



HAL
open science

Usage et valeur économique des technologies de traitement big data

Charles Xavier

► **To cite this version:**

Charles Xavier. Usage et valeur économique des technologies de traitement big data. Economies et finances. Université Paris-Saclay, 2023. Français. NNT : 2023UPASI014 . tel-04411457

HAL Id: tel-04411457

<https://theses.hal.science/tel-04411457>

Submitted on 23 Jan 2024

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.

Usage et valeur économique des technologies de traitement big data

Use and economic value of big data processing technologies

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 630 Droit Economie Management (DEM)

Spécialité de doctorat : Sciences Economiques

Graduate School : Economie - Management. Référent : Faculté de droit, économie et gestion

Thèse préparée dans l'unité de recherche **Université Paris-Saclay, Réseaux Innovation Territoires et Mondialisation, 92330, Sceaux, France.**, sous la direction de **Anne PLUNKET**, Professeur des Universités

Thèse soutenue à Paris-Saclay, le 11 décembre 2023, par

Charles XAVIER

Composition du Jury

Membres du jury avec voix délibérative

David FLACHER Professeur, Université Technologique de Compiègne	Président
Marc BAUDRY Professeur, Université Paris Nanterre	Rapporteur
Valerio STERZI Professeur, Université de Bordeaux	Rapporteur
Grazia CECERE Professeur, Institut Mines-Télécom Business School	Examinatrice
François ACQUATELLA Maître de conférences, Université Paris Dauphine-PSL	Examineur

Titre : Usage et valeur économique des technologies de traitement big data

Mots clés: Big Data, Données massives, MapReduce, Brevets, Patent Pledge, Innovation ouverte, Google

Résumé : L'objectif principal de cette thèse est d'étudier l'émergence des technologies de traitement Big Data et leur impact sur les innovations du secteur. Cette analyse est réalisée à travers l'étude d'une des principales technologies Big Data : MapReduce, technologie qui a fait l'objet d'un dépôt de brevet par l'entreprise américaine Google (Alphabet depuis 2015) en 2004.

Cette thèse se compose de trois chapitres. Dans le premier chapitre, une méthodologie scientométrique est appliquée pour construire une base de données des technologies Big Data autour du brevet MapReduce. Une étude sectorielle est ensuite proposée.

Le second chapitre étudie l'impact de l'innovation ouverte (fondation Apache et les « patent pledge ») au sein du secteur des technologies Big Data grâce à une analyse économétrique des citations du brevet MapReduce et de ses follow-on innovations.

Enfin, le troisième chapitre étudie la valeur privée et sociale des patent pledge dont MapReduce dans le portefeuille technologique de Google ainsi que des Pledge de IBM et Microsoft par rapport aux brevets non pledge de leurs portefeuilles et de brevets de contrôle ayant les mêmes dates de priorité et les mêmes technologies.

Title: Use and economic value of big data processing technologies

Keywords: Big Data, MapReduce, Patents, Patent Pledge, Open innovation, Google

Abstract: The main objective of this thesis is to study the emergence of Big Data processing technologies and their impact on innovations in the sector. This analysis is carried out through the study of one of the main Big Data technologies: MapReduce, a technology patented by the American company Google (Alphabet since 2015) in 2004.

This thesis is divided into three chapters. In the first chapter, a scientometric methodology is applied to build a database of Big Data technologies around the MapReduce patent. An industry study is then proposed.

The second chapter examines the impact of open innovation (Apache Foundation and patent pledges) within the Big Data technology sector, through an econometric analysis of citations of the MapReduce patent and its follow-on innovations.

Finally, the third chapter examines the private and social value of patent pledges, including MapReduce in Google's technology portfolio, as well as IBM's and Microsoft's pledges, compared to non-pledge patents in their portfolios and control patents having similar priority dates and technologies.

REMERCIEMENTS

Je souhaite tout d'abord à remercier ma directrice de thèse le Professeur Anne Plunket, qui m'a accompagné, soutenu, encouragé tout au long de cette grande aventure, durant toutes ces années de recherche, où elle s'est montrée toujours disponible, s'adaptant à mon emploi du temps compliqué. Sa finesse d'analyse, sa rigueur et son empathie m'ont été précieux dans la réalisation de ce travail. Je la remercie sincèrement pour tout ce qu'elle a pu m'apporter en conseils, réflexions et qualités humaines. Son encadrement a été un étayage fondamental qui m'a permis, malgré les tiraillements de la vie entrepreneuriale, de vous présenter cette thèse.

Je tiens à remercier chaleureusement les rapporteurs de ma thèse les Professeurs Marc Baudry et Valerio Sterzi pour le temps qu'ils ont accordé à la lecture de ma thèse. Je tiens à les remercier pour les commentaires précieux qu'ils m'adresseront et qui me permettront de faire progresser ces travaux recherches.

Je tiens à remercier le professeur David Flacher d'avoir accepté d'être le Président de mon jury de thèse. Je tiens également à remercier tous les membres du jury les professeurs Marc Baudry, Valerio Sterzi, Grazia Cecere et François Acquatella pour leur présence lors de la soutenance de cette thèse et pour les observations qu'ils me feront afin d'améliorer mon travail de recherche.

Je remercie bien sûr les différents membres des comités de suivi.

En premier lieu le Professeur Grazia Cecere qui a participé à l'ensemble de mes comités et dont l'écoute et les recommandations dans la démarche scientifique m'ont permis de clarifier ma pensée et d'élaborer un cheminement plus méthodique.

Je remercie également les Professeurs suivants : Aurélie Duzert qui a participé aux comités de suivi 2018 et 2019, Nicolas Soulié qui a participé au comité 2020, Thierry Pénard et Fabrice Le Guel qui ont participé au comité 2021 et Lorenzo Cassi qui a participé au comité 2022.

Par ailleurs, il me faut avouer que cumuler une thèse avec la gestion d'une entreprise a été beaucoup plus contraignant que prévu. Cela a nécessité du temps, de l'énergie et dans tous les moments de difficulté j'ai pu compter, à chaque fois, sur mes proches.

Je souhaite vivement les remercier.

Je commencerai par Johan Politzer, analyste fonctionnel chez Engie IT avec qui j'ai pu échanger sur les technologies Big Data et qui m'a apporté son éclairage dans ce domaine.

Je remercie aussi mes collaborateurs et associés qui, dans le cadre des missions confiées à ma société Metrics 720, ont permis que je puisse libérer du temps pour mon travail de recherche.

J'ai également une très forte pensée pour ma famille. Je remercie mes parents Alfred et Christiane qui m'ont donné le goût du travail bien fait, l'envie d'apprendre encore et toujours. En facilitant mon quotidien, ils m'ont permis de me consacrer à mes recherches.

