node: CRAIG and node: Cena
There are 2 edges between node: MATT and node: AnthonyTimberlake
There are 4 edges between node: JamesHuff and node: MDAWAFFE
There are 3 edges between node: CRAIG and node: GeofF.Morris
There are 46 edges between node: MATT and node: PODZ
There are 2 edges between node: DarrellSchulte and node: MICHAELH
There are 6 edges between node: MATT and node: OwenWinkler
There are 2 edges between node: Kitty and node: MATT
There are 2 edges between node: MATT and node: RyanJ.Markel
There are 2 edges between node: HanniRoss and node: CRAIG
There are 7 edges between node: MDAWAFFE and node: CRAIG
There are 2 edges between node: DavidHouse and node: PODZ
There are 8 edges between node: CRAIG and node: MDAWAFFE
There are 2 edges between node: JenniferHodgdon and node: JenniferHodgdon
There are 2 edges between node: TechGnome and node: Vogel,Andrew(vogelap)
There are 3 edges between node: GeofF.Morris and node: MATT
There are 2 edges between node: LORELLE and node: LORELLE
There are 4 edges between node: MATT and node: ChristianBalzer
There are 2 edges between node: WESTI and node: MATT
There are 19 edges between node: PODZ and node: CARTHIK
There are 2 edges between node: MichaelRenzmann and node: PODZ
There are 2 edges between node: GeofF.Morris and node: Cena
There are 6 edges between node: JamesHuff and node: CARTHIK
There are 2 edges between node: KAFKAESQUI and node: MICHAELH
There are 2 edges between node: JosiahRitchie and node: JosiahRitchie
There are 6 edges between node: CARTHIK and node: SKIPPY
There are 3 edges between node: MorganDoocy and node: MatthewThomas
There are 2 edges between node: MATT and node: WESTI
There are 6 edges between node: JosiahRitchie and node: MATT
There are 18 edges between node: PODZ and node: CRAIG
There are 3 edges between node: MorganDoocy and node: CARTHIK
There are 12 edges between node: LORELLE and node: MATT
There are 2 edges between node: KAFKAESQUI and node: SKIPPY
There are 2 edges between node: MarkStyles and node: JerryPeek
There are 5 edges between node: GeofF.Morris and node: CRAIG
There are 4 edges between node: CARTHIK and node: MichaelB
There are 3 edges between node: LORELLE and node: MICHAELH
There are 2 edges between node: RyanDuff and node: LORELLE
There are 17 edges between node: CARTHIK and node: CARTHIK

There are 3 edges between node: CARTHIK and node: XavierBorderie
There are 8 edges between node: SKIPPY and node: PODZ
There are 3 edges between node: OwenWinkler and node: CRAIG
There are 4 edges between node: Moshu and node: SKIPPY
There are 2 edges between node: MichaelB and node: MATT
There are 6 edges between node: JosiahRitchie and node: Cena
There are 4 edges between node: CRAIG and node: DougalCampbell
There are 2 edges between node: MDAWAFFE and node: MorganDoocy
There are 3 edges between node: PS and node: MICHAELH
There are 2 edges between node: LORELLE and node: BrooksRimes
There are 3 edges between node: TorBjornrud and node: PODZ
There are 7 edges between node: PODZ and node: SKIPPY
There are 2 edges between node: virgilhuston and node: PODZ
There are 3 edges between node: ChrisWaigl and node: MorganDoocy
There are 3 edges between node: JosiahRitchie and node: MatthewThomas
There are 21 edges between node: CARTHIK and node: CRAIG
There are 5 edges between node: CRAIG and node: MatthewThomas
There are 6 edges between node: OwenWinkler and node: MATT
There are 5 edges between node: Cena and node: CARTHIK
There are 2 edges between node: ChrisCoggburn and node: CRAIG
There are 4 edges between node: MDAWAFFE and node: MDAWAFFE
There are 2 edges between node: JoeyB and node: MATT
There are 3 edges between node: SKIPPY and node: LORELLE
There are 4 edges between node: RyanDuff and node: MATT
There are 2 edges between node: Moshu and node: MATT
There are 3 edges between node: OwenWinkler and node: ChrisWaigl
There are 2 edges between node: JenniferHodgdon and node: AndrewRivett
There are 2 edges between node: ChristianBalzer and node: Cena
There are 3 edges between node: MorganDoocy and node: OwenWinkler
There are 2 edges between node: SKIPPY and node: MatthewThomas
There are 4 edges between node: ArthurJennings and node: PODZ
There are 6 edges between node: PODZ and node: RyanBoren
There are 2 edges between node: PODZ and node: Kitty
There are 4 edges between node: DougalCampbell and node: MATT
There are 5 edges between node: RyanBoren and node: SKIPPY
There are 29 edges between node: PODZ and node: MATT
There are 2 edges between node: CRAIG and node: KAFKAESQUI
There are 3 edges between node: Cena and node: RyanJ.Markel
There are 2 edges between node: MatthewThomas and node: Cena
There are 3 edges between node: SKIPPY and node: CARTHIK
There are 2 edges between node: CARTHIK and node: ChristianBalzer(TigerDE2)
There are 2 edges between node: LORELLE and node: JeremyVisser
There are 18 edges between node: CRAIG and node: CARTHIK
There are 2 edges between node: JenniferHodgdon and node: LloydBudd
There are 9 edges between node: PODZ and node: MICHAELH
There are 2 edges between node: MATT and node: DavidEads
There are 6 edges between node: MichaelB and node: MICHAELH
There are 6 edges between node: MorganDoocy and node: MATT
There are 4 edges between node: JenniferHodgdon and node: MartinSturm
There are 5 edges between node: CARTHIK and node: MatthewThomas
There are 3 edges between node: CARTHIK and node: Moshu
There are 3 edges between node: LloydBudd and node: JenniferHodgdon
There are 3 edges between node: KAFKAESQUI and node: MATT
There are 3 edges between node: PODZ and node: ChrisWaigl
There are 3 edges between node: Kitty and node: PODZ
There are 5 edges between node: CARTHIK and node: JamesHuff
There are 26 edges between node: CARTHIK and node: PODZ
There are 7 edges between node: LORELLE and node: SKIPPY
There are 2 edges between node: PODZ and node: JohnWatson
There are 26 edges between node: CARTHIK and node: MATT

There are 2 edges between node: ArthurJennings and node: AaronBrazell
There are 3 edges between node: PODZ and node: Vogel,Andrew(vogelap)
There are 7 edges between node: LORELLE and node: JenniferHodgdon
There are 8 edges between node: MDAWAFFE and node: OwenWinkler
There are 2 edges between node: CRAIG and node: Me
There are 9 edges between node: MatthewThomas and node: CARTHIK
There are 4 edges between node: KSuominen and node: LORELLE
There are 6 edges between node: CARTHIK and node: Cena
There are 9 edges between node: MorganDoocy and node: CRAIG
There are 2 edges between node: PODZ and node: vwebmaster
There are 2 edges between node: WESTI and node: MICHAELH
There are 7 edges between node: SKIPPY and node: MATT
There are 2 edges between node: SKIPPY and node: Vogel,Andrew(vogelap)
There are 2 edges between node: PODZ and node: DavidHouse
There are 7 edges between node: ChristianBalzer and node: MATT
There are 4 edges between node: MichaelB and node: CARTHIK
There are 13 edges between node: PODZ and node: OwenWinkler
There are 4 edges between node: MatthewThomas and node: PODZ
There are 2 edges between node: MATT and node: GeofF.Morris
There are 3 edges between node: CARTHIK and node: ChrisWaigl
There are 2 edges between node: RyanBoren and node: MICHAELH
There are 3 edges between node: MichaelB and node: PODZ
There are 4 edges between node: SKIPPY and node: RyanBoren
There are 3 edges between node: MatthewThomas and node: **CRAIG**
There are 31 edges between node: MATT and node: CRAIG
There are 4 edges between node: RyanBoren and node: MATT
There are 2 edges between node: AnthonyTimberlake and node: PODZ
There are 2 edges between node: PODZ and node: TorBjornrud
There are 3 edges between node: RobertDeaton and node: MATT
There are 14 edges between node: PODZ and node: LORELLE
There are 21 edges between node: MATT and node: LORELLE
There are 3 edges between node: MichaelRenzmann and node: CRAIG
There are 6 edges between node: KAFKAESQUI and node: LORELLE
There are 2 edges between node: SKIPPY and node: OwenWinkler
There are 17 edges between node: CRAIG and node: PODZ
There are 13 edges between node: Cena and node: CRAIG
There are 6 edges between node: OwenWinkler and node: LORELLE
There are 3 edges between node: CRAIG and node: MorganDoocy
There are 2 edges between node: MichaelB and node: TechGnome
There are 4 edges between node: MichaelB and node: LORELLE
There are 6 edges between node: LORELLE and node: CARTHIK
There are 3 edges between node: CARTHIK and node: KSuominen
There are 10 edges between node: Cena and node: MATT
There are 9 edges between node: MATT and node: MICHAELH
There are 2 edges between node: ChrisLott and node: MICHAELH
There are 2 edges between node: miahz and node: ChrisWaigl
There are 2 edges between node: LORELLE and node: DavidEads
There are 2 edges between node: RyanDuff and node: SKIPPY
There are 4 edges between node: JosiahRitchie and node: CRAIG
There are 4 edges between node: AaronBrazell and node: RobertDeaton
There are 2 edges between node: OwenWinkler and node: lokrin
There are 2 edges between node: Moshu and node: virgilhuston
There are 2 edges between node: MatthewThomas and node: LORELLE
There are 2 edges between node: MatthewThomas and node: MATT
There are 2 edges between node: ChrisWaigl and node: OwenWinkler
There are 24 edges between node: CARTHIK and node: LORELLE

Annexe 7: “codebook” correspondant à la liste de discussion des hackers

Contenu	Répertoire de plugin	22	Qualification négative	Info pas trouvée	46		Codex utile	23
	Répertoire de plugin/Demande de vérification du contenu.	1		Info inexistante	3		Codex très bon outil	6
	Répertoire de plugin/ Demande d'ajout de contenu.	0		Info incomplète	17		Exemplarité du Codex	4
		0		Info fausse	13			
Formes d'organisation observées	Régulation des tâches	23		Info pas claire/imprécise	13		Codex centralise information	13
	Régulation des comportements	10			92		Autre forme de documentation suggérée	31
	Régulation des comportements/Code	36	Qualification neutre	Info en cours de recherche	3		Autre forme de régulation de la programmation	1
	Régulation des comportements/Forum	1		Info mise en doute	16		Documentation des plugins	1
	Répartition des rôles	9						
	Délégation à Docs	15	Qualification positive	Info trouvée	3	Conséquences de la citation	Modification du Codex effective	21
	Délégation aux hackers	6		Info bonne	5		Modification du Codex promise par l'auteur du mail	28
	Plainte pour négligence des « tâches humbles »	12		Info qui fait autorité	39		Modification du Codex commentée	1
	Pas de négligence des tâches humbles	4			44		Demande de modification du Codex	46
	Plainte pour excès d'exigence des utilisateurs	2	Remarque générale sur le Codex	Codex difficile à cause de l'anglais	0		Demande de création de page	13
	Réputation/Reconnaissance acquise par les contributions au Codex	4		Codex inaccessible	4		Création explicite de page au Codex	2
	Codex mémoire de la coordination et des échanges	9		Codex inaccessible pour cause de panne	5			
	Aucune organisation assumée	2		Codex difficilement accessible pour cause de ralentissement	1		Demande de vérification du contenu	6
	Désordre/organisation distribuée	5		Codex où information difficilement accessible à cause moteur de recherche	4		Demande de vérification du contenu/Architecture du Codex	1
	Communauté assumée	3		Codex : problème à cause blanchissage (etc.)	1		Demande de vérification du contenu/Expertise	3
	Proposition d'aide	3		Moteur wiki	13		Citation pour donner information technique	41
	Documentation technique à expertise reconnnue	3						
	Principe wiki bon	6		Panne disparue	1		Régulation des tâches	23
	Common knowledge	2					Régulation des tâches/Coordination d'événements	0
	Modèle de Wikipedia	2		Codex mauvais	12		Répertoire de plugin	22
	Modèle du Manuel de PhP	1		Codex mauvais/Mauvais formatage de l'information	2			
	Articulation wiki/reste communauté	14		Codex mauvais/Problème d'architecture de l'information	3		Read The Fucking Manual	9
	Catégorisation et langage commun	1		Codex mauvais/Problème de clarté de l'information	0		Use Google	2
				Codex mauvais/Toutes les catégories	0		Spam	2
	Info personnelle dont on fait la publicité.	21		Codex à améliorer	35		Référencement wiki	1
	Info d'autrui dont on fait la publicité	7		Codex imprécis	0		Disparition ML-Doc	2
				Plainte pour négligence des « tâches humbles »	12		Moteur de recherche	2
	Connotation débutant	8		Codex ignoré par les contributeurs notables	3		Codex outil de publicité	2
	Connotation expert	13		Codex mal utilisé	4			

Nous n’avons placé que le « codebook » de la liste la plus importante.