Je remercie aussi ma sœur Joséphine, qui a partagé mes doutes et mes tourments en restant positive et mes frères Pradeep et Antoine ainsi que ma belle-sœur Gwladys pour leur soutien moral et leurs encouragements.

Pour terminer, j'ai une pensée pour ma nièce et mon neveu, Vidhya et Ishan encore très jeunes. Je souhaiterais qu'ils soient plus tard inspirés par ce travail dans leur vie personnelle et professionnelle.

TABLE DES MATIERES

REMERCIEMENTS	ii
TABLE DES MATIERES	iv
LISTE DES TABLEAUX	vii
LISTE DES FIGURES	viii
INTRODUCTION GENERALE	1
CHAPITRE 1. Emergence des technologies de traitement des données big data – une analyse de brevets	10
1.1 Méthodologie permettant de définir un périmètre d’analyse	10
1.1.1 Intérêt de la démarche	10
1.1.2 Méthodologies pour définir un périmètre d’analyse	12
1.2 Définition du périmètre d’analyse des technologies de traitement de données Big Data	16
1.2.1 Revue de la littérature	16
1.2.2 Présentation de la méthodologie utilisée	19
1.3 Analyse des trois périmètres	28
1.3.1 Etude de la temporalité	28
1.3.2 Étude des principaux déposants	32
1.3.3 Localisation des déposants	37
1.3.4 Co-brevets	40
1.3.5 Etude des champs d’applications.	42
1.4 Conclusion	48
1.5 Références	50
CHAPITRE 2. Innovation ouverte et collaboration dans le secteur des technologies de traitement des données « Big Data » : propriété intellectuelle et enjeux stratégiques	53
2.1 Introduction	53
2.2 Les incitations à collaborer avec ses concurrents	56
2.2.1 Les modalités de coopération entre concurrents	58
2.2.2 Coopétition et émergence d’un standard technologique	63

2.3	Innovation ouverte et collaborations au sein des technologies « Big Data »	68
2.3.1	Innovation ouverte « communautaire » : la fondation Apache	69
2.3.2	Innovation ouverte « unilatérale » : les patents pledge	75
2.4	Stratégie d'analyse empirique	83
2.4.1	Modèles d'estimation et hypothèses	83
2.4.2	Constitution des deux groupes de travail	86
2.5	Estimations agrégées : citations reçues par brevet	92
2.5.1	Variables et descriptions	92
2.5.2	Résultats des estimations agrégées	97
2.6	Estimations désagrégées : citations reçues par brevet par année	107
2.6.1	Variables et descriptions	108
2.6.2	Résultats des estimations désagrégées	110
2.7	Conclusion	115
2.8	Références bibliographiques	119
CHAPITRE 3. Impact des patents pledge sur la valeur des portefeuilles technologiques des entreprises		123
3.1	Les stratégies d'innovation ouverte	124
3.1.1	Le principe de l'innovation ouverte	124
3.1.2	La stratégie d'innovation à « frontières ouvertes »	126
3.1.3	La stratégie d'innovation « open-source »	128
3.2	Positionnement de la technologie MapReduce dans le champ de l'innovation ouverte	131
3.2.1	Définition d'un cadre d'analyse	131
3.2.2	Analyse des mécanismes d'innovation ouverte de la technologies MapReduce	133
3.3	Impact des patents à travers la littérature	136
3.3.1	Les patent pledge pour promouvoir les technologies vertes	136
3.3.2	Les patent pledge pour promouvoir les véhicules électriques	140
3.4	Stratégie d'analyse empirique des patent pledge : le cas de trois acteurs des TIC	144
3.4.1	Modèles d'estimation et hypothèses	144
3.4.1	Constitution des deux groupes de travail	145
3.4.2	Description des portefeuilles des brevets de Google, IBM et Microsoft	148
3.5	Estimations agrégées avec groupe de contrôle	155
3.5.1	Variables et statistiques descriptives	155
3.5.2	Statistiques descriptives	157

3.5.3	Résultats des estimations avec contrôles	158
3.6	Estimations agrégées centrées sur des portefeuilles de brevets	161
3.6.1	Statistiques descriptives	161
3.6.2	Résultat des estimations	162
3.7	Conclusion	171
3.8	Références	173
CONCLUSION GENERALE		175

LISTE DES TABLEAUX

Tableau 1.1 : nombre de brevets « communs » par périmètre	26
Tableau 1.2 Structure du périmètre 1 avec brevets « communs »	27
Tableau 1.3 Top 20, Top 50 et de l'ensemble des trois périmètres	34
Tableau 1.4 : nombre de brevets déposés par le Top 20 des déposants	35
Tableau 1.5: nombre de brevets déposés de l'ensemble du Top 20 des entreprises	36
Tableau 1.6 : nombre de brevets déposés sur la période 2010-2017 – détail par périmètre.....	36
Tableau 1.7 : Top 10 localisation des déposants – ensemble des trois périmètres	38
Tableau 1.8: nombre de déposants par co-brevet	40
Tableau 1.9 : co-brevets ayant deux déposants – période de dépôt.....	41
Tableau 1.10: Top 10 des paires de co-dépôts.....	41
Tableau 1.11: répartition des brevets du Top10 des co-dépôts.....	42
Tableau 1.12 : répartition des brevets à travers la nomenclature WIPO – (3 périmètres).....	43
Tableau 1.13: répartition des brevets sur le Top 10 des classes et sous-classes de CPC.....	46
Tableau 2.1 Liste des brevets associés à la technologie MapReduce.....	87
Tableau 2.2 Nombre de brevets traités ayant au moins un brevet de contrôle	88
Tableau 2.3 liste des sponsors de la fondation Apache.....	93
Tableau 2.4 Statistiques descriptives	95
Tableau 2.5. Test de comparaison des moyennes entre le groupe traité et les contrôles.....	96
Tableau 2.6 : Impact de la technologie Pledge MapReduce et Follow-on sur le nombre de citations reçues.....	97
Tableau 2.7: impact de la technologie Pledge MapReduce et Follow-on sur le nombre de citations reçues.....	98
Tableau 2.8: Rôle de la Fondation Apache et type de follow-on innovation sur le nombre de citations	100
Tableau 2.9 : Rôle de la Fondation Apache et type de follow-on innovation sur le nombre de citations sans autocitation.....	101
Tableau 2.10 : Rôle de la Fondation Apache et type de follow-on innovation sur le nombre de citations	104
Tableau 2.11 : Rôle de la Fondation Apache et type de follow-on innovations sur le nombre de citations sans autocitation.....	106
Tableau 2.12 : Statistiques descriptives	108
Tableau 2.13 : Test de comparaison des moyennes entre le groupe traité et les contrôles.....	109
Tableau 2.14 : Nombre de citations reçues par année.....	111
Tableau 2.15: Nombre de citations reçues par année avec niveaux de citations des follow-on innovations	113
Tableau 2.16 : Nombre de citations reçues par année avec niveaux de citations des follow-on innovations.....	117
Tableau 3.1 Eco-brevets – date d'intégration et nombre de brevets par entreprise (source : Contreras et al., 2018)	137
Tableau 3.2 : Périmètre des brevets Google, IBM et Microsoft ayant au moins un brevet de contrôle.....	146
Tableau 3.3 Statistiques descriptives – Brevets Google, IBM, Microsoft et contrôles.....	157

Tableau 3.4 Tests de comparaison des moyennes	158
Tableau 3.5 Ensemble des brevets (Google, IBM et Microsoft) et leurs contrôles	159
Tableau 3.6 Impact des pledge Google par rapport à l'ensemble des brevets et contrôles.....	160
Tableau 3.7 Impact des pledge IBM par rapport à l'ensemble des brevets et contrôles.....	161
Tableau 3.8 Statistiques descriptives	162
Tableau 3.9 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft	164
Tableau 3.10 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft.....	165
Tableau 3.11 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft	166
Tableau 3.12 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft	168
Tableau 3.13 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft	169
Tableau 3.14 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft	170

LISTE DES FIGURES

Figure 0.1: Estimation du volume de données numériques créées ou répliquées par an dans le monde, en zettaoctets (Sources : IDC, Seagate, Statista, 2021)	1
Figure 0.2 : Évolution de la recherche du mot clé "MapReduce" sur internet (en indice) entre 2004 et 2015 (Sources : Google Trends, 2023)	3
Figure 1.1 principe de l'algorithme MapReduce (Charles Xavier, 2023)	20
Figure 1.2 Trajectoire du brevet MapReduce : nombre de brevets et classification CPC	22
Figure 1.3. Principe d'extraction des « follow-on innovations » du brevet MapReduce	22
Figure 1.4. Définition des trois périmètres d'analyse	25
Figure 1.5. Synthèse des trois périmètres d'analyse	26
Figure 1.6. Nombre de brevets par an déposés sur l'office USPTO – ensemble des trois périmètres	29
Figure 1.7 Périmètres	30
Figure 1.8. Taux d'accroissement du nombre de brevets déposés par an à l'office USPTO	31
Figure 1.9. Top 50 des déposants de brevets sur la période 1994-2020	33
Figure 1.10 Périmètres par période.	38
Figure 1.11. Top 10 par champs d'activités	44
Figure 1.12. Indicateur PatentScope sur le Top 10 des sous-classes de CPC – trois périmètres	488
Figure 2.1 Historique de l'écosystème des technologies de traitement	54
Figure 2.2 Processus d'émergence d'un standard technologique	66
Figure 2.3 Terminaux mobiles intelligents - utilisation des principaux systèmes d'exploitation 1998–2006 (source : Gueguen, 2009)	67
Figure 2.4 Cycle de vie d'un projet au sein de la fondation Apache	72
Figure 2.5. Parc de voitures à batterie électrique et de voitures hybrides rechargeables dans le monde de 2005 à 2022 (source Statista, 2023)	80

<i>Figure 2.6 Impact de la technologie MapReduce et ses brevets associés sur les innovations ultérieures – stratégie empirique (Charles Xavier, 2023)</i>	86
<i>Figure 2.7 Composition du groupe traité</i>	87
<i>Figure 2.8 Brevets traités & contrôles - nombre de dépôts de brevets par an</i>	89
<i>Figure 2.9 Brevets traités & contrôles - champs technologiques (WIPO)</i>	89
<i>Figure 2.10 Brevets traités et contrôles</i>	90
<i>Figure 2.11 : synthèse de l'impact des brevet pledge MapReduce et follow-on innovations</i>	114
<i>Figure 3.1 Cercle vertueux de l'innovation fermée (source : Chesbrough, 2003)</i>	124
<i>Figure 3.2 : le continuum d'innovation ouverte (Euchner, 2013)</i>	132
<i>Figure 3.3 : la technologie MapReduce dans le continuum d'innovation ouverte (source : Euchner, 2013, modifié par Xavier, 2023)</i>	134
<i>Figure 3.4 nombre de brevets déposés par Google entre 1994 et 2017 (source : USPTO)</i>	148
<i>Figure 3.5 nombre de brevets déposés par IBM entre 1994 et 2017 (source : USPTO)</i>	149
<i>Figure 3.6 nombre de brevets déposés par Microsoft entre 1994 et 2017 (source : USPTO)</i>	149
<i>Figure 3.7 Top 3 des technologies (WIPO) des brevets Google entre 1994 et 2017 (source : USPTO)</i>	150
<i>Figure 3.8 Top 3 des technologies (WIPO) des brevets IBM entre 1994 et 2017 (source : USPTO)</i>	150
<i>Figure 3.9 Top 3 des technologies (WIPO) des brevets Microsoft entre 1994 et 2017 (source : USPTO)</i>	151
<i>Figure 3.10 : nombre de citations reçues par les brevets Google non pledge entre 1994 et 2017 – agrégées sur les 5 années après la publication du brevet (source : USPTO)</i>	152
<i>Figure 3.11 nombre de citations reçues par les brevets Google pledge entre 1994 et 2017 - agrégées sur les 5 années après la publication du brevet (source : USPTO)</i>	152
<i>Figure 3.12 nombre de citations reçues par les brevets IBM non pledge entre 1994 et 2017 - agrégées sur les 5 années après la publication du brevet (source : USPTO)</i>	153
<i>Figure 3.13 nombre de citations reçues par les brevets IBM pledge entre 1994 et 2017 - agrégées sur les 5 années après la publication du brevet (source : USPTO)</i>	153
<i>Figure 3.14 nombre de citations reçues par les brevets Microsoft non pledge entre 1994 et 2017 - agrégées sur les 5 années après la publication du brevet (source : USPTO)</i>	154
<i>Figure 3.15 nombre de citations reçues par les brevets Microsoft pledge entre 1994 et 2017 - agrégées sur les 5 années après la publication du brevet (source : USPTO)</i>	154

INTRODUCTION GENERALE

L'émergence des nouvelles technologies de l'information et de la communication (TIC) ainsi que la démocratisation de l'accès à internet ont entraîné dans leur sillage une production massive de données individuelles.