Annexe 8: “codebook” du codage de la liste de discussion du forum

Contenu	Répertoire de plugin	3	Qualification négative	Info pas trouvée	4		Codex utile	12
	Répertoire de plugin/ Demande d'ajout de contenu.	0		Info inexistante	2		Codex très bon outil	5
				Info incomplète	2		Exemplarité du Codex	2
Formes d'organisation observées	Régulation des tâches	4		Info fausse	1			
	Régulation des comportements	10		Info pas claire/imprécise	3		Codex utilisé par les utilisateurs	2
	Régulation des comportements/Forum	13			12			
	Délégation à Docs	1	Qualification neutre	Info en cours de recherche	1		Codex centralise information	6
	Délégation aux hackers	1		Info mise en doute	3		Autre forme de documentation suggérée	6
	Plainte pour négligence des « tâches humbles »	0						
	Réputation/Reconnaissance acquise par les contributions au Codex	6	Qualification positive	Info trouvée	1	Conséquences de la citation	Modification du Codex effective	1
	Codex mémoire de la coordination et des échanges	1		Info bonne	2		Modification du Codex effective/Mise à jour du Codex	0
	Aucune organisation assumée	1					Modification du Codex promise par l'auteur du mail	2
	Désordre<organisation distribuée	2	Remarque générale sur le Codex	Codex difficile à cause de l'anglais	1		Modification du Codex commentée	0
	Communauté assumée	2		Codex inaccessible	1		Demande de modification du Codex	0
				Codex inaccessible pour cause de panne	13		Demande de modification du Codex/Délégation à Docs	0
	Info personnelle dont on fait la publicité	13		Codex difficilement accessible pour cause de ralentissement	0		Demande de modification du Codex /Mise à jour du Codex	2
	Info d'autrui dont on fait la publicité	1		Codex où information difficilement accessible à cause moteur de recherche	5		Modification du code suggérée	0
				Codex : problème à cause blanchissage (etc.)	3		Demande de création de page	0
	Forum-Codex/Opposition	3			23		Création explicite de page au Codex	0
	Forum-Codex/Complémentarité	17		Panne disparue	6			0
	Forum-Codex/Panne liens wiki	0					Standardisation des représentations	0
	Forum-Codex/AideACoordination	4		Codex mauvais	5			
	CColléLiensInsuffisant	5		Codex mauvais/Mauvais formatage de l'information	1		Régulation des tâches	4
	Forum-Codex/Juxtaposition	1		Codex mauvais/Problème d'architecture de l'information	5			0
	Forum-Codex/MauvaiseCoordination	1		Codex mauvais/Problème de clarté de l'information	1	Autre/Secondaire	Référence commune	0
	Arbitrage insertion info vers le Codex	4		Codex à améliorer	10		Problème technique sur le Codex	0
	Arbitrage information reste dans le Forum	3		Codex imprécis	0		RFTM	2
	Forum-Codex/Travail commun des équipes	1		Codex ignoré par les utilisateurs	10		Vague	0
				Codex ignoré par les contributeurs notables	0		Use Google	2
	Connotation débutant	11		Codex mal utilisé	1		Codex dupliqué par d'autres sites	2
	Connotation expert	22		Codex accessible par moteur de recherche interne	1			

Annexe 9 : tableau de quelques régressions du chapitre 7

Nous avons effectué plusieurs régressions au chapitre 7, mais il était impossible de placer en annexe ici tous les tableaux de chacune d'elles, ce qui n'a pas toujours intéressant étant donné que certaines, avec des R2 particulièrement forts et ressemblant quasiment à une droite, se suffisent à elles-mêmes. Nous avons choisi de n'en présenter ici que quelques-unes d'importance ou présentant des imperfections. Le fichier complet est disponible au besoin.

Régression de la partie 7.6.2.

Variable	Observations	Obs. avec données manquant	Obs. sans données manquante	Minimum	Maximum	Moyenne	Ecart-type
0,6286	50	0	50	0,175	1,000	0,508	0,210
0,5455	50	0	50	0,000	1,000	0,478	0,341

Coefficients d'ajustement :

Observations	50,000
Somme des poids	50,000
DDL	48,000
R^2	0,779
R^2 ajusté	0,775
MCE	0,010
RMCE	0,100
MAPE	16,072
DW	2,021
Cp	2,000
AIC	-228,394
SBC	-224,570
PC	0,239

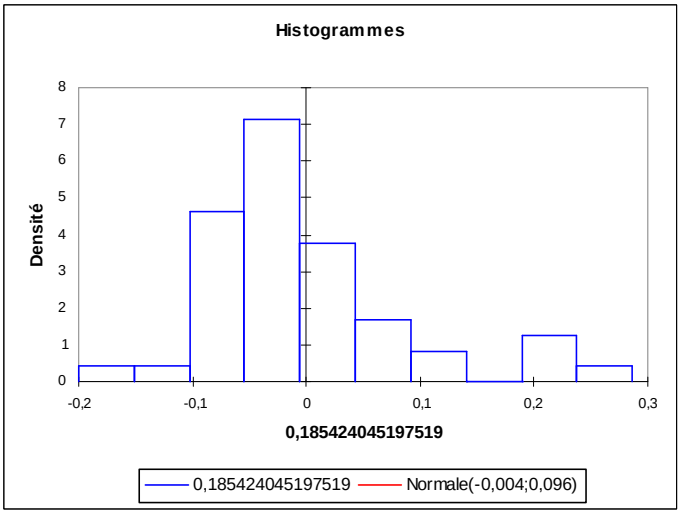
Analyse de la variance :

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	1	1,691	1,691	169,431	< 0,0001
Erreur	48	0,479	0,010		
Total corrigé	49	2,170			

Source	Valeur	Ecart-type	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
Constante	0,769	0,024	31,393	< 0,0001	0,719	0,818
0,5455	-0,545	0,042	-13,017	< 0,0001	-0,629	-0,461

Coefficients normalisés :

Source	Valeur	Ecart-type	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
0,5455	-0,883	0,068	-13,017	< 0,0001	-1,019	-0,746


Histogrammes
Densité
8
7
6
5
4
3
2
1
0
-0,2
-0,1
0
0,1
0,2
0,3
0,185424045197519
0,185424045197519
Normale(-0,004;0,096)

Régression de la partie 7.3.2.

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
6	58	0	58	0,000	28,000	7,897	6,106
2	58	0	58	0,000	8,000	3,638	2,222

Coefficients d'ajustement :

Observations	58,000
Somme des poids	58,000
DDL	56,000
R^2	0,665
R^2 ajusté	0,659
MCE	12,700
RMCE	3,564
MAPE	59,853
DW	1,920
Cp	2,000
AIC	149,378
SBC	153,499
PC	0,359

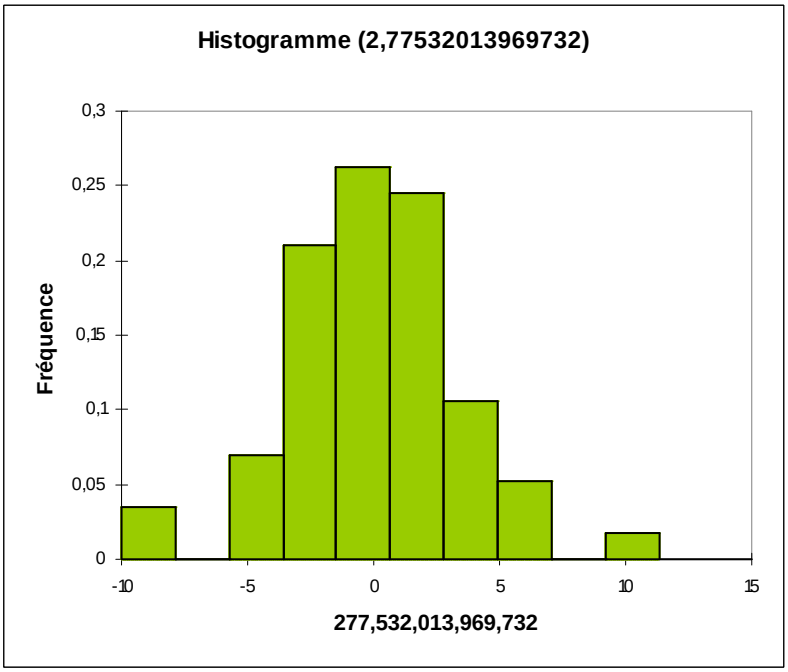
Analyse de la variance :

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	1	1414,173	1414,173	111,351	< 0,0001
Erreur	56	711,206	12,700		
Total corrigé	57	2125,379			

Source	Valeur	Ecart-type	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
Constante	-0,259	0,903	-0,287	0,776	-2,069	1,551
2	2,242	0,212	10,552	< 0,0001	1,816	2,667

Coefficients normalisés :

Source	Valeur	Ecart-type	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
2	0,816	0,077	10,552	< 0,0001	0,661	0,971


Histogramme (2,77532013969732)
Fréquence
0,3
0,25
0,2
0,15
0,1
0,05
0
-10
-5
0
5
10
15
277,532,013,969,732

Régression de la partie 7.5.1. (figure 61)

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
22	57	0	57	1,000	116,000	19,702	21,321
13	57	0	57	0,000	231,000	30,070	42,100

Coefficients d'ajustement :

Observations	57,000
Somme des poids	57,000
DDL	55,000
R^2	0,736
R^2 ajusté	0,731
MCE	122,282
RMCE	11,058
MAPE	91,275
DW	1,706
Cp	2,000
AIC	275,925
SBC	280,011
PC	0,283

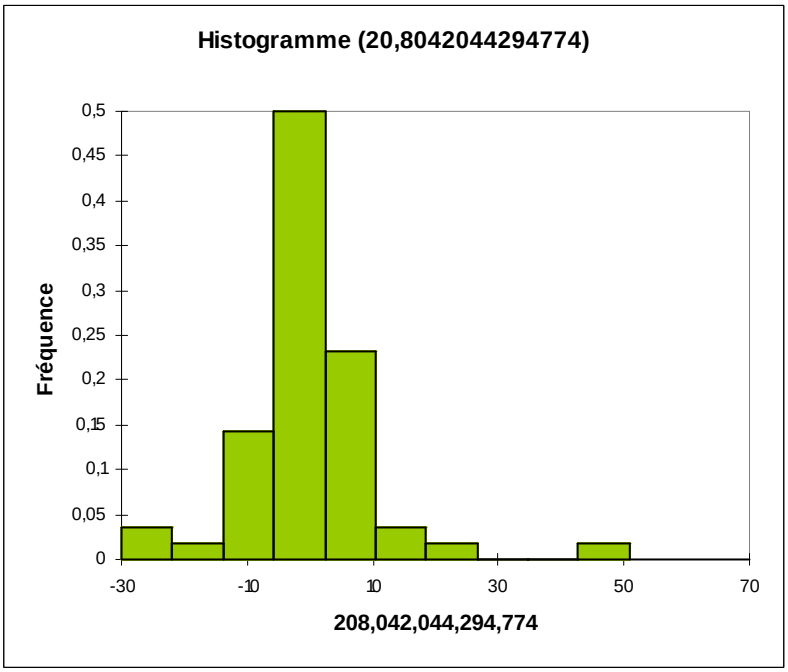
Analyse de la variance :

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	1	18732,445	18732,445	153,191	< 0,0001
Erreur	55	6725,485	122,282		
Total corrigé	56	25457,930			

Source	Valeur	Ecart-type	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
Constante	6,638	1,805	3,677	0,001	3,020	10,256
13	0,434	0,035	12,377	< 0,0001	0,364	0,505

Coefficients normalisés :

Source	Valeur	Ecart-type	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
13	0,858	0,069	12,377	< 0,0001	0,719	0,997


Histogramme (20,8042044294774)
Fréquence
0,5
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
-30
-10
10
30
50
70
208,042,044,294,774

Régression de la partie 7.6.1. (figure 77)

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
1,157452789	57	0	57	0,000	2,251	1,139	0,476
2,079441542	57	0	57	0,000	3,584	2,170	0,872

Coefficients d'ajustement :

Observations	57,000
Somme des poids	57,000
DDL	55,000
R^2	0,372
R^2 ajusté	0,361
MCE	0,145
RMCE	0,381
MAPE	27,734
DW	1,802
Cp	2,000
AIC	-108,144
SBC	-104,058
PC	0,673
Press RMCE	0,394

Analyse de la variance :

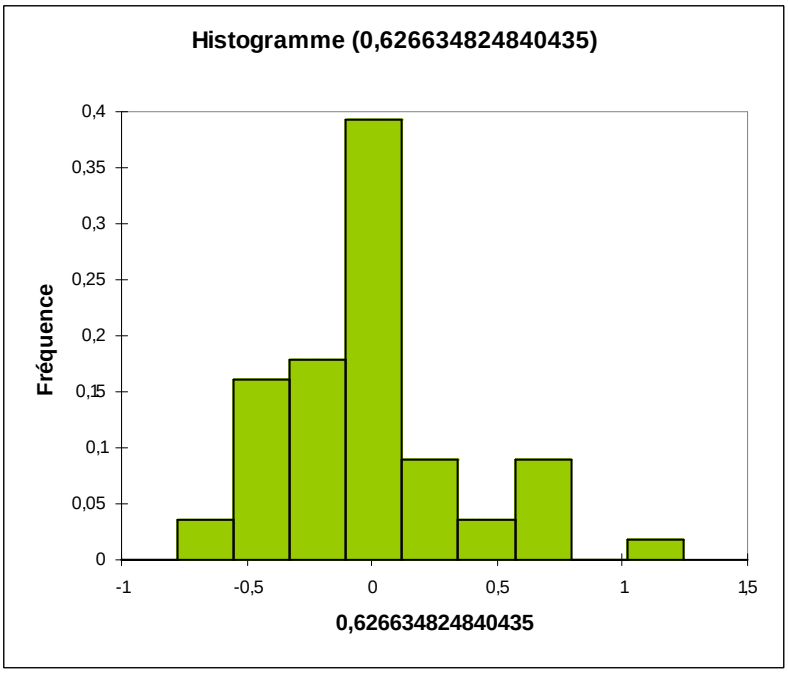
Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	1	4,725	4,725	32,609	< 0,0001
Erreur	55	7,969	0,145		
Total corrigé	56	12,694			

Paramètres du modèle :

Source	Valeur	Ecart-type	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
Constante	0,416	0,136	3,053	0,003	0,143	0,689
2,079441542	0,333	0,058	5,710	< 0,0001	0,216	0,450

Coefficients normalisés :

Source	Valeur	Ecart-type	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
############	0,610	0,107	5,710	< 0,0001	0,396	0,824


Histogramme (0,626634824840435)
Fréquence
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
-1
-0,5
0
0,5
1
1,5
0,626634824840435

Régression de la partie 7.6.1. (figure 82)