Ces données ou « data » sont collectées depuis de nombreuses années dans les secteurs de la banque et de l'assurance qui, historiquement, ont eu besoin de s'appuyer sur des éléments tangibles pour développer leurs outils métiers (outils d'évaluation des risques, de calcul d'indemnisation des sinistres...). Par leur nature, ces deux secteurs d'activité ont eu un accès aisé aux données de leurs clients et ils ont rapidement compris l'intérêt de les exploiter.

L'accès aux données individuelles a été pendant plusieurs décennies un sujet de réflexion récurrent pour de nombreux secteurs d'activité n'ayant pas accès à ce type d'information car difficilement récupérable ou nécessitant des moyens techniques et financiers colossaux.

La généralisation de l'accès à internet va modifier cet état de fait et, la collecte des données individuelles va se trouver grandement simplifiée grâce à la numérisation de l'information. Ce ne sont plus les entreprises qui collectent les données auprès de leurs clients mais les utilisateurs qui vont spontanément fournir ces informations en remplissant un formulaire numérique et ceci afin d'accéder à un service digital (e-commerce, réseaux sociaux, programme de fidélité...).

Le volume de données mondial va s'accroître rapidement comme le montre la Figure 0.1 :

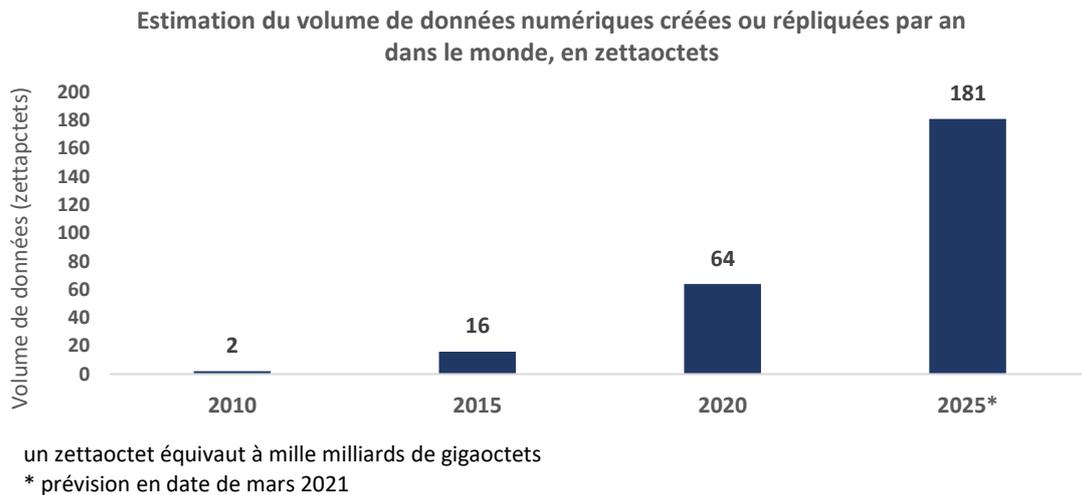


Figure 0.1: Estimation du volume de données numériques créées ou répliquées par an dans le monde, en zettaoctets (Sources : IDC, Seagate, Statista, 2021)

Des entreprises de tout type de secteurs se sont ainsi constituées d'immenses bases de données. Un nouveau terme a alors fait son apparition : le « Big Data » (ou « données massives ») pour désigner cette masse de données individuelles collectées à travers différents points de contacts utilisateurs.

A l'instar des secteurs banque et assurance, l'exploitation des données Big Data a pour objectif de mieux comprendre les usages et les attitudes des utilisateurs. Cette analyse revêt ainsi une double « promesse » pour les entreprises :

- A court terme : permettre de mieux cibler leurs clients et donc de mieux vendre.
- A long terme : anticiper les décisions de leurs clients et donc d'avoir de meilleurs outils d'analyses prédictives (Brasseur, 2013).

Ces deux points sont depuis toujours des enjeux majeurs dans les stratégies de développement des entreprises, de sorte qu'elles n'hésitent pas à modifier leur organisation interne afin de s'adapter aux exigences qu'impose l'utilisation des outils Big Data (McAfee & Brynjolfsson, 2012). Depuis quelques années, on assiste à une importante mutation structurelle au sein des entreprises dans le but d'intégrer l'analyse des Big Data dans leur schéma de gouvernance. Cela se traduit par l'apparition de nouveaux métiers dans les organigrammes des sociétés tels que le « Chief Data Officer » en charge de l'exploitation transverse de l'ensemble des données.

Aujourd'hui, on constate que cette évolution a été portée par un certain nombre d'entreprises et notamment les grands groupes industriels soucieux de ne pas être dépendants des géants du web dans ce domaine (Fléchaux, 2014). Les différents retours d'expériences relayés par les médias montrent bien l'engouement autour de ces « stratégies orientées Big Data ». Il est donc pertinent de s'intéresser sur l'usage et l'impact des technologies de traitement Big Data dans le développement des entreprises et sur la possibilité de quantifier la valeur économique qui en découle.

1. Problématiques

Afin d'intégrer les traitements Big Data dans leur processus de travail, les entreprises ont dû faire face à un premier défi : comprendre et s'appropriier les nouvelles technologies associées à ce terme. En effet, l'explosion du volume de données collectées a entraîné une nouvelle réflexion autour des techniques permettant de mieux traiter cette nouvelle masse d'information. Les outils traditionnels tels que les bases de données relationnelles (type langage SQL) n'étaient

plus adaptés pour traiter les données de type Big Data que l'on caractérise par les « 3V » (Rawat & Yadav, 2021) :

- Volume : caractérisant la quantité importante des données
- Variété : caractérisant la forte variété des données
- Vélocité : caractérisant la vitesse de production des données

Des technologies spécifiques au traitement Big Data ont donc été développées et apparaissent des termes tels que « Hadoop », « MapReduce » qui reflètent de nouveaux procédés dans le stockage, le traitement et l'analyse de données. L'objectif de ces solutions est de traiter un volume important de données (« Volume ») se trouvant sur des terminaux variés comme des ordinateurs, smartphones, tablettes (« Variété »), et produites en temps réel (« Vélocité »).

La compréhension de ces nouvelles technologies est devenue une thématique centrale pour les entreprises comme en témoigne l'évolution de la croissance des recherches sur internet au début du XXIème siècle autour des principales solutions Big Data.

La Figure 0.2 illustre cette évolution pour la technologie de traitement MapReduce.

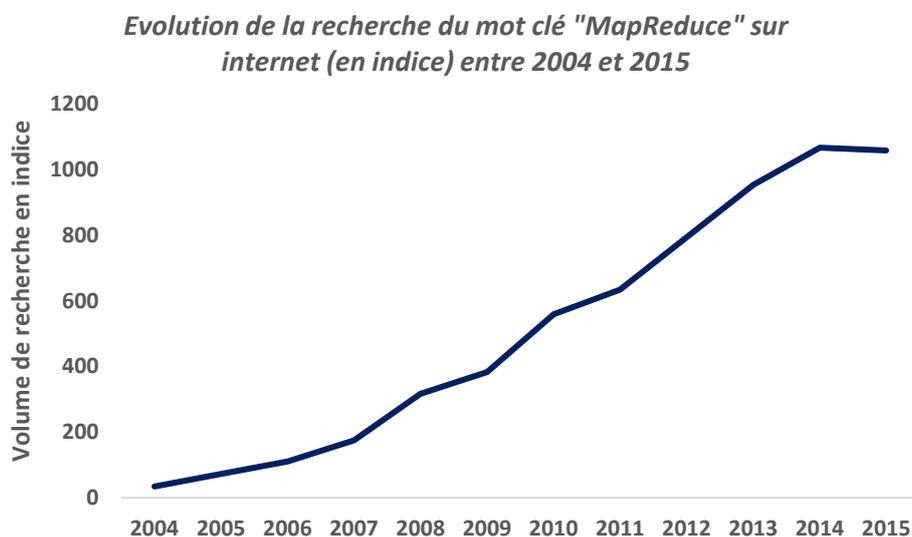


Figure 0.2 : Évolution de la recherche du mot clé "MapReduce" sur internet (en indice) entre 2004 et 2015 (Sources : Google Trends, 2023)

Afin de comprendre l'usage des technologies de traitement Big Data dans les entreprises, cette thèse étudie la diffusion de ces nouvelles technologies au sein du paysage industriel grâce à l'étude d'un des principaux indicateurs de propriété industrielle : le brevet.

2. Cadre théorique

Le brevet permet d'octroyer à une personne (ou à une entreprise) qui le dépose un droit exclusif d'exploitation d'une invention sur une période définie (au maximum vingt ans). Au terme de cette période, le déposant accepte que son invention tombe dans le domaine public et puisse être dupliquée par un tiers. Le brevet favorise ainsi l'innovation en créant un monopole temporaire et il contribue également à la diffusion de la connaissance en permettant l'accès aux technologies une fois la période d'exclusivité révolue (Cohendet et al., 1998). En analysant les brevets, il va être possible de comprendre d'une part l'émergence des technologies sur un secteur en étudiant les dépôts de brevets et d'autres part, la diffusion des technologies en analysant l'utilisation des brevets par les entreprises (contrat de licence, vente, mise à disposition...). Sur ce dernier point, il apparaît que les recettes liées à l'exploitation des brevets par les entreprises sont un facteur non négligeable. En 2001, l'entreprise IBM annonçait avoir perçu plus de 1,9 milliard de dollars de redevances sur ses brevets (Chesbrough, 2003). Dans cette optique, les entreprises vont également être incitées à déposer des brevets afin d'exploiter financièrement leurs technologies, ce qui peut conduire à des litiges juridiques concernant la paternité des inventions. On peut citer l'exemple du procès entre les entreprises Apple et Samsung concernant leurs brevets sur la technologie contenue dans les smartphones. D'une part, Apple a accusé Samsung d'avoir reproduit le design de l'iPhone pour la conception de son smartphone Galaxy et d'autre part, Samsung a reproché à Apple d'avoir copié la technologie de la batterie de son smartphone (Certes, 2018). Dans ce contexte, le rôle du brevet évolue, il n'est plus uniquement un outil d'incitation à l'innovation mais également un outil stratégique comme l'ont souligné Noel & Schankerman, (2013). Leur étude met en évidence le rôle défensif du brevet car avoir un large portefeuille de brevets permet de limiter les procédures judiciaires. Par ailleurs, la croissance régulière des dépôts de brevets depuis une vingtaine d'années montre bien l'importance de cet enjeu pour les entreprises (source : Wipo¹). De plus, un large portefeuille de brevets permet également d'améliorer son pouvoir de négociation dans la mise en place de ces accords ayant pour objectif de diminuer les coûts de transactions (Heald, 2005).

Le brevet est donc un outil pertinent pour caractériser l'émergence et l'usage d'une nouvelle technologie à travers son rôle dans l'incitation à l'innovation et comme outil stratégique.

¹ https://www.wipo.int/pressroom/fr/articles/2022/article_0013.html

3. Cadre analytique

Le brevet est un titre de propriété industrielle, et par conséquent, pour jouir de ce titre l'inventeur doit en faire la demande auprès des autorités compétentes. Comme le brevet est délivré pour un territoire géographique donné, l'inventeur doit d'abord considérer le lieu de dépôt qui sera le lieu d'exploitation de l'invention (Barré & Laville, 1994). En effet, des offices tels que « United States Patent and Trademark Office » (USPTO) ou « l'Institut National de la Propriété Industrielle » (INPI) sont habilités à délivrer des brevets uniquement pour les pays auxquels ils sont rattachés, respectivement les Etats-Unis et la France pour les offices précédemment cités. L'inventeur a donc le choix de breveter son invention dans un ou plusieurs pays différents, il devra pour cela en faire la demande à chacun des offices habilités. Il est important de préciser que chaque pays a sa propre réglementation en matière de dépôt de brevet et de ce fait l'obtention d'un brevet pour une invention dans un pays ne garantit en rien la délivrance d'un brevet pour cette même invention dans un autre pays. A titre d'exemple, l'office américain USPTO accorde des brevets pour des procédés intégrant des algorithmes mathématiques, ce qui n'est pas le cas de l'office français INPI du fait de l'article L.611-10² du code français de la propriété intellectuelle qui exclut de la brevetabilité « les théories scientifiques et les méthodes mathématiques », mais également « les programmes d'ordinateur ».

Ces aspects réglementaires sont clairement un frein dans l'obtention d'un brevet sur différents territoires, néanmoins certains offices permettent, grâce à une demande unique, de déposer un brevet dans plusieurs pays, on peut citer l'exemple de « l'Office Européen des Brevets » (OEB) qui compte 38 États membres. Pour effectuer une demande de brevet, l'inventeur va devoir renseigner l'ensemble des éléments caractérisant son invention et notamment les sources sur lesquelles il s'est appuyé (citations de brevets, publications...) et les technologies contenues dans l'invention (Barré & Laville, 1994). Une fois la demande de brevet déposée, un examinateur de l'office de brevet va étudier la brevetabilité de l'invention en effectuant notamment une recherche d'antériorité. Cette recherche s'attache à vérifier si des publications antérieures (articles, conférences ou brevets) à la demande de brevet ne pourraient pas invalider la brevetabilité de l'invention considérée. Cette procédure prend du temps et il est courant que plusieurs années s'écoulent entre le moment où la demande de brevet a été déposée et le moment où le brevet est octroyé.