Statistiques descriptives :

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
1,157452789	57	0	57	0,000	2,251	1,139	0,476
3,555348061	57	0	57	0,000	5,793	3,168	1,379

Coefficients d'ajustement :

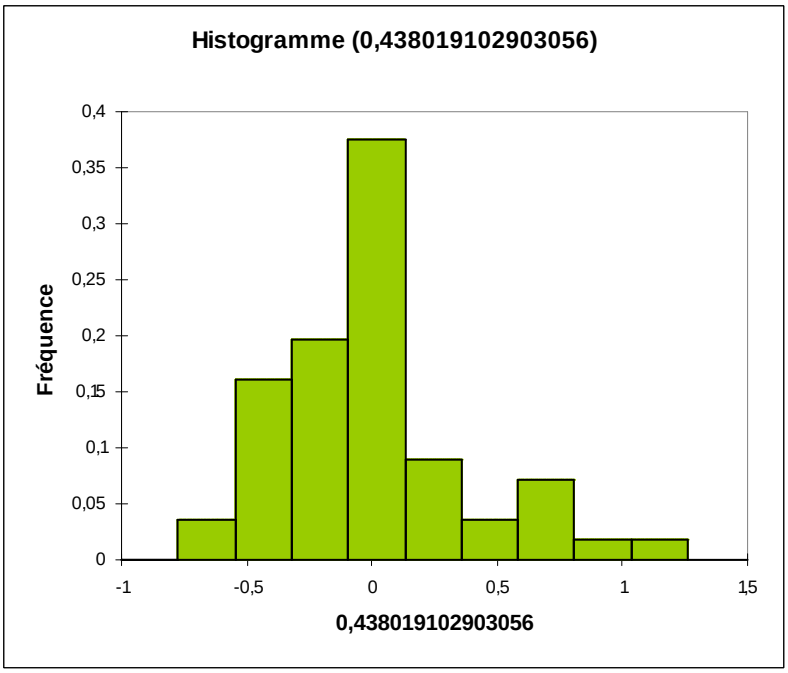
Observations	57,000
Somme des poids	57,000
DDL	55,000
R^2	0,397
R^2 ajusté	0,386
MCE	0,139
RMCE	0,373
MAPE	25,927
DW	1,800
Cp	2,000
AIC	-110,400
SBC	-106,314
PC	0,647

Analyse de la variance :

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	1	5,034	5,034	36,146	< 0,0001
Erreur	55	7,660	0,139		
Total corrigé	56	12,694			

Paramètres du modèle :

Source	Valeur	Ecart-type	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
Constante	0,450	0,125	3,608	0,001	0,200	0,700
3,555348061	0,217	0,036	6,012	< 0,0001	0,145	0,290


Histogramme (0,438019102903056)
Fréquence
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
-1
-0,5
0
0,5
1
1,5
0,438019102903056

Régression de la partie 7.6.2. (figure 73)

Statistiques descriptives :

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
2,595067437	57	0	57	0,000	15,009	3,199	3,317
8	57	0	57	1,000	36,000	11,737	7,943

Coefficients d'ajustement :

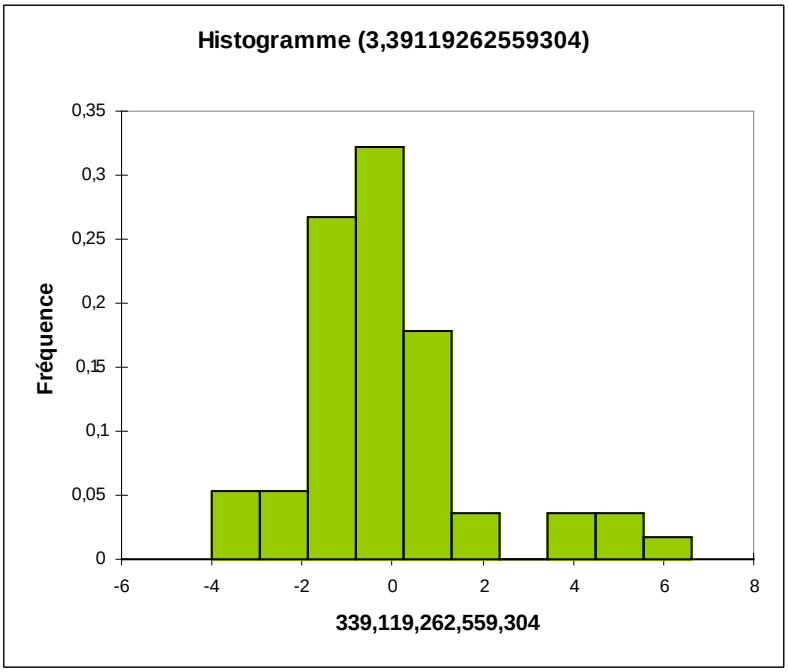
Observations	57,000
Somme des poids	57,000
DDL	55,000
R^2	0,637
R^2 ajusté	0,630
MCE	4,070
RMCE	2,017
MAPE	65,003
DW	1,092
Cp	2,000
AIC	81,974
SBC	86,060
PC	0,390

Analyse de la variance :

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	1	392,212	392,212	96,363	< 0,0001
Erreur	55	223,859	4,070		
Total corrigé	56	616,072			

Paramètres du modèle :

Source	Valeur	Ecart-type	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
Constante	-0,711	0,480	-1,483	0,144	-1,673	0,250
8	0,333	0,034	9,816	< 0,0001	0,265	0,401


Histogramme (3,39119262559304)
Fréquence
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
-6
-4
-2
0
2
4
6
8
339,119,262,559,304

Bibliographie

Cette bibliographie ne correspond pas à tout ce que nous avons pu consulter ou lire durant notre travail de thèse. L'évolution de notre travail ainsi que le caractère inégal de la littérature sur les wikis par exemple ont conduit à rejeter un nombre conséquent de textes. Nous avons divisé notre bibliographie en plusieurs sections au caractère distinct marqué.

Ouvrages et articles généraux, en particulier en management des systèmes d'information et théories des organisations

BUREAU Sylvain (2007): "*La diffusion des technologies : une approche par la professionnalisation. Le cas des technologies web*", thèse de doctorat de l'École Polytechnique en Économie et Sciences Sociales, spécialité Gestion, Paris.

CORIAT Benjamin et WEINSTEIN Olivier (1995), *Les nouvelles théories de l'entreprise*, Le Livre de Poche, 218 p. .

DOSSE François (1997), *L'empire du sens*, La découverte, Paris, 432 p. .

GIRIN Jacques (1990), "L'analyse empirique des situations de gestion : élément de théorie et de méthode", in A.-C. MARTINET (coord.), *Epistémologie et Sciences de Gestion*, Paris, Economica, pp.141-182.

GUNNLAUGSDOTTIR Johanna (2003), Seek and you will find, share and you will benefit: organizing knowledge using groupware systems, *International Journal of Information and management*, 23, pp. 363-380.

HAMERI Ari-Pekka et PUITTINEN Rainer (2003), WWW-enabled knowledge management for distributed engineering projects, *Computers in Industry*, 50, pp. 165-177.

LAUDON et LAUDON (2006), *Management des systèmes d'information*, Pearson Education, 638 p.

SAUSSOIS Jean-Michel (2007), *Théories des organisations*, Repères, Paris, 120 p.

SCHNEIDER Henrik (2005), "Rapid ICT Change and Workplace Knowledge Obsolescence". Berkman Center Research Publication No. 2005-04, Harvard. SSRN: http://ssrn.com/abstract=870084.

VON HIPPEL Eric (2005), *Democratizing Innovation*, MIT Press, Cambridge, 204 p.

WAHLSTROM Billie et SRUTON Chris (1997), "Constructing texts/understanding texts : Lessons from antiquity and the middle ages", *Computers and Composition*, 14-3, pp. 311-328.

WEILL Peter et ROSS Jeanne W., (2004), "IT Governance on One Page". MIT Sloan Working Paper No. 4517-04. http://ssrn.com/abstract=664612

Ouvrages et articles de méthodologie, et théories mobilisées

Ouvrages et articles traitant d'histoire, d'enquête, méthodes qualitatives et quantitatives

ALEXY Oliver (2008), "On the Double Importance of Time in IT-Related Event Studies", Working Paper. SSRN: http://ssrn.com/abstract=1029592

BANISTER Frank (2002), "The Dimension of Time: Historiography in Information System Research", *Electronic Journal of Business Research Methods*, 1-1, pp. 1-10.

BEAUD Stéphane et WEBER Florence, *Guide de l'enquête-terrain*, La Découverte, Paris, 2003, 357 p.

BLOCH Marc (1993), *Le métier d'historien*, Armand Colin, Paris, 158 p.

BORZEIX Anni, COLLARD Damien, RAULET-CROSET Nathalie (2006) « Participation, insécurité, civilité : quand les habitants s'en mêlent », *Les Cahiers de la sécurité*, 61-2, pp. 55-84.

BORZEIX Anni (2007), « Jeux d'échelle », *Le Libellio d'AEGIS*, 3-2, pp. 25-28. http://www.crg.polytechnique.fr/v2/fic/LeLibellio_Printemps2007.pdf

BOULLIER Dominique (1989), « Archéologie des messageries », *Réseaux*, 36, pp. 9-29.

DHERENT Catherine (2002), *Les archives électroniques, manuel pratique*, La Documentation Française, Paris, 104 p. .

CAMPBELL-KELLY Martin et GARCIA-SWARTZ Daniel D. (2005), « The history of the Internet : The Missing Narratives », Working Paper. SSRN: http://ssrn.com/abstract=867087

DEGENNE Alain, FORSE Michel (1994), *Les réseaux sociaux*, Armand Colin, Paris, 263 p. .

DUMEZ Hervé (2006), « L'analyse de données qualitatives assistée par ordinateur (ADQAO), », *Le Libellio d'AEGIS*, 2-3, pp. 36-40.

DUMEZ Hervé (2008), « L'ethnographie virtuelle : reconstituer le contexte des interactions en ligne », *Le Libellio d'AEGIS*, 4-1, pp. 39-43.

DUMEZ Hervé et JEUNEMAITRE Alain (2005), « La démarche narrative en économie », *Revue économique*, 56-4, pp 983-1006.

DUMEZ Hervé et JEUNEMAITRE Alain (2006) "Reviving narratives in economics and management : towards an integrated perspective of modelling, statistical inference and narratives.", *European Management Review*, 3-1, pp. 32-43.

DUMEZ Hervé (1999), « Mes critères d'évaluation d'une thèse en gestion sur étude de cas », Working Paper CRG.

DUMEZ Hervé (2009), « Comment peut-on avoir des idées ? », *Le Libellio d'AEGIS*, 5-1, pp. 1-10.

FLINN Andrew (2007), "Community Histories, Community Archives: some opportunities and Challenges", *Journal of the Society of Archivists*, 28-2, pp 151-176.

FRIDENSON Patrick (2006), « Une entreprise de sélection : Google », *Entreprises et Histoire*, 43, p. 47-57.

GEORGE Alexander L. et BENNETT Andrew (2005), *Case studies and theory development in the social sciences*, MIT Press, Cambridge, 331 p.

GUERREAU Alain (2004), « Statistique pour historiens », Document de travail Ecole des Chartes, http://elec.enc.sorbonne.fr/statistiques/stat2004.pdf (consulté le 01/01/09).

HINE Christine (2000), *Virtual Ethnography*, Sage Publishers, Londres, 192 p. .

HOLLAN James, HUTCHINS Edwin, et KIRSH David (2000), Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research, *ACM Transactions on Computer-Human Interaction*, 7-2, pp. 174–196.

KIESER Alfred (1994), "Why Organization Theory Needs Historical Analysis – And How This Should Be Performed", *Organization Science*, 5-4, pp. 608-620.

LANGNER Irene (1997), "An introduction to Internet Mailing-list Research", Working Paper.

LEMERCIER Claire et ZALC Claire (2008), *Méthodes quantitatives pour l'historien*, La Découverte, Paris, 120 p. .

MATON, Eric (2006), « Une analyse des diagrammes présents dans les rapports annuels des grandes entreprises françaises en 1980, 1990 et 2001 », *Entreprises et Histoire*, 44, pp. 66-86.

MILES Matthew et HUBERMAN A. Michael (1994), *Qualitative Data Analysis*, Sage Publications, Thousand Oaks, 338 p.

MONOD Emmanuel (2002), « Epistémologie de la recherche en système d'information », in ROWE F. (2002), *Faire de la recherche en systèmes d'information*, Vuibert, Paris, .

PASSERON Jean-Claude et REVEL Jacques (2005), *Penser par cas*, Editions de l'EHESS, Paris, 291 p.

POINT Sébastien et VOYNNET-FOURBOUL Catherine (2006), « Le codage à visée théorique », *Recherche et Applications en Marketing*, 21-4, pp. 61-78.

RAGIN Charles and BECKER Howard S. (1992), *What is a Case? Exploring the Foundations of Social Inquiry*. Cambridge University Press, 256 p. .

RUZE Emmanuel (2009), « Les archives de la Toile. Le cas de la communauté WordPress (2003-2008) », *Entreprises et Histoire*, 55, en cours d'impression.

RYGIEL Philippe, « Des archives numériques sans historiens ? Un point de vue », *Matériaux*, 79, 2005, pp 11-13 .

ROBLES Gregorio, GONZÁLEZ-BARAHONA J. M., et HERRAIZ I. (2005). "An empirical approach to Software Archaeology", in *Proceedings of the International Conference on Software Maintenance (poster session).* http://opensource.mit.edu/papers/robles-koch-barahona_cvsanaly.pdf

SIMON Herbert (1999), "The many shapes of knowledge", *Revue d'économie industrielle*, 88, pp. 23-39.

THIETART Raymond-Alain (2003), *Méthodes de recherche en Management*, Dunod, Paris, 537 p.

DE VAUJANY François-Xavier (2006), "Between eternity and actualization: the co-evolution of the fields of communication in the Vatican," Munich Personal RePEc Archive Paper 4082.