² https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000019298703

Durant cette procédure, l'examineur va pouvoir également s'appuyer sur la classification internationale des brevets (CIB) dans ses recherches de technologies antérieures. La classification internationale des brevets « est un système hiérarchique de symboles indépendants de la langue pour le classement des brevets et des modèles d'utilité selon les différents domaines technologiques auxquels ils appartiennent » (source : Wipo³). Cette classification est composée de 8 sections :

- A : Nécessités courantes de la vie
- B : Techniques industrielles diverses ; transports
- C : Chimie ; métallurgie
- D : Textiles ; papier
- E : Constructions fixes
- F : Mécanique ; éclairage ; chauffage ; armement ; sautage
- G : Physique
- H : Electricité

Au sein de chacune de ces sections, on va trouver des sous-sections (code à 3 chiffres), des groupes (code à 4 chiffres) et enfin des sous-groupes (code complet de la technologie). Plus on descend dans la ramification et plus la définition de la technologie sera fine. Il est à noter qu'il existe une classification coopérative des brevets (CPC) qui est une extension de la classification internationale des brevets. Cette classification est gérée conjointement par les offices des brevets européens (EPO) et américains (USPTO) (source : EPO⁴). La classification coopérative des brevets ajoute une section supplémentaire, Y, qui regroupe les nouveaux développements technologiques.

Il apparaît qu'à travers cette classification il est possible de rechercher très précisément des technologies sur des domaines spécifiques. La littérature économique s'est grandement appuyée sur cette classification internationale des brevets pour mener de nombreuses analyses. On peut citer l'étude du portefeuille de brevets de l'entreprise Tesla réalisée par Baudry & Dumont (2020), ils ont utilisé la classification coopérative des brevets (CPC) afin de restreindre leur périmètre d'analyse aux classes technologiques ciblées par les brevets Tesla. On peut également prendre l'exemple de Contreras, Hall and Helmers (2018) qui ont étudié les « éco-brevets ». Ils

³ <https://www.wipo.int/classifications/ipc/fr/preface.html>

⁴ <https://www.epo.org/fr/classification-cooperative-des-brevets-cpc>

ont utilisé la classification internationale des brevets afin d'identifier des brevets ayant des technologies similaires aux « éco-brevets ».

Il apparaît que le brevet possède un cadre analytique propice pour étudier les technologies d'un secteur d'activité. En effet, grâce aux informations contenues dans les documents de brevet, il est possible d'étudier la localisation des inventeurs de technologies, les caractéristiques technologiques grâce la classification internationale et enfin, la diffusion des technologies.

4. Plan de la thèse

Afin d'étudier l'usage et la valeur économique des traitements Big Data, nous allons étudier l'une des technologies majeures des technologie de traitement Big Data : MapReduce. La technologie MapReduce a fait l'objet d'un dépôt de brevet auprès de l'office américain USPTO sous le nom « *System and method for efficient large-scale data processing* » par l'entreprise américaine Google en 2004 (source : patents.google⁵). Cette thèse va donc étudier la technologie MapReduce à travers trois chapitres.

Chapitre I : Emergence des technologies de traitement des données Big data – une analyse de brevets

A l'image des secteurs des biotechnologies et des nanotechnologies, une méthodologie scientométrique sur le brevet MapReduce a été appliquée pour définir l'émergence des technologies de traitement Big Data. Une base de données de 108 968 brevets a été constituée à partir des données de l'office USPTO sur la période 2004 à 2017. Afin de caractériser l'émergence des technologies de traitement Big Data, trois périmètres d'analyse distincts ont été créés : périmètre 1 : brevet MapReduce et ensemble des brevets follow-on innovation associé (brevets citant le brevet MapReduce et ceux qui citent les citant) ; périmètre 2 : ensemble des brevets ayant les mêmes classes technologiques que les brevets cités par MapReduce ; périmètre 3 : ensemble des brevets dont les résumés contiennent des mots-clés sur le traitement de données. Ce chapitre permet d'étudier l'écosystème des technologies Big Data à travers l'analyse des brevets follow-on innovation du brevet MapReduce et en utilisant comme référence d'autres technologies concurrentes ayant les mêmes spécifications techniques. Ces travaux s'inscrivent dans la littérature des études scientométriques en

⁵ <https://patents.google.com/patent/US7650331>

appliquant une méthodologie établie à l'analyse des technologies Big Data et en enrichissant le champ lexicographique par rapport aux travaux existants dans le domaine.

Chapitre II. Innovation ouverte et collaboration dans le secteur des technologies de traitement des données « big data » : propriété intellectuelle et enjeux stratégiques

Au sein de la fondation Apache, des entreprises concurrentes telles que Google ou Microsoft collaborent à l'amélioration des technologies. De plus, de nombreux brevets sur les technologies Big Data sont en accès libre sous la forme de « patent pledge », renforçant les aspects collaboratifs autour de ces technologies. Ce second chapitre s'intéresse à ces deux mécanismes d'innovation ouverte favorisant les collaborations afin de quantifier leur impact sur l'innovation au sein du secteur des technologies Big Data. Des estimations économétriques sur les citations de 32 320 brevets de l'office USPTO ont été réalisées. Cet échantillon de brevets a été divisé en deux groupes, d'une part le groupe « traité » qui est l'ensemble des brevets associés à la technologie MapReduce et à ses follow-on innovations et d'autre part, le groupe « contrôle » constitué par des technologies de traitement de données mais n'ayant aucun lien avec la technologie MapReduce. Ce second chapitre vient compléter la littérature existante sur le thème de la « coopération » en l'agrémentant avec des analyses spécifiques sur la fondation Apache et les « patent pledge ».

Chapitre 3. Impact des *patent pledge* sur la valeur des portefeuilles technologiques des entreprises

Un nombre croissant d'entreprises choisissent d'ouvrir leur technologie sous la forme de « patent pledge ». Ce chapitre se consacre à l'étude de cette ouverture et cherche notamment à comprendre si cette stratégie a permis de valoriser la propriété intellectuelle de trois entreprises leaders du secteur de l'informatique et des télécommunications que sont Google, IBM et Microsoft en étudiant l'ensemble de leurs brevets sur les technologies de traitement de données sur la période 1994-2017. Dans ce cadre, une base de données de 1 355 856 brevets a été constituée et deux types d'analyses économétriques ont été effectués : 1. Une analyse qui compare la valeur des « patent pledge » des trois entreprises à un groupe de contrôle ayant des technologies semblables ; 2. Une analyse qui compare la valeur des brevets pledge des trois entreprises aux brevets non pledge de leur portefeuille (101 878 brevets). Ce chapitre vient prolonger les études réalisées sur les « patent pledge » en les transposant au secteur de l'informatique et des télécommunications et vient ainsi enrichir la littérature existante sur les

études des propriétés intellectuelles ouvertes à l'image de celles réalisées sur les « éco-brevets » ou sur l'entreprise Tesla.