SUDDABY Roy (2006), "What Grounded Theory is Not", *Academy of Management Journal*, 49-4, pp. 633–642.

BHOWMICK Tanuka (2006), "Building an Exploratory Visual Analysis Tool for Qualitative Researchers", Working Paper.http://www.geovista.psu.edu/publications/2006/Bhowmick_AutoCarto_QualRes_06.pdf :

ULICHNY Polly (1991), « The role of Hypothesis Testing in Qualitative Research. A Researcher Comments", *TESOL Quarterly*, 25-1, pp. 200-202.

VEYNE Paul (1971), *Comment on écrit l'histoire*, Seuil, Paris, 438 p. .

YIN Robert K. (2003), *Case Study Research, Design and Methods*, Sage Publications, Thousand Oaks, 181 p.

YAMADA et al. (2005), « Clustering Interaction Histories in Support of Collaboration », Proceedings of the HCI International Conference (HCII), Las Vegas, CD-ROM.

Articles et ouvrages concernant les théories mobilisées

BRIAN Arthur W. (1994), *Increasing Returns and Path Dependence in the Economy*, University of Michigan Press, Ann Arbor, 224 p.

BRIAN Arthur W. (1994), “Inductive Reasoning and Bounded Rationality”, *American Economic Review*, 84, pp. 406–411.

BATIFOULIER Philippe (2001), *Théorie des Conventions*, Economica, Paris, 328 p.

BOLTANSKI Luc, et THEVENOT Laurent (1991), *De la justification. Les économies de la grandeur*, Gallimard, Paris, 483 p .

BOURGINE Paul et NADAL Jean-Pierre (2004), *Economie cognitive*, Springer, Berlin, 479 p.

BROUSSEAU Eric and PÉNARD Thierry (2007), “The Economics of Digital Business Models: A Framework for Analyzing the Economics of Platforms”, *Review of Network Economics*, 6-2, pp. 81-114.

COHEN et al. (1995), “Routines and other recurring action patterns in organizations. Contemporary research issues”, Santa Fe Working Paper.

DAVID Paul (1985), “Clio and the Economics of QWERTY”, *American Economic Review*, 75-2, pp. 332-37.

DE BESSES Marie (1999), « Les logiciels libres, un système de circulation fertile en apprentissage, *Terminal*, n°80-81, , pp 257-278.

DESSALLES, Jean-Louis (2006), “Generalised signalling: a possible solution to the paradox of language”, in Cangelosi A., Smith A. D. M. et Smith K. (eds), *The evolution of language*, World Scientific, Singapore, pp 75-82.

DESSALLES Jean-Louis (2007), *Why We Talk. The Evolutionary Origins of Language*, Oxford University Press, Oxford, 400 pp.

DUMEZ Hervé (2004), *Gouverner les organisations*, L’Harmattan, Paris, 439 p.

DOUGLAS Mary (1999), *Comment pensent les institutions?*, La Découverte, Paris, 180p.

EYMARD-DUVERNAY François (2004), *L’économie politique de l’entreprise*, La Découverte, Paris, 120p.

DUBUISSON (1998), « Regard d'un sociologue sur la notion de routine dans la théorie évolutionniste : A quoi sert la notion de routine ? », *Sociologie du travail* , 40-4, pp. 491-502.

DUMEZ Hervé (2005), « Equifinalité, étude de cas et modèle de l’enquête », *Le Libellio d’AEGIS*, 2-1, pp. 18-21.

FAVEREAU Olivier (2001), “Theory of Information: from bounded rationality to interpretative reason”, in PETIT Pascal (2001), *Economics and Information*, Kluwer, Dordrecht, pp 93-120.

FAVEREAU Olivier et LAZEGA Emmanuel (2002), *Conventions and structures in economic organization. Markets, networks and hierarchies*, Edward Elgar, Cheltenham, 361 p.

FELIN Teppo et FOSS Nicolai J. (2004), “Organizational routines: A skeptical look”, DRUID Working Paper.

FOSS Nicolaj J. et MAHNKE Volker (2003), "Knowledge Management: What Can Organizational Economics Contribute?", DRUID Working Papers 03-02.

GIGERENZER Gerd et SELTEN Reinhard (2002), *Bounded Rationality, The Adaptive Toolbox*, MIT Press, Cambridge, MA, 277 p. .

GOMEZ P.-Y. (1994), *Qualité et économie des conventions*, Economica, Paris, 270 p.

HOLLAN J., Hutchins, E., and Kirsh, D. (2000), “Distributed cognition: toward a new foundation for human-computer interaction research”, *ACM Trans. Comput.-Hum. Interact.* 7-2, pp. 174-196

HORN François (2001), « La diversité de l'économie du logiciel: pluralité et dynamique de quatre mondes de production », *Revue d'économie industrielle*, 95 , pp. 37-60.

HUTCHINS Edwin (1991) "Organizing Work by Adaptation" *Organization Science* 2-1, pp. 14-39.

KECHIDI Med (2005), « Une taxinomie des figures collectives de la coordination de l'action: de la convention à la routine », Cahier du GRES n°2005-7.

LESOURNE Jacques, ORLEAN André et WALLISER Bernard (2002), *Leçons de microéconomie évolutionniste*, Odile Jacob, Paris, 260p.

LORENZ Edward (2001) “Models of cognition, The Contextualisation of Knowledge and Organisational Theory”, *Journal of Management & Governance*, 5-3&4, pp. 307-330.

NOVARESE Marco et RIZZELLO Salvatore (2004). "The intermingling between cognitive economics and experimental economics: a few remarks on history methodology and applications," CESMEP Working Papers 200406.

REYNAUD Bénédicte (2004), *Les Règles économiques et leurs usages***,** Odile Jacob, Paris, 240 p.

RUZE Emmanuel (2004), "Changements organisationnels, implémentation des ERP, et gestion du savoir : une réinterprétation économique.", Colloque "Métamorphose des organisations" de Nancy 2.

SHIN Minsoo (2004), "A framework for evaluating economics of knowledge management systems", *Information and Management*, 42-1, pp. 179 - 196.

SIMON Herbert. A. (1971), "Designing Organizations for an Information-Rich World", in GREENBERGER Martin, *Computers, Communication, and the Public Interest*, The Johns Hopkins Press, Baltimore, 335 pp.

SMITH Eric, BOWLES Samuel, GINTIS Herbert (2000), “Costly Signaling and Cooperation”, Santa Fe Institute Working Paper, Santa Fe.

SUQUET Jean-Baptiste (2007), "Règles et conventions : l’approche économique", Le Libellio d’AEGIS, pp. 40-46.

WARGLIEN Massimo, NARDUZZO Alessandro et ROCCO Elena (1997). "Talking about routines in the field: the emergence of organizational capabiliies in a new cellular phone network company," CEEL Working Papers 9706.

Ouvrages et articles sur les wikis

Articles tirés de revues à comité de lecture, ou chapitres d’ouvrages.

ALBORS J., RAMOS J.-C. et HERVAS J.-L. (2008), “New learning network paradigms: Communities of objectives, crowdsourcing, wikis and open source, *International Journal of Information Management*, 28-3, pp. 194-202.

ANDRUS D. Calvin (2005), “The Wiki and the Blog:Toward a Complex Adaptive Intelligence Community”, *Studies in Intelligence*, 49-3.

BARTON Matthew D. (2005), “The future of rational-critical debate in online public spheres”, *Computers and Composition* 22, pp. 177–190.

BEAN Luan (2005), “Wiki: A speedy new tool to manage projects”, *Journal of Corporate Accounting & Finance*, 16-5, pp 3-8.

BROWN John Seely (2004), “Minding and Mining the periphery”, *Long Range Planning*, 37, pp 143–151.

CAPOCCI A. et al. (2006), “Preferential attachment in the growth of social networks: the case of Wikipedia”, *Physical Review*, 74-3

CARDON Dominique et LEVREL Julien (2009), "La vigilance participative. Une interprétation de la gouvernance de Wikipédia", *Réseaux*, 155, pp. 51-89.

CHALLBORN Carl et REIMANN Teresa (2006), “Wiki products: a comparison”, *International Review of Research in Open and Distance Learning*, 6-2, pp 1-4.

CHENG Shang-Wen, NORD Robert L., STAFFORD Judith A. (2005), “WICSA Wiki WAN Party: capturing experience in software architecture best practices, ACM SIGSOFT Software Engineering Notes archive”, 30-1.

CHESNEY Thomas (2006), “An empirical examination of Wikipedia's credibility”, *First Monday* 11-11.

CIFFOLILLI Andrea (2003), “Phantom authority, self–selective recruitment and retention of members in virtual communities: The case of Wikipedia”, *First Monday*, 8-12.

CUNNINGHAM Ward et LEUF Bo (2001), *The Wiki Way: Collaboration and Sharing on the Internet*, Addison-Wesley, Reading, Massachusetts, 464 p. .

DEN BESTEN Matthijs et DALLE Jean-Michel (2008), "Keep it simple, a Companion for Simple Wikipedia", *Industry & Innovation*, 15-2, pp. 169-178.

DEN BESTEN Matthijs L., ROSSI Alessandro, GAIO Loris, LOUBSER Max et DALLE, Jean-Michel (2008), « Mining for Practices in Community Collections: Finds from Simple Wikipedia", in RUSSO Barbara, *Open Source Development, Communities and Quality*, Springer, Berlin, pp. 105-120.

DUGUID Paul (2006), "Limits of self–organization: Peer production and "laws of quality"" *First Monday*, 11-10.

FROOMKIN Michael (2003), "Habermas@discourse.net:Toward a Critical Theory of Cyberspace", *Harvard Law Review*, 116-3, pp. 751-873.

GLASER Anja et EBERSBACH Markus (2004), "Towards Emancipatory Use of a Medium: The Wiki", *International Journal of Information Ethics*, 2, pp. 1-9.

GOLDSPINK Christopher (2008), "Social Self-Regulation in Computer Mediated Communities, The case of Wikipedia", *International Journal of Agent Technologies and Systems*, 1-1, pp. 19-33.

GORMAN G.E. (2005), "Editorial: Is the wiki concept really so wonderful?", *Online Information Review*, 29-3, pp. 225–226.

GUZDIAL Mark et al. (2001), "Beyond adoption to invention: Teacher-created collaborative activities in higher education", *The Journal of the Learning Sciences*, 10-3, pp. 265-279.

HARRER Andreas et ZEINI Sam (2006), "Evaluation of communication in web-supported learning communities, an analysis with triangulation research design", *International Journal of Web Based Communities*, 2-4, pp. 428-446.

HOLLOWAY Todd et alii (2007), "Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors", *Complexity*, 12-3, pp. 30 - 40.

JONES John (2008) « Patterns of Revision in Online Writing: A Study of Wikipedia's Featured Articles », *Written Communication*, 25 pp. 262 - 289.

KING William R. (2007), "IT Strategy and Innovation: Recent Innovations in Knowledge Management", *Information Systems Management*, 24-1, pp. 91-93.

KOLBITSCH, Josef et MAURER Hermann (2005): "Community Building around Encyclopaedic Knowledge", *Journal of Computing and Information Technology*, 13, pp. 175-190.

KOLBITSCH Josef et MAURER Hermann (2006), “The Transformation of the Web: How Emerging Communities Shape the Information We Consume”, *Journal of Universal Computer Science*, 12-2, pp. 187-213.

KORFIATIS Nikolaos Th. et al. (2006), “Evaluating authoritative sources using social networks: An insight from Wikipedia”, *Online Information Review*, 30-3, pp. 252-262.

LATTEMANN Christoph et Stieglitz Stefan (2005), “Framework for Governance in Open Source Communities”, *Proceedings of the 38th Hawaii International Conference on System Sciences*. http://opensource.mit.edu/papers/Governance-in-OpenSourceProjects.pdf

LEE Maria R. et LAN Yi-Chen (2007), “From Web 2.0 to Conversational Knowledge Management : Toward Collaborative Intelligence”, *Journal of Entrepreneurial Research*, 2-2, p. 47-62.

LESSIG Lawrence (1999),”The Limits in Open Code: Regulatory Standards and the Future of the Net”, *Berkeley Technology Law Journal*, 14, p 759. http://ssrn.com/abstract=184870

LEVREL, Julien (2006). « Wikipedia : Un dispositif médiatique de publics participants », in: *Réseaux*, 138, pp.185-218.

OKOLI Chitu et OHB Wonseok (2007), « Investigating recognition-based performance in an open content community: A social capital perspective », *Information & Management*, 44-3, 240-252.

PEREIRA Carla Sofia et SOARES Antonio Lucas (2007), “Improving the quality of collaboration requirements for information management through social network analysis, *International Journal of Information Management*, 27, pp. 86-103.

PETERSEN Soren Mork (2008), “Loser Generated Content: From Participation to Exploitation”, *First Monday*, 13-3.

PFEIL Ulrike, PANAYIOTIS Zaphiris et ANG Chee Siang (2006), “Cultural Differences in Collaborative Authoring of Wikipedia”, *Journal of Computer-Mediated Communication*, 12-1, pp. 88-113.

QUIGGIN John (2006), “Blogs, wikis and creative innovation”, *International Journal of Cultural Studies*, 9-12, pp. 481 - 496.

SALAUN Jean-Michel (2004), « Chronique inachevée d’une réflexion collective sur le document », *Communication et langages*, n°140, pp. 9-17.

SCHROER Joachim et HERTEL Guido (2009), "Voluntary Engagement in an Open Web-based Encyclopedia: Wikipedians, and Why They Do It", *Media Psychology*, 12-1, pp. 96-120. http://opensource.mit.edu/papers/Schroer_Hertel_Wikipedia_Motivation.pdf

SUNSTEIN Cass R. (2006), *Infotopia: How Many Minds Produce Knowledge*, Olin Working Paper. SSRN: http://ssrn.com/abstract=924249 .

VIEGAS Fernanda B. et al. (2007), « The hidden Order of Wikipedia », in SCHULER D. (Ed.), *Online Communities and Social Computing*, Springer, pp 445-454.

WAGNER Christian et BOLLOJU Narasimha (2005), “Supporting Knowledge Management in Organizations with Conversational Technologies: Discussion Forums, Weblogs, and Wikis”, *Journal of Database Management*, 16-2, pp. 1-8.

WAGNER Christian (2006), “Breaking the Knowledge Acquisition Bottleneck through Conversational Knowledge Management”, *Information Resources Management Journal*, 19-1, pp. 70-83.

WAGNER Christian et MAJCHRZAK Anne (2007), “Enabling Customer-Centricity Using Wikis and the Wiki Way”, *Journal of Management Information Systems*, 23-3, pp. 17 - 43.

WAGNER Christian et PRASARNPHANICH Pattarawan (2007), “Innovating Collaborative Content Creation: The Role of Altruism and Wiki Technology”, *Proceedings of the 40th Hawaii International Conference on System Sciences*.

ZLATIC V., Bozicevic M., Stefancic H., et Domazet M. (2006), “Wikipedias: Collaborative web-based encyclopedias as complex networks”, *Physical Review*, 74 –2.

<u>« Working papers » ayant subi le test d’une conférence ou d’un colloque.</u>

ADLER B. Thomas et al. (2008), “Measuring Author contributions to the Wikipedia”, *Proceedings of Wikisym 08*, Porto.

ADLER B. Thomas et al. (2008), “Assigning Trust to Wikipedia Content”, *Proceedings of Wikisym 08*, Porto.

AGUIAR Ademar et DAVID Gabriel (2005), “WikiWiki Weaving Heterogeneous Software Artifacts”, *Proceedings of WikiSym 2005*, San Diego.

ARAZY Ofer, MORGAN Wayne, PATTERSON Raymond A. (2006), “Wisdom of Crowds: Decentralized Knowledge Construction in Wikipedia”, *16th Annual Workshop on Information Technologies & Systems (WITS) Paper*, Milwaukee.

ARONSSON Lars (2002), “Operation of a Large Scale, General Purpose Wiki Website: Experience from susning.nu's first nine months in service”, *6th International ICCC/IFIP Conference on Electronic Publishing*, Karlovy Vary, Czech Republic, pp. 6–8.

BAC Christian et al. (2005), “Why and how to contribute to libre software when you integrate them into an in-house application ?”, *Proceedings of the 1st International Conference on Open Source Systems*, Genova, Italy.

BELLOMI Francesco et BONATO Roberto (2005), “Network Analysis for Wikipedia”, *Proceedings of "Wikimania 2005*, Frankfurt, Main.

BRANDLE Andreas (2005), “Too Many Cooks Don't Spoil the Broth”, *Wikimania2005*, Frankfurt, Main.

BROUDOUX Evelyne (2002), Outils informatiques d'écriture et de lecture: "nouvelles conditions" au devenir auteur, *Colloque "écritures en ligne, pratiques et communautés"*, , université de Rennes 2.

BURIOL Luciana S. et al. (2006), « Temporal Analysis of the Wikigraph », *IEEE/WIC/ACM International Conference on Web Intelligence*, Hong Kong, pp. 45 – 51.

BUFFA Michel (2006), “Intranet Wikis”, *WWW Conference 2006*, Edinburgh.

CABY-GUILLET Laurence, GUESMI Samy et MALLARD Alexandre (2009), "Wiki professionnel et coopération en réseaux. Une étude exploratoire", *Réseaux*, 154, pp. 195-227

CHAU Thomas et MAURER Frank (2005), “A case study of wiki-based experience repository at a medium-sized software company”, *Proceedings of the 3rd international conference on Knowledge capture*, Banff, Alberta, pp. 185-186.

CHENG Shang-Wen Cheng, NORD Robert L., STAFFORD Judith A. (2005), "WICSA Wiki WAN Party: capturing experience in software architecture best practices", *ACM SIGSOFT Software Engineering Notes archive*, 30-1.

CRAIG Thomas (2005), “Open Source Is Coming: Here's How to Deal With It”, *2005 Pacific Northwest Software Quality Conference*, Portland, Oregon.

CROWSTON Kevin and HOWISON James (2005), "The social structure of Free and Open Source software development", *First Monday*, 10-2.

CROWSTON Kevin and FAGNOT Isabelle (2008), “The motivational arc of massive virtual collaboration”, *International Working Conference on Virtuality and Society: Massive Virtual Communities*, Leuphana.

CUBRANIC Davor, HOLMES Reid, YING Annie T.T., et MURPHY Gail. C. (2003), “Tools for light-weight knowledge sharing in open-source software development”, *3rd Workshop on Open Source Software Engineering*, Portland, Oregon.

DA LIO Elizabeth et al. (2005), “TWiki-based facilitation in a newly formed academic community of practice”, *WikiSym2005*, San Diego.

DANIS Catalina et SINGER David (2008), “A wiki instance in the enterprise: opportunities, concerns and reality”, in *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work*, San Diego.

DEKEL Uri (2007), “A Framework for Studying the Use of Wikis in Knowledge Work Using Client-Side Access Data”, *WikiSym07*, Montreal.

DESILETS Alain et al. (2005), “Are wiki usable?”, *WikiSym’05*, San Diego.

DI LORIO Angelo et ZACCHIROLI Stefano (2006), “Constrained Wiki: an Oxymoron?”, *Wikisym 06*, Odense, Denmark.

DI LORIO Angelo et VITALI Fabio (2005), “Web authoring: a closed case?”, *Proceedings of the 38th Hawaii International Conference on System Sciences*, Hilton Waikoloa Village Island, Hawaï.

EFIMOVA Lilia (2005), “Beyond personal webpublishing: An exploratory study of conversational blogging practices”, *Proceedings of the 38th Hawaii International Conference on System Sciences*, Hilton Waikoloa Village Island, Hawaï.

EMIGH William et HERRING Susan C. (2005), “Collaborative Authoring on the Web: A Genre Analysis of Online Encyclopedias”, *Proceedings of the 38th Hawaii International Conference on System Sciences*, Hilton Waikoloa Village Island, Hawaï.

ETO Kouichirou et al (2005), “Qwik-Web: integrating mailing-list and WikiWikiWeb for group communication”, *WikiSym 2005*, San Diego.

FUCH-KITTOWSKI Frank et KOHLER André (2005), “Wiki Communities in the Context of Work Processes”, *Wikisym 2005*, San Diego.

HASAN Helen et PFAFF Charmaine (2006), “The Wiki: an environment to revolutionise employees' interaction with corporate knowledge”, *Proceedings of the 18th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artefacts and environments*, Sydney, pp. 377-380.

HENRIKSSON Jani, MIKKONEN T., and VADÉN T. (2008), “Experiences of Wiki use in Finnish companies”, in *Proceedings of the 12th international Conference on Entertainment and Media in the Ubiquitous Era*, Tampere, Finland, pp.150-153.

LIH Andrew (2004), “The Foundations of Participatory Journalism and the Wikipedia Project”, *Conference paper for the Association for Education in Journalism and Mass Communications Communication Technology and Policy Division*, Toronto.

LOUBSER Max et DEN BESTEN Mattijs (2008), “Wikipedia Admins and Templates : The organizational Capabilities of a Peer Production Effort”, *European Academy of Management*, Ljubljana et Bled, Slovénie.

FERNANDEZ Alejandro (2004), “Scaki - The Scaffolding Wiki, Proceedings of the Database and Expert Systems Applications”, *15th International Workshop on (DEXA'04)*, Saragosse.

FORTE Andrea et BRUCKMAN Amy (2008), “Scaling consensus: Increasing Decentralization in Wikipedia Governance”, *Proceedings of the 41st Hawaii International Conference on System Sciences*, Waikoloa, Hawaï, 1-10.

GEORGE Andrew (2007), “Avoiding Tragedy in the Wiki-Commons”, *Wikimania 2007*, Tapei.

GIORDANO Richard (2007), “An investigation of the use of a wiki to support knowledge exchange in public health”, *Proceedings of the 2007 international ACM conference on Supporting group work*, Sanibel Island, USA, pp 269-272.

GUNNAR Ristroph (2005), “Case Studies of Online Deliberation: The Debian Project and Wikipedia”, *Online Deliberation 2005 Conference*, Stanford.

HALLORAN T. J. et alii (2005), “BeyondCode: Content Management and the Open Source Development Portal”, *3rd Workshop on Open Source Software Engineering*, Portland.

HANSEN Derek, ACKERMAN Mark S. , RESNIK Paul J., MUNSON Sean (2007), “Virtual community Maintenance with a Collaborative Repository”, *Proceedings of ASIST’07*, Milwaukee.

HASAN Helen et PFAFF Chairmain C. (2006), “Emergent Conversational technologies that are Democratizing Innovation Systems in Organisations: the case of the corporate Wiki”, *Proceedings of the Information Systems Foundations (ISF): Theory, Representation and Reality Conference*, Canberra.

HARRER Andreas (2008), “Visualizing Wiki-Supported Knowledge Building: Co-Evolution of Individual and Collective Knowledge”, *Proceedings of Wikisym 08*, Porto.

HESTER Andrea (2008), “Innovating with organizational wikis: factors facilitating adoption and diffusion of an effective collaborative knowledge management system”, *Proceedings of the 2008 ACM SIGMIS CPR conference on Computer personnel doctoral consortium and research*, Charlottesville, USA, pp 161-163.

JACQUEMIN Bernard et al. (2008), “Managing conflicts between users in Wikipedia”, *Social aspects of the Web 08*, Innsbruck, Austria, pp. 87-99.

KINZLER Daniel (2005), “Wiki-Sense - Mining the Wiki”, *WikiMania 2005*, Frankfurt am Main, Germany.

KITTUR, A., CHI, E. H., PENDLETON, B. A., SUH, B, et MYTKOWICZ, T., “Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie”, *in 25th Annual ACM Conference on Human Factors in Computing Systems (CHI 2007)*, San Jose, USA.

KITTUR Aniket (2007), SUH, B., PENDLETON, B. A., et CHI, E. H. (2007). “He says, she says: conflict and coordination in Wikipedia, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, USA.

KITTUR Aniket et KRAUT Robert (2008), “Harnessing the wisdom of crowds in wikipedia: quality through coordination”, *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, San Diego, USA, pp. 37-46.

KRAMER Mark et al. (2008), “Wiki Trust Metrics based on Phrasal Analysis”, *Proceedings of Wikisym 08*, Porto.

KROGH Georg von, Haefliger Stefan, SpaethMay Sebastian (2003), “Collective Action and Communal Resources in Open Source Software Development: The Case of Freenet”, Working Paper. http://opensource.mit.edu/papers/vonkroghhaefligerspaeth.pdf

KROWNE Aaron et BAZAZ Anil (2004), “Authority Models for Collaborative Authoring”, *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04)*, Big Island, Hawaii.

LAWLER Cormac (2005), “Wikipedia as a learning community: content, conflict and the “common good””, *Wikimania 2005*, Frankfurt am Main, Germany.

MAJCHRZAK Ann et al. (2006), “Corporate Wiki Users: Results of a Survey”, *Wikisym 06*, Odense, Denmark.

MAKICE Kevin (2006), *PoliticWiki: Exploring Communal Politics*, *WikiSym 2006*, Odense, Denmark.

MARTLAND David (2003), “Building effective open source communities using software tools”, *Proceedings of the Workshop on Open Source Software Movements and Communities*, Gothenburg, Sweden.

MOULTHROP Stuard et al. (2002), “Self-assembling hypertexts, weblogs, and wikis”, *Proceedings of the thirteenth ACM conference on Hypertext and hypermedia*, College Park, USA.

MUNSON Sean A. (2008), “Motivating and Enabling Organizational Memory with a Workgroup Wiki”, *Proceedings of Wikisym 08*, Porto.

NEUS Andreas (2001), “Managing Information Quality in Virtual Communities of Practice”, *The 6th International Conference on Information Quality at MIT*, Cambridge, USA.

NEWELL Sue (2005), “Sharing learning through documents:conflicting outcomes”, *Proceedings of the 38th International Conference on System Sciences*, Hilton Waikoloa Village, Hawaii.

NICHOLSON Peter J. (2006), “The changing role of intellectual authority, Notes for remarks to the Association of Research Libraries”, *148th Membership Meeting*, Ottawa.

NUNES Sérgio et al. (2008), “WikiChanges - Exposing Wikipedia Revision Activity”, *Proceedings of Wikisym 08*, Porto.

ORTEGA Felipe et al., (2008), “On the Inequality of Contribution to Wikipedia”, *Proceedings of the Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, Hilton Waikoloa Village, Hawaï, pp 304-314.

ORTEGA Felipe et GONZALEZ-BARAHONA Jesus M. (2007), « Quantitative Analysis of the Wikipedia Community of Users », *WikiSym 07*, Montreal.

PENTZOLD Christian et SEIDENGLANZ Sebastian (2006), “Foucault@Wiki: First Steps Towards a Conceptual Framework for the Analysis of Wiki Discourses”, *WikiSym 2006*, Odense, Denmark.

RADZIWILL Nicole M. et SHELTON Ami L. (2004), “TWiki as a Platform for Collaborative Software Development Management”, *Proceedings of the SPIE*, 5496, pp. 609-617.

RAITMAN Ruth et AUGAR Naomi (2005), “Employing Wikis for Online Collaboration in the E-Learning Environment: Case Study”, *Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05)*, Sydney.

RAYGAN Robert et GREEN David G. (2002), “Internet Collaboration: Twiki”, *SoutheastCon, 2002, Proceedings IEEE*, pp. 137-141.

RAYNES-GOLDIES Kate et FONO David (2005), “Wiki Use by Political Parties: A Case Study”, in *Proceedings of the Online Deliberation Conference 2005*, Stanford California.