5. Références

- Barré, R., & Laville, F. (1994). La bibliométrie des brevets : une mesure de l'activité technologique. *Economie et Statistique*, 275(1), 71–81. <https://doi.org/10.3406/estat.1994.5890>
- Baudry, M., & Dumont, B. (2020). Que nous apprennent les brevets de Tesla Motors sur sa stratégie PI ? https://www.openscience.fr/IMG/pdf/iste_techinn20v5n4_6.pdf
- Brasseur, C. (2013). Enjeux et usages du big data (*Lavoisier*).
- Certes, N. (2018). Après 7 ans de procès, Apple et Samsung trouvent un accord. <https://www.lemondeinformatique.fr/actualites/lire-apres-7-ans-de-proces-apple-et-samsung-trouvent-un-accord-72205.html>.
- Chesbrough, H. (2003). The logic of Open Innovation: *Managing intellectual property*.
- Cohendet, P., Pénin, J., Farcot, M., Henri Tudor, C., & Matthieufarcot, L. (1998). Entre incitation et coordination : Repenser le rôle du brevet d'invention dans une économie fondée sur la connaissance. <https://www.researchgate.net/publication/255599313>
- Contreras, J., Hall, B., & Helmers, C. (2018). Green Technology Diffusion: A Post-Mortem Analysis of the Eco-Patent Commons. <https://doi.org/10.3386/w25271>
- Fléchaux, R. (2014). Big Data : près d'une entreprise française sur deux a un projet. <https://www.silicon.fr/big-data-entreprise-francaise-projet-95288.html>.
- Heald, P. J. (2005). A Transaction Costs Theory of Patent Law. *In Ohio State Law Journal* (vol. 66).
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*.
- Noel, M., & Schankerman, M. (2013). Strategic Patenting and Software Innovation. *The Journal of Industrial Economics*, 61(3), 481–520. <https://doi.org/10.1111/joie.12024>
- Rawat, R., & Yadav, R. (2021). Big Data: Big Data Analysis, Issues and Challenges and Technologies. *IOP Conference Series: Materials Science and Engineering*, 1022(1), 012014. <https://doi.org/10.1088/1757-899X/1022/1/012014>

CHAPITRE 1. EMERGENCE DES TECHNOLOGIES DE TRAITEMENT DES DONNEES BIG DATA – UNE ANALYSE DE BREVETS

Pour comprendre l'émergence d'un secteur d'activité dans une industrie, il est important de s'intéresser aux nouvelles technologies qui ont été développées spécifiquement pour ce secteur. Dans le cas du secteur des technologies de traitement de données, l'explosion du volume de données collectées a entraîné une nouvelle réflexion autour des techniques permettant de mieux traiter une grande quantité d'information : les données Big Data. Les caractéristiques spécifiques à ces données (« 3V » ; voir introduction) ont entraîné l'émergence de nouvelles technologies de traitement de données telles que « MapReduce » ou « Hadoop », des technologies qui ont introduit de nouveaux procédés dans le stockage, le traitement et l'analyse des données.

Afin de comprendre l'usage des technologies de traitement de données Big Data, il est donc important de s'intéresser à l'émergence de ces nouvelles technologies au sein des paysages industrielles, et de définir dans un premier temps un périmètre d'analyse pertinent.

1.1 Méthodologie permettant de définir un périmètre d'analyse

1.1.1 Intérêt de la démarche

Lors de l'étude d'une technologie, il est nécessaire de s'intéresser aux champs d'application de celle-ci. En effet, une technologie peut avoir plusieurs débouchés donc de nombreux champs d'application. Par le caractère générique de ces technologies, il n'est pas aisé de délimiter le périmètre des champs et applications. Or sans ce travail préalable, il est difficile d'étudier une industrie ou une communauté d'utilisateurs. Loin d'être spécifique aux technologies de traitement de données « Big Data », cette problématique est propre à toutes les technologies émergentes et génériques comme cela a été le cas pour les nanotechnologies ou les biotechnologies.

L'étude des nanotechnologies est un exemple intéressant pour comprendre les difficultés liées à la définition d'un périmètre d'analyse. Les nanotechnologies consistent à exploiter des matériaux de taille moléculaire (échelle de 1 à 100nm) pour créer de nouveaux produits ou processus (Meyer & Persson, 1998). Ces technologies sont notamment utilisées en physique avec les nanomatériaux qui contiennent au minimum 50% de nanoparticules, comme c'est le

cas des nanotubes de carbone dans les domaines de l'électronique (Launois P et al., 2004). En médecine, la nanomédecine est le domaine qui exploite spécifiquement les innovations introduites par les nanotechnologies. Elles permettent notamment de construire des procédés moléculaires destinés à interagir, traiter ou reconstituer un tissu ou un organe particulier dans le corps humain (Boisseau & Loubaton, 2011). Enfin des débouchés dans le domaine de l'agriculture sont également à l'étude notamment dans le domaine des engrais avec comme objectif de limiter leur nocivité (Manjunatha et al., 2016). Il apparaît que les champs d'applications des nanotechnologies sont nombreux et si on s'intéresse aux pays producteurs de ces technologies, on constate que ces technologies étaient développées historiquement par des entreprises américaines ou allemandes mais des pays émergents tels que la Chine ou l'Inde tendent également à accroître leur production scientifique (brevets et articles scientifiques) dans le domaine (Hullmann, 2007).