ROBBINS (2003) “Adopting Open Source Software Engineering (OSSE) Practices by Adopting OSSE Tools”, in FELLER J., FITZGERALD B., HISSAM S. & LAKHAM K. (Eds.), *Making Sense of the Bazaar: Perspectives on Open Source and Free Software, O’Reilly & Associates*, Sebastopol, Canada.

REAGLE Joseph (2007), “Do As I Do: Authorial Leadership in Wikipedia”, *Wikisym 07*, Montreal.

ROTH Camille, “Viable Wikis, Struggle for life in the wikisphere”, *Proceedings of the 2007 International Symposium on Wikis*, Montreal.

ROTH Camille et al. (2008), “Measuring Wiki viability”, *Proceedings of the 2008 International Symposium on Wikis*, Porto.

SERAKIOTOU Niki (2008), “(Wiki + ResTechs) = (fresh documentation + organic knowledge management + training materials + good, cheap technical writers)”, *Proceedings of the 36th annual ACM SIGUCCS conference on User services conference*, Portland, USA, pp. 173-180.

STVILIA Besiki et al. (2005), “Information quality discussions in Wikipedia”, *Proceedings of ICKM05*, Sydney.

TANG Libby Veng-Sam et al. (2008), “A Method for Measuring Co-authorship Relationships in MediaWiki”, *Proceedings of Wikisym 08*, Porto.

VOSS Jakob (2005), “Measuring Wikipedia”, *10th International Conference of the International Society for Scientometrics and Informetrics*, Stockholm.

WILKINSON Dennis et HUBERMAN Bernardo (2007), “Cooperation and Quality in Wikipedia”, *WikiSym 07*, Montreal.

WANG Hao-Chuan, LU Chun-Hung, YANG Jun-Yi, HU Hsin-Wen, CHIOU Guey-Fa, et CHIANG Yueh-Tzu (2005), “An Empirical Exploration of Using Wiki in an English as a Second Language Course”, *Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies*, Kaohsiung, Taiwan.

WONG W. et WONG, L. (2005), “Using Wikiweb for Community Information Sharing and e-Governance”, In C. Crawford et al. (Eds.), *Proceedings of Society for Information Technology and Teacher Education International Conference 2005*, Chesapeake, Canada, pp. 2125-2128.

XIAO WenPeng et al. (2007), “On-line collaborative software development via wiki”, *Proceedings of the 2007 international symposium on Wikis*, Montreal, pp. 177 – 183.

ZHANG Wei et STORCK John (2001), “Peripheral Members in Online Communities”, *Americas Conference on Information Systems (ACIS 2001)*, Dallas.

ZHANG Yuejiao (2006), “Wiki means more: hyperreading in Wikipedia”, *Proceedings of the seventeenth conference on Hypertext and hypermedia*, Odense, Denmark, 23 - 26.

VIEGAS Fernanda B., WATTENBERG Martin, et DAVE Kushal (2004), “Studying cooperation and conflict between authors with HistoryFlow visualizations”, *Proceedings of the SIGCHI conference on Human factors in computing systems*, Vienne, Austriche.

Autres travaux pertinents ou autres articles sans comité de lecture (Working paper non soumis à conférences, mémoires, etc.)

ANDERSEN Espen (2004), “Using Wikis in a Corporate Context”, In Hohenstein A. and Wilbers K. (eds), *Handbuch E-Learning*, WoltersKluwer, Cologne, pp. 5-15.

ANGELES Michael (2004), “Using a Wiki for Documentation and Collaborative Authoring”, Working Paper LLRX.com.

BESCHASTNIKH Ivan et al. (2008), “Wikipedian Self-Governance in Action: Motivating the Policy Lens”, Association for the Advancement of Artificial Intelligence Working paper, Washington.

CHRISTEN Antoine (2003), « La gestion des connaissances dans les projets open source », Mémoire de Master en Informatique et organisation, HEC Lausanne.

CORREA Paulo et al. (2005), *Wikipedia: A Techno-Cult of Ignorance*, Ouvrage en ligne, http://www.aetherometry.com/antiwikipedia/awp_index.html

DESILETS Alain et PAQUET Sébastien (2005), “Wiki as a Tool for Web-based Collaborative Story Telling in Primary School: a Case Study”, Working Paper Ed-Media.

EBERSBACH Anja, GLASER Markus, HEIGL Richard, et ANDREA Adelung (2005), « Wiki : Web Collaboration », Springer, Berlin, 383 p.

FARKAS Meredith (2005), “Using Wikis to Create Online Communities”, WebJunction Working Paper.

GIJBERS Victor (2004), “Ideals of Knowledge Media from Plato to Wikipedia”, Working Paper. http://lilith.gotdns.org/~victor/writings/0051ideals.pdf

KORFIARTIS Nikolaos et NAEVE Ambjorn (s.d.), “Evaluating Wiki Contributions using social Networks: a Case Study on Wikipedia”, Royal Institute of Technology (KTH) Working Paper, Stockholm. http://www.tech2click.net/Korfiatis_Naeve-1.pdf

KLEIN Mark et IANDOLI Luca (2008), “Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The Mit Collaboratorium”, MIT Sloan Research Paper, Cambridge. SSRN: http://ssrn.com/abstract=1099082

MONTESSUIT Nicolas, *De la supercherie wiki*, pamphlet en ligne : http://pagesperso-orange.fr/cyber-marxiste/gestation/DEL2.html

PAQUET Sébastien (2003), *A Socio-technical Approach to Sharing Knowledge across Disciplines, Thèse de doctorat*, Université de Montréal.

PEDAUQUE Roger T. (2006), *Document et modernité*, Presses de l’ENSSIB, Lyon.

POLBORN, Mattias K. (2007), “Competing for Recognition through Public Good Provision” CESifo Working Paper Series No. 1920, Munich. SSRN: http://ssrn.com/abstract=924351.

SABEL Mikalai, GARG Anurag, et BATTITI Roberto (2005), “Wiki Rep: Digital Reputation in Virtual Communities”, Technical Report DIT-05-050, University of Trento.

SCHWALL Johannes (2003), “The wiki phenomenon”, Working Paper "Internet Literature seminar", Munster.

SZYBALSKI Andy (2005), “Why it’s not a wiki world (yet)”, Working Paper.

UNDP (2005), “Building Online Communities of Practice : The International Open Source Network Model”, E-Note 5.

ZACKLAD Manuel (2007), « Espace documentaire participatif et gouvernance », HAL Working Paper, http://archivesic.ccsd.cnrs.fr/docs/00/20/24/23/PDF/espace_participatif_et_gouvernance-Final_AA_Good.pd

ZHANG Xiaoquan Michael et ZHU Feng (2009), “Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia”, NET Institute Working Paper No. 07-22. SSRN: http://ssrn.com/abstract=1021450

Ouvrages et articles sur l’open-source, les communautés virtuelles, l’économie du logiciel

ALEXY Oliver et LEITNER Martin (2008), “Norms, Rewards, and their Effect on Motivation: The Case of Open Source Software Development”, Working paper, SSRN: http://ssrn.com/abstract=1007689

ASHTA Arvind et ASSADI Djamchid (2008), “Do Social Cause and Social Technology Meet? Impact of Web 2.0 Technologies on Peer-to-Peer Lending Transactions”, Working paper. SSRN: http://ssrn.com/abstract=1281373

AURAY Nicolas (2005), « Les configurations du marché du logiciel et les transformations du capitalisme », à paraître dans Eymard-Duvernay, F. Favereau, O., Salais, R., Thévenot, L. *Institutions et conventions*, La Découverte, Paris.

AURAY Nicolas (2005), "Le sens du juste dans un noyau d'experts : Debian et le puritanisme civique", *in Internet. Une utopie limitée. Nouvelles régulations, nouvelles solidarités*, CONEIN B., MASSIT-FOLLEA F et PROULX S, Presses de l'Université Laval, Montréal, 352 p. .

BAECKER R.M. et al. (1995), *Readings in human-computer interaction: toward the year 2000*, Morgan Kaufmann Publishers, San Francisco, 950 pp. .

BARCELLINI Flore, DÉTIENNE Françoise, BURKHARDT Jean-Marie, SACK Warren (2005), "Thematic Coherence and Quotation Practices in OSS Design-Oriented Online Discussions", GROUP’05, Sanibel Island, Florida.

BENGHOZI Pierre-Jean (2006), « Les communautés virtuelles : structuration sociale ou outil de gestion ? », *Entreprises et Histoire*, 43, pp. 67-81.

BENKLER Yoshai (2001), “Coase’s Penguin, or, Linux and the Nature of the Firm”, *Yale Law Journal*, 112-3, pp. 367-445.

BENUSSI Lorenzo (2006), “The evolution of free/libre open source software”, Working Paper http://opensource.mit.edu/papers/Benussi(2006)_The_evolution_of_FLOSS_1.pdf

BERDOU Evangelia (2007), "Managing the Bazaar: Commercialization and peripheral participation in mature, community-led Free/Open source software projects", Thèse de doctorat en "Médias et communication", LSE, http://opensource.mit.edu/papers/PhD_Berdou.pdf .

BITOUZET Christine (2001), « Le concept de communauté : un levier pour le développement du commerce électronique ? Analyse et études de cas de la construction de l’offre de services en ligne », Thèse de doctorat en gestion de l’Ecole Polytechnique.

BITZER Jürgen et al. (2007), “Intrinsic motivation in open source software development” *Journal of Comparative Economics*, 35-1, pp. 160-169.

BOMSEL Olivier, « Qu’est-ce que le numérique ? » *Entreprises et Histoire*, 43, juin, pp 5-14.

BOYLE James (1998), “Foucault in Cyberspace: Surveillance, Sovereignty, and Hard-Wired Censors”, *University of Cincinnati Law Review*, 66-1, pp. 177-206.

BRABHAM, Daren C. (2008), “Moving the Crowd at iStockphoto: The Composition of the Crowd and Motivations for Participation in a Crowdsourcing Application”, *First Monday*, 13-6.

CAMP Jean L. et SYME Serena (2001), “Code as Governance, The Governance of Code”, Working Paper Harvard University”.

CHAN Tzu-Ying et LEE Jen-Fang (2004), “A Comparative study of Online User Communities Involvement In Product Innovation and Development”, *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, New York, pp. 2039-2044.

CHARTIER Roger (2006), « L’écrit sur l’écran, ordre du discours, ordre des livres et manières de lire », *Entreprises et Histoire*, 43, pp 3-11.

CHEN Haiqiang, SHEN Huawei , XIONG Jin , TAN Songbo , CHENG Xueqi, (2006),"Social Network Structure behind the Mailing Lists", ICT-IIIS, Pékin. http://elvis.slis.indiana.edu/irpub/TREC/t15_notebook/notebook.papers/cas-ict.ent.pdf

CIFFOLILLI Andrea (2004), “The Economics of Open Source Hijacking and Declining Quality of Digital Information Resources: A Case for Copyleft”, EconWPA,. http://opensource.mit.edu/papers/ciffolili.pdf

COFFIN Jill (2000), « Analysis of open source principles in diverse collaborative communities », *First Monday*, 11-6.

CONEIN Bernard (2003), « Cognition distribuée, groupe social et technologie cognitive », *Réseaux*, 124, 55-79.

CONEIN Bernard (2004), « Relations de conseil et expertise collective:comment les experts choisissent-ils leurs destinataires dans les listes de discussion? », *Recherches sociologiques*, 35-3, pp. 61-74.

CONEIN Bernard et LATAPY Matthieu (2008), « Les usages des réseaux de communication électroniques: le cas de l’Open-Source », *Sociologie du Travail*, 50-3, , pp 331-352.

CROWSTON Kevin et HOWISON James (2004), "The perils and pitfalls of mining sourceforge", *26th International Conference on Software Engineering*, Edinburgh, Scotland.

CROWSTON Kevin et HOWISON James (2004), “The Social Structure of Free and Open Source Software development”, Syracuse FLOSS research working paper.

CROWSTON Kevin, Wei K., LI Qing, et HOWISON James (2006), "Core and Periphery in Free/Libre and Open Source Software Team Communications", *Proceedings of the 39th Annual Hawaii international Conference on System Sciences*, Kauai, Hawaï.

CROWSTON Kevin, LI Qing, WEI Kangning, ESERYEL Yeliz, HOWISON James (2007) , “Self-organization of teams for free/libre open source software development”, *Information and Software Technology*, 49-6, pp. 564-575.

DAHLANDER Linus et MAGNUSSON Mats G. (2005), “Relationships between open source software companies and communities: Observations from Nordic firms”, *Research Policy*, 34-4, pp. 481-493. http://ssrn.com/abstract=797404

DALLE Jean-Michel, DAVID Paul, GHOSH Rishab Aiyer, STEINMUELLER W. E. (2004), “Advancing Social Science Research on the Free and Open Source Software Mode of Production?”, Stanford Working Paper.

DANG-NGUYEN Godefroy et PENARD Thierry (1999), « Don et coopération dans Internet, une nouvelle organisation économique? », *Terminal*, 80-81, pp 101-116.

DE BESSES Marie (1999), « Les logiciels libres, un système de circulation fertile en apprentissage », *Terminal*, 80-81, pp 257-278.

DE SANCTIS Géraldine et al. (2004) , *Structure and Interaction in Online Discussion Forums*, in *Gouverner les organisations*, DUMEZ Hervé, L’Harmattan, Paris, pp. 101-140.

DEMAZIERES Didier, HORN François et ZUNE Marc (2007), “The functioning of a free software community, Entanglement of Three Regulation Modes, Control, Autonomous, and Distributed”, *Sciences Studies*, 20-2, pp. 34-54.

DEMIL Benoît, LECOCQ Xavier (2005), “Neither market nor hierarchies or network: the emerging bazaar governance”, AIMS, Angers, 2005.

DENG Feng (2008), “What is 'Open'? An Economic Analysis of Open Institutions”, MPRA Paper 8888, Munich, Germany. SSRN: http://ssrn.com/abstract=1138203

DORAT Rémi, LATAPY Matthieu, CONEIN Bernard et AURAY Nicolas (2007), « Multi-level analysis of an interaction network between individuals in a mailing list », *Annales des Télécommunications*, 62, pp. 325-349.

DUBOSSON Magali et BROILLET Alexandra (2008), “Analyzing Web 2.0 Internet Users in Order to Drive Innovation in Distribution Strategy of Luxury Watches”, Working Paper. SSRN: http://ssrn.com/abstract=1126222

FELLER Joseph et FITZGERALD Brian (2003), *Understanding Open-Source Software development*, Pearson, London, 211 pp. .

FROOMKIN A. Michael (2005), “Building the Bottom Up From the Top Down”, Workshop on “governance, Regulations and Powers on the Internet”, University of Miami Legal Studies Research Paper No. 2009-02.

GENSOLLEN Michel (2005), « Des réseaux aux communautés: la transformation des marchés et des hiérarchies », in PROULX Serge, POISSANT Louise et SENECAL Michel (eds.), *Communautés virtuelles : penser et agir en réseau*, Presses de l'Université Laval, Montréal, 107-132.

FREEMAN Stéphanie (2007), “The Material and Social Dynamics of Motivation: Contributions to Open Source Language Technology Development”, *Science Studies* 20-2, pp 55-77.

GINSBERG Deborah, KRIBBLE Meg et SHUCHA Bonnie J. (2009), “Inspiring Innovation: Planning, Implementing, and Evaluating the Web 2.0 Challenge”, Univ. of Wisconsin Legal Studies Research Paper No. 1077. SSRN: http://ssrn.com/abstract=1354822

GLOTT Rüdiger, MEISZNER Andreas et SOWE Sulayman K. (2007), “Report on the Learning Environment of FLOSS Communities”, Working Paper. http://opensource.mit.edu/papers/FLOSSCom_WP2_Phase_1_Report_v070709_1.pdf

GUERIN Francis (2004), «Le concept de communauté : une illustration exemplaire de la production des concepts en sciences sociales ?», 13ème Conférence de l’AIMS, Tours.

HEMETSBERGER Andrea et REINHARDT Christian (2004), "Sharing and Creating Knowledge in Open-Source Communities. The case of KDE", The Fifth European Conference on Organizational Knowledge, Learning, and Capabilities, Innsbruck, Austria http://opensource.mit.edu/papers/hemreinh.pdf

HENG Stefan, MEYER Thomas et STOBBE Antje (2007), “Implications of Web 2.0 for Financial Institutions: Be a Driver, Not a Passenger”. *Deutsche Bank Research, E-conomics*, 63. SSRN: http://ssrn.com/abstract=1003973

HORN François (2004), *L’économie du logiciel*, Repères La Découverte, Paris, 120p.

HUYSMANS Marleen et YUWEIN Lin (2005), « Learn to Solve Problems: a virtual Ethnographic Case Study of Learning in a GNU/LINUX Users Group », *E-Jovs*, 7, pp. 56-69.

IMHORST Christian (2005), “Anarchy and Source Code - What does the Free Software Movement has to do with Anarchism?”, Working Paper. http://opensource.mit.edu/papers/imhorst.pdf

JENKINS Peter S. (2003), “Leafleting and Picketing on the 'Cydewalk' - Four Models of the Role of the Internet in Labour Disputes”, *UCLA Journal of Law and Technology*. SSRN: http://ssrn.com/abstract=470682

JULLIEN Nicolas, CLÉMENT-FONTAINE Mélanie, et DALLE Jean-Michel (2002), *Projet RNTL: «Nouveaux modèles économiques, nouvelle économie du logiciel», Rapport final*.

KARINBEGG Borei (1999), “The Rewards of Managing an Electronic Mailing List”, *Library Trends*, 47-4, pp. 686-698. http://www.ideals.uiuc.edu/bitstream/handle/2142/8249/librarytrendsv47i4_opt.pdf?sequence=3#page=89

KOUSHIK Srinivas, BIRKINSHAW Julian et CRAINER Stuart (2009), “Using Web 2.0 to Create Management 2.0.”, *Business Strategy Review*, 20-2, pp. 20-23. SSRN: http://ssrn.com/abstract=1400336

KRISHNAMURTHY Sandeep (2002), “Cave or Community?: An Empirical Examination of 100 Mature Open Source Projects”, *First Mo*nday, 7-6.

KRISHNAMURTHY Sandeep et TRIPATHI Arvind K. (2006), “Bounty Programs in Free/Libre/Open Source Software”, in BITZER Jurgen, *The Economics of Open Source Software Development*, Lavoisier, Paris, pp. 165-183.

KRISHNAMURTHY Sandeep (2005), “The Launching of Mozilla Firefox- A Case

Study in Community-Led Marketing", Floss Working Paper. http://opensource.mit.edu/papers/sandeep2.pdf

KROGH Georg von, SPAETH Sebastian, et LAKHANI Karim R. (2003), "Community, joining, and specialization in open source software innovation: a case study", *Research Policy*, 32, pp. 1217–1241.

LANZARA Giovan Francesco et MORNER Michèle (2003), "The Knowledge Ecology of Open-Source Software Projects", 19th EGOS Colloquium http://opensource.mit.edu/papers/lanzaramorner.pdf

LAKHANI Karim R. et WOLF Robert G. (2003), "Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects", MIT Sloan Working Paper No. 4425-03. SSRN: http://ssrn.com/abstract=443040

LAKHANI Karim R., VON HIPPEL Eric (2003), "How open source software works: "free" user-to-user assistance", *Research Policy*, 32, pp. 923–943.

LATTEMANN Christophe and STIEGLITZ Stefan (2006), "Coworker Governance in Open-Source Projects", in BITZER Jurgen, *The Economics of Open Source Software Development*, Lavoisier, pp. 149-164. http://opensource.mit.edu/papers/Governance-in-OpenSourceProjects.pdf

LEE Gwendolyn K. et COLE Robert (2003), "From a Firm-Based to a Community-Based Model of Knowledge Creation: The Case of the Linux Kernel Development", *Organization Science*, 14-6, pp. 633–649. SSRN: http://ssrn.com/abstract=950262

LERNER Joshua etTIROLE Jean (2002), "The Simple Economics of Open-Source", *The Journal of Industrial Economics*, 50-2, , pp 197-234.

LERNER Joshua et TIROLE Jean (2004), "Economic Perspectives on Open-Source", *Advances in the Study of Entrepreneurship, Innovation, & Economic Growth*, 15, pp. 33-69.

LIN Yuwei (2005), "Hybrid Innovation: How Does the Collaboration Between the FLOSS Community and Corporations Happen? », *Knowledge, Technology and Policy*, 18-4, pp. 86-100.

LOPEZ-FERNANDEZ L., ROBLES G., et GONZALEZ-BARAHONA J. M. (2004), "Applying Social Network Analysis to the Information in CVS Repositories", *International Workshop on Mining Software Repositories (MSR)*. http://opensource.mit.edu/papers/llopez-sna-short.pdf

LJUNGBERG J. (2000), "Open Source Movements as a Model for Organizing", *European Journal of Information Systems*, 9-4, pp. 208-216.

MABILLOT David (2007), "User Generated Content: Web 2.0 Taking the Video Sector by Storm", *Communications & Strategies*, 65, p. 39-49. SSRN: http://ssrn.com/abstract=1008860

MADANMOHAN T.R. et Siddhesh NAVELKAR (2004), "Roles and Knowledge Management in Online Technology Communities: An Ethnography Study", *International Journal of Web Based Communities*, 1, 1, pp. 71-89.

MADINIER Hélène (2006), « La montée en puissance des réseaux sociaux », *Revue électronique suisse de science de l'information*, Congrès Online information, 3.

MAGGIONI Mario A. (2002), "Open Source Software Communities and Industrial Districts: a Useful Comparison?", Working Paper. http://opensource.mit.edu/papers/maggioni.pdf

MARKUS Lynne (2007), "The governance of free/open source software projects: monolithic, multidimensional, or configurational?", *Journal of Management and Governance*, 11-2, pp. 151-163.

MARRELLA Fabrizio et YOO Christopher S. (2007), "Is Open Source Software the New Lex Mercatoria?", *Virginia Journal of International Law*, 47-4, SSRN: http://ssrn.com/abstract=1007236

MASAFUMI Kotani (2004), "Manufacturing Organizational Memory: Logged Conversation Thread Analysis", *Issues in Informing Science & Information Technology*, 1, pp 1037-1042. http://informingscience.org/proceedings/InSITE2004/130kotan.pdf

MASSEY Bart (2005), « Longitudinal Analysis of LongTimescale Open Source Repository Data, *PROMISE'05*, Ottawa. http://promise.site.uottawa.ca/proceedings/pdf/2.pdf

MATSUMURA Naohiro et SASAKI Yoshihiro (2005), "Leader Qualification in Managing Nonprofit Organization", Working Paper. http://www.ii.ist.i.kyoto-u.ac.jp/jsai2005ws/proceedings/matsumura.pdf

MEISZNER A., (2007), "Communication tools in FLOSS communities: a look at FLOSS communities at large, beyond the development team". *Web Based Communities Conference 2007*.http://kn.open.ac.uk/public/getfile.cfm?documentfileid=10780

MERGEL Ines A., SCHWEIK Charles M. et FOUNTAIN Jane E. (2009), The Transformational Effect of Web 2.0 Technologies on Government. SSRN: http://ssrn.com/abstract=1412796

METIU Anca (2008), « Knowledge Creation Across Boundaries : Innovative Work in Software Development Teams », ESSEC Working Paper, Cergy-Pontoise.

METIU Anca et OBODARU O. (2008) "Women's Professional Identity Formation in the Free/Open Souirce Software Community" . Essec Research Center Working Paper DR‑08009, Cergy-Pontoise.

NEUS A. (2001), "Managing Information Quality in Virtual Communities of Practice", *Proceedings of the 6th International Conference on Information Quality at MIT*. Boston.

NICHOLS David M. et TWIDALE Michael B. (2006), « Usability Processes in Open Source Projects », *Software Process, improvement and Practice*, 11-2, pp. 149-162.

O'MAHONY S. O. et FERRARO F. (2007), "The Emergence of Governance in an Open Source Community"**,** *The Academy of Management Journal*, 50-5, pp. 1079 – 1106.

NOISETTE Perline et NOISETTE Thierry (2004), *La bataille du logiciel libre*, La Découverte, Paris, 128 p.

OCDE (2007), *Participative Web and User-Created Content, Web 2, Wikis and Social Networking*, OECD Publishing, 124 p.

O'MAHONY Sioban (2007), "The governance of open source initiatives: what does it mean to be community managed?", *Journal of Management and Governance*, 11-2, pp. 139-150.

OREG Shaul et NOV Oded (2008), "Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values", *Computers in Human Behavior*, 24-5, pp. 2055-2073.

OSTERLOH Margit et ROTA Sandra G. (2007), "Open Source Software Development - Just Another Case of Collective Invention?", *Research Policy*, 36-2, pp. 157-171. SSRN: http://ssrn.com/abstract=561744

VAN WENDEL DE JOODE Ruben (2006), "Organizing for continuity in a digital information commons: the institutions in open source communities", *IASCP Europe Regional Meeting. Building the European Commons: from Open Fields to Open Source*, Brescia, Italy, pp. 1-17.

MADANMOHAN T.-R. et NAVELKAR Sidesh (2004), "Roles and Knowledge Management in Online Technology Communities : An Ethnographic Study", *International Journal of Web Based Communities*, 1-1, pp. 71 – 89.

O'REILLY Tim (2007), "What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software", *Communications & Strategies*, 1, p. 17. SSRN: http://ssrn.com/abstract=1008839

RAYMOND Eric (1998), *La Cathédrale et le Bazar*, ouvrage en ligne : http://www.lifl.fr/~blondeel/traduc/cathedral-bazaar/mainfile.html∞

RAYNA Thierry et STRIUKOVA Ludmila (2008), "Web 2.0 is Cheap: Supply Exceeds Demand", Imperial College Internet Centre Reports. SSRN: http://ssrn.com/abstract=1371077

ROBERTS Jeffrey et al. (2007), "Understanding the Motivations, Participation and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects", Marshall School of Business Working Paper No. IOM 01-06. SSRN: http://ssrn.com/abstract=918518

ROBLES Gregorio-Martinez et GONZALEZ-BARAHONA Jesus M. (2004), "Executable source code and non-executable source code: analysis and relationships", *Proceedings of the Source Code Analysis and Manipulation, Fourth IEEE International Workshop*, London.

RUZE Emmanuel (2009), "Formes de régulations et de gouvernance formelles des communautés en ligne. Ce que dit la littérature et ce qu'il faudrait savoir." Rapport de recherche de la Chaire innovation et régulation des services numériques, Ecole Polytechnique.

SACK Warren, DETIENNE Françoise, DUCHENEAUT Nicolas, BURKHARDT Jean-Marie, MAHENDRAN Dilan et BARCELLINI Flore (2006), « A Methodological Framework for Socio-Cognitive Analyses of Collaborative Design of Open Source Software », *Computer Supported Cooperative Work*, 15, 2-3, pp. 229-250.

SADOWSKI Bert M., SADOWSKI-RASTERS Gaby et DUYSTERS Geert (2008), "Transition of governance in a mature open software source community: Evidence from the Debian case," *Information Economics and Policy*, 20-4, pp. 323-332.

SHAH Rajiv C. et KESAN Jay P. (2003), "Incorporating Societal Concerns into Communication Technologies". *IEEE Technology & Society*, pp. 28-33,. SSRN: http://ssrn.com/abstract=577561

SCHIFF Aaron (2002), "The Economics of Open Source Software: A Survey of the Early Literature", Review of Network Economics, 1-1, pp. 66-74. SSRN: http://ssrn.com/abstract=338000

SENYARD Anthony and MICHLMAYR Martin (2004), "How to Have a Successful Free Software Project", *Proceedings of the 11th Asia-Pacific Software Engineering Conference (APSEC'04)*, pp. 84 - 91. http://opensource.mit.edu/papers/senyardmichlmay.pdf

SOWE Sulayman, STAMELOS Ioannis, et ANGELIS Lefteris (2006), "Identifying knowledge brokers that yield software engineering knowledge in OSS projects Information and Software Technology", Journal of Systems and Software, 81-3, pp. 431-446. http://opensource.mit.edu/papers/IST-Vol-48-11-2006.pdf

SOWE Sulayman K, STAMELOS Ioannis et ANGELIS Lefteris (2007), "Understanding knowledge sharing activities in free/open source software projects: An empirical study", The *Journal of Systems and Software* 48 pp. 1025–1033. http://opensource.mit.edu/papers/JSS.pdf

VON HIPPEL Eric et VON KROGH Georges (2003), "Open Source Software and the "Private-Collective" Innovation Model:Issues for Organization Science", *Organization Science*, 14-2, pp. 209-223.

WATSON Richard T., BOUDREAU Marie-Claude, GREINER Martina, WYNN Donald, YORK Paul and GUL Rusen (2005), "Governance and global communities", *Journal of International Management*, 11-2, pp 125-142.

WYNN Donald E. (2003), "Organizational Structure of Open Source Projects: A Life Cycle Approach", *Proceedings of the 7th Annual Conference of the Southern Association for Information Systems*, Savannah, USA.

YUWEI Lin (2006), "A Techno-Feminist Perspective on the Free/Libre Open Source Software Development", in TRAUTH E. M., *Encyclopedia of Gender and Information Technology*, IGI Publishing, Hershey, USA, pp. 1148-1153.

Articles de presse

BENKLER Yoshai (2009), « Yochaï Benkler : Dépasser l'analyse de la topologie des réseaux », *Internet Actu*.

CARDON Dominique (2005), « De l'Innovation ascendante », *Internet Actu*.

SANGER Larry (2004), « Why Wikipedia Must Jettison its Anti-elitism », *Kuro5hin.org*.

UNDP (2005), "Building Online Communities of Practice : The International Open Source Network Model", *E-Note* n° 5,. http://www.apdip.net/apdipenote/5.pdf

WEISS Aaron (2005), "The power of collective intelligence", *netWorker* 9-3. pp. 16–23.

Liste des figures

Figure 1 : Matrice de synthèse des outils de communication sur Internet (BAECKER et al. 1995)..... 25
Figure 2: Capture d'écran du site principal de la communauté (effectuée le 5 juin 2008) 27
Figure 3: le wiki de la communauté Linux « Ubuntu » (capture d'écran du 5/07/09)..... 60
Figure 4 : archives du forum de la communauté capturé par Internet Archives (http://web.archive.org/web/*/http://wordpress.org/support/ , consulté le 02/09/08)..... 99
Figure 5: capture d'écran de l'ancien wiki de la communauté. (source : http://web.archive.org/web/20050224095754/wiki.wordpress.org/ ;, date : 24 février 2005)..... 100
Figure 6 : capture d'écran du système d'archivage officiel de la communauté WordPress (source : http://lists.automattic.com/mailman/listinfo) 101
Figure 7: historique des contributions de l'administrateur Michael Hancock (le 25/10/08) . 102
Figure 8 : système d'archivage des listes de discussion de « wp-hackers » (le 29/08/08)..... 104
Figure 9 : archives de « hackers », mois de juin 2004, présentées par fils de discussion..... 105
Figure 10 : capture d'écran d'un archivage de « minute » de chat IRC..... 108
Figure 11 : la première capture d'écran disponible de la communauté..... 123
Figure 12: Larry King, célèbre présentateur américain, utilisateur de WordPress 124
Figure 13 : schéma simplifié du système d'information de WordPress..... 125
Figure 14: capture d'écran du site d' Automattic (http://automattic.com/, le 07/09/09)..... 127
Figure 15 : les premières ébauches de documentation..... 131
Figure 16 : état du wiki le 24 mars 2004..... 133
Figure 17: les premières éditions sur le Codex le 30 juillet 2004..... 136
Figure 18: état du Codex après la « WordPress Backup Week », le 30/07/05..... 143
Figure 19 : changement d'apparence du Codex avec son statut de documentation officielle 146
Figure 20 : état du Codex le 01/09/07..... 149
Figure 21 : état final du Codex le 24/12/2008..... 151
Figure 22 : évolution des interventions des administrateurs sur la liste «wp- docs »..... 155
Figure 23 : datation des sorties des nouvelles versions significatives du logiciel 157
Figure 24 : séries chronologiques de l'activité chez les informaticiens..... 158
Figure 25 : agrégation des listes principales d'informaticiens et addition des séries chronologiques des informaticiens..... 158
Figure 26 : séries chronologiques des listes de discussions des projets informatiques secondaires 159
Figure 27 : série chronologique de la sommation de l'activité des projets informatiques secondaires 160
Figure 28 : graphique de fréquentation de la communauté, par Alexa de 2003 à 2007..... 161
Figure 29 : graphique de fréquentation de wordpress.org, par Alexa, période de août 2007 à février 2009 162
Figure 30 : série chronologique du nombre de sujets postés et du nombre de réponses « posts » (échelle logarithmique) 163
Figure 31 : évolution des téléchargements et de la liste de discussion « WP-Pro » 163
Figure 32 : série chronologique de la liste de discussion de la documentation 165
Figure 33 : graphique montrant l'évolution de la création d'articles..... 166
Figure 34 : croissance du nombre d'articles créés (en pourcentage cumulé)..... 167
Figure 35 : reconstitution de l'évolution du nombre d'éditions sur le wiki..... 169

Figure 36 : séries chronologiques de synthèse des activités de support et de programmation, échelle logarithmique 170
Figure 37 : séries chronologiques de l'activité des listes de discussions de support aux utilisateurs 172
Figure 38 : distribution de la taille des articles sur le wiki en forme de loi de puissance..... 185
Figure 39 : distribution des contributions des volontaires « bleus » et « rouges » 187
Figure 40 : tris à plat des temps de présence sur le Codex et de la densité des contributions 188
Figure 41 : tri à plat de la durée totale de présence et croisement avec la variable « temps de présence avec contributions ». 189
Figure 42 : corrélations entre les contributions et les discussions, et entre les types de discussions..... 191
Figure 43: tris à plat des différents types d'éditions sur le wiki..... 192
Figure 44 : effectifs des administrateurs et des non-administrateurs sur les pages du Codex 193
Figure 45 : corrélations entre le nombre de contributeurs et le nombre d'éditions, et entre les éditions non-mineures et le nombre total d'éditions 194
Figure 46 : corrélations entre le nombre de contributeurs et les éditions mineures et non-mineures 195
Figure 47 : capture d'écran de la page utilisateur de l'administrateur Carthik où apparaissent les « sub pages » sous forme de liste..... 231
Figure 48 : série chronologique des meetings IRC hebdomadaires 246
Figure 49 : série chronologique du nombre d'indésirables sanctionnés 250
Figure 50 : capture d'écran de : http://codex.wordpress.org/WordPress_Responsibilities.... 258
Figure 51 : articulations entre routines et problèmes de management..... 265
Figure 52 : nombre de contributeurs, différencié entre administrateurs et non administrateurs 272
Figure 53 : présentation du rapport entre le nombre d'administrateurs et de non-administrateurs. 273
Figure 54 : relation linéaire entre le nombre d'administrateurs et de non-administrateurs ... 274
Figure 55 : tris à plats du nombre d'interventions sur la liste et des variances des interventions sur cinq ans..... 275
Figure 56 : tris à plat du nombre de mois où chaque intervenant est intervenu, et du nombre moyen de messages par mois 276
Figure 57 : médiane et écart moyenne-médiane dans la population des intervenants 277
Figure 58 : corrélation entre l'indicateur de dispersion (variance) et l'indicateur de dissymétrie (écart moyenne-médiane) et entre variance et nombre d'interventions..... 277
Figure 59 : le nombre d'interventions de chacun est corrélé au nombre de mois d'intervention (échelle logarithmique) 278
Figure 60 : distribution en loi de puissance des interventions sur la liste de discussions..... 279
Figure 61 : séries chronologiques de l'évolution du nombre d'interventions des administrateurs et mise en évidence d'une relation linéaire entre les interventions des deux populations 280
Figure 62 : pourcentage d'interventions des administrateurs sur la liste 281
Figure 63 : on observe des corrélations entre le nombre des administrateurs et non-administrateurs et le chiffre de leurs interventions 282
Figure 64 : séries chronologiques du nombre moyen de message par contributeur, différencié à droite entre administrateurs et non-administrateurs 283
Figure 65 : le nombre d'intervenants fait augmenter le nombre et la moyenne des interventions..... 283
Figure 66 : pourcentage des interventions des quatre et six plus gros intervenants mensuels285

Figure 67 : les plus gros intervenants (quatre et six plus gros) entraînent les autres interventions 286
Figure 68 : la corrélation entre le nombre d'interventions des six et quatre plus gros intervenants et le nombre d'interventions des administrateurs montre que ce sont eux qui forment l'essentiel des plus grosses contributions 286
Figure 69 : le nombre d'interventions des quatre plus gros intervenants augmente la variance des interventions 287
Figure 70 : baisse de la variance mensuelle des contributions et différenciation entre variance des interventions des administrateurs et non-administrateurs 287
Figure 71 : corrélation entre la variance des contributions et l'écart moyenne-médiane (présentation avec régression et série chronologique des valeurs) 288
Figure 72 : corrélation entre la variance mensuelle des contributions, le nombre de contributions, et le nombre moyen de messages par contributeur 289
Figure 73 : corrélation entre le nombre d'intervenants et la dispersion des interventions..... 290
Figure 74 : mise en relation de la variance des interventions et de la démographie des catégories d'intervenants 290
Figure 75 : série chronologique de la taille mensuelle des fils et relation linéaire entre nombre de courriels et quantité d'information 292
Figure 76 : corrélation entre le nombre total mensuel de discussions et le nombre de fils par mois (relation linéaire et séries chronologiques avec échelle logarithmique)..... 293
Figure 77 : mise en relation du nombre d'intervenants mensuel, du nombre de fils mensuels et de la taille moyenne des fils 293
Figure 78 : mise en relation de la démographie des catégories d'intervenants et du nombre de fils de discussions qu'ils ont initié 294
Figure 79 : le nombre de fils mensuels est corrélé au nombre d'administrateurs, mais aussi au nombre de non-administrateurs 295
Figure 80 : mise en relation du nombre de fils initiés par chaque catégorie d'intervenants et du nombre de leurs interventions 295
Figure 81 : comparaison des causes de l'augmentation du nombre de fils par mois : le nombre des contributions des administrateurs l'explique mieux 296
Figure 82 : mise en relation de la variance mensuelle des interventions (échelle logarithmique), du nombre de courriels mensuels et de la taille moyenne mensuelle des fils. 297
Figure 83 : série chronologique du nombre de fils initiés par les administrateurs et les non-administrateurs, et présentation en pourcentage 298
Figure 84 : régressions mettant en évidence l'éviction de la proportion d'interventions des non-administrateurs suite à l'augmentation de la proportion de prise d'initiative des administrateurs 299
Figure 85 : visualisation du réseau des intervenants avec les plus prolifiques au centre..... 302
Figure 86 : visualisation du réseau des intervenants avec les plus prolifiques en périphérie 303
Figure 87 : mise en relation des degrés entrants et sortants et tri à plat des valeurs observées 306
Figure 88 : aperçu d'une partie de la base de données 310
Figure 89 : le nombre de fils initiés par les administrateurs est corrélé imparfaitement avec le nombre d'interventions des non-administrateurs, et vice-versa 311
Figure 90 : le nombre de fils initiés par les administrateurs est corrélé de façon positive avec le nombre de non-administrateurs ayant parlé. Il n'y a aucune relation entre le pourcentage de fils initiés par les administrateurs et le pourcentage de non-administrateurs. 312

Figure 91 : corrélation négative entre le nombre d'administrateurs et le pourcentage de contributions des non-administrateurs. Corrélation positive entre le nombre d'administrateurs ayant parlé et le nombre de fils initiés par eux. 312
Figure 92 : pas de corrélation entre le pourcentage d'administrateurs et le pourcentage de contributions des non-administrateurs. Relation décroissante entre le nombre de fils par mois et le pourcentage de contributions des non-administrateurs..... 313
Figure 93 : série chronologique de la valeur des pourcentages des interventions des non-administrateurs et du nombre de fils initiés par eux..... 314
Figure 94 : graphique de l'évolution de la liste de discussion du forum utilisateurs 323
Figure 95 : pourcentage de citations ayant trait au wiki sur la liste de discussions du forum 324
Figure 96 : pourcentage de fils avec au moins une citation ayant trait au wiki 324
Figure 97 : capture d'écran de la page principale du forum..... 328
Figure 98 : graphe de la liste de discussion du forum 332
Figure 99 : nombre de citations sur les listes d'informaticiens..... 343
Figure 100 : pourcentage de citations sur les listes d'informaticiens 343
Figure 101 : nombre de fils de discussions et nombre de citations concernées par les mots clefs se référant à la documentation sur « wp-hackers »..... 349
Figure 102 : capture d'écran du fichier Excel où fut effectué le codage des données 357
Figure 103 : mise en relation du nombre d'intervention sur la liste de discussion et du nombre de jours de contributions sur le Codex 369

Liste des tableaux

Tableau 1: synthèse de l'usage des sources numériques..... 118
Tableau 2 : liste des catégories les plus fréquentes (plus de 50 pages, par ordre décroissant) 195
Tableau 3 : synthèse des routines. 262
Tableau 5 : somme des demi-degrés entrants et sortants chez les principaux intervenants... 304
Tableau 6 : caractéristiques quantitatives du codage des listes de discussions..... 320
Tableau 7 : caractéristiques quantitatives du codage de la liste « wp-forum »..... 323
Tableau 8 : caractéristiques quantitatives du codage des listes des informaticiens 342