Le secteur des biotechnologies est également un exemple probant de l'interdisciplinarité des technologies. Les biotechnologies peuvent être définies comme un « ensemble de techniques qui visent l'exploitation industrielle des microorganismes, des cellules animales, végétales, et de leurs constituants, ou encore, en raccourci, l'exploitation « par le vivant » de matières généralement organiques » (Douzou et al., 2001). Historiquement, des techniques de biotechnologies ont été exploitées dans le domaine médicale dès le début des années 1940 aux Etats-Unis et en Grande-Bretagne avec la production de pénicilline, qui a été rendue possible grâce à l'utilisation de techniques microbiologiques. Plus récemment, c'est dans un domaine spécifique de la recherche médicale, la génétique, que s'illustrent les procédés utilisant les biotechnologies. L'utilisation de ces technologies permet la mise au point d'hormones protéiniques, d'anticorps monoclonaux et de certaines enzymes qui vont traiter directement les gènes responsables des pathologies (Stryjewska et al., 2013). Si les domaines de la médecine et de la biologie sont les champs d'application historique des biotechnologies, on constate que d'autres secteurs d'activité ont également profité de ces nouvelles technologies. L'agriculture est un des secteurs d'activité, hors médicale, qui a le plus fortement profité des techniques développées par les biotechnologies. En effet, des semences ont pu être génétiquement modifiées grâce aux techniques de biologie moléculaire afin d'obtenir des plantes ayant des propriétés spécifiques, telles que, par exemple, la résistance à des insectes nuisibles (Wieczorek, 2003). Enfin l'industrie pétrochimique, pour répondre aux enjeux environnementaux grandissants, a également développé des techniques de production utilisant des innovations issues des biotechnologies. On peut citer l'exemple de certains types de biopolymères obtenus

par fermentation de matières glucosés et qui permettent la fabrication d'emballages biodégradable (Wenda et al., 2011).

Les exemples récents des nanotechnologies et des biotechnologies mettent en évidence la capacité des technologies à être transversales à travers différents domaines d'activité. Pour comprendre l'émergence d'une nouvelle technologie, il est donc important de définir, dans un premier temps, un périmètre d'analyse adéquate pour mieux appréhender la complexité des acteurs et des industries.

1.1.2 Méthodologies pour définir un périmètre d'analyse

Selon Porter (1985), « une technologie est l'application concrète des connaissances scientifiques et techniques à la conception, à la mise au point et à la fabrication d'un produit ». L'étude des savoirs scientifiques et techniques va donc permettre de définir un périmètre d'analyse pertinent des technologies considérées. Ce type d'analyse, qui a pour objectif de mesurer et d'analyser la science, fait partie du domaine de la scientométrie. Les études scientométriques ont largement été utilisées dans la littérature économique pour quantifier l'émergence de nouvelles technologies au sein d'une entreprise, d'un secteur d'activité, ou encore d'un pays (Meyer & Persson, 1998).

La scientométrie étudie notamment deux indicateurs clés :

- Les publications : indicateur détaillant les connaissances scientifiques.
- Les brevets : indicateur décrivant les connaissances techniques.

L'analyse de ces deux indicateurs a permis de mieux comprendre et de caractériser de nouvelles technologies émergentes, à commencer par les nanotechnologies et des biotechnologies citées plus haut.

L'étude des publications dans le secteur des nanotechnologies a permis de mettre en évidence l'interdisciplinarité de ces nouvelles technologies (Meyer & Persson, 1998). L'étude de Meyer et al. s'appuie sur une approche fondée sur une recherche lexicographique avec une analyse des publications ayant le préfixe « nano » dans leur titre. Cette approche se base sur les travaux de Braun et al., (1997) qui ont utilisé cette méthodologie pour caractériser la forte émergence des nanotechnologies au début des années 1990. Les travaux de Meyer et al. complètent ceux de Braun et al. en analysant plus précisément le type de publication et leurs localisations. Sur

les types de publication, ils ont pu constater un développement important des articles sur les nanotechnologies dans les champs de la chimie et des sciences de l'ingénieur. Ils ont ainsi montré que le taux de publication des articles sur les nanotechnologies était deux fois plus important dans le domaine de la chimie que ceux des sciences physiques (le domaine historique des nanotechnologies) entre 1991 et 1996. Sur les localisations des articles, leur étude souligne le leadership de trois pays : Les Etats-Unis, le Japon et l'Allemagne. De plus, ils ont complété cette analyse de la localisation par une étude des collaborations entre les principaux pays publiant des articles sur les nanotechnologies. Cette analyse spécifique des collaborations entre pays a permis de mettre en évidence que certains pays ont tendance à avoir des partenaires privilégiés (par exemple la Chine avec l'Allemagne) tandis que d'autres collaborent avec un plus grand nombre de pays (Etats-Unis, Allemagne). Les publications permettent donc de caractériser une technologie à travers l'analyse de plusieurs critères tels que les champs d'application, la localisation ou encore les collaborations.

Le deuxième indicateur clé des études scientométriques est le brevet, et à l'instar des publications, les brevets permettent également la caractérisation des technologies émergentes. Dans le cas des biotechnologies, une analyse de 48 272 brevets spécifiques à ces nouvelles technologies sur la période 1982 à 1993, a permis de démontrer une accélération de l'utilisation de ces nouveaux procédés dans le domaine de l'agriculture à la fin des années 1980 (Banerjee et al., 2000). Ce résultat a pu être mis en évidence en comparant le nombre de brevets liés aux biotechnologies déposés dans le secteur de l'agriculture et le nombre de dépôts total dans ce secteur. Il apparaît que ce ratio a été multiplié par six au cours de la décennie des années 1980, alors que comparativement, ce ratio n'est que de trois pour les industries pharmaceutiques, secteur historique des Biotechnologies. De plus, l'étude de Banerjee et al. (2000) a également permis de montrer que certains pays étaient dépendants d'autres pays dans la production de biotechnologies. En effet, un brevet est systématiquement identifié par son pays de dépôt initial. On peut donc comparer pour un pays la production locale de brevets et les brevets d'origine étrangère sur une même technologie spécifique. Si les brevets d'origine étrangère sont supérieurs à la production locale, on peut alors supposer une dépendance du pays envers des technologies étrangères. Dans le cas des biotechnologies, l'analyse de Banerjee et al. a montré que la production domestique de brevets en Chine, Grande Bretagne, Hongrie et Australie sur les biotechnologies était largement inférieure sur la période 1982-1993 par rapport à la production de brevets étrangers, d'où une forme de dépendance de ces pays dans ce nouveau domaine technologique.

L'analyse scientométrique repose sur une méthodologie spécifique qui permet de caractériser une technologie émergente. Celle-ci s'appuie sur trois étapes distinctes :

- Etape 1 - Définition d'un corpus scientifique et technique : cette première étape consiste à étudier l'ensemble des nouvelles technologies à travers la mise en place d'un lexique scientifique et technique. Ce lexique permet notamment de répertorier tous les mots-clés associés aux nouvelles technologies étudiées.
- Etape 2 - Validation par des experts métiers : lors de cette étape, le lexique scientifique et technique est confronté à des experts métiers de ces nouvelles technologies afin de valider les différents mots-clés répertoriés et les compléter si nécessaire.
- Etape 3 - Création d'une base de données de publications et/ou de brevets spécifiques : les mots-clés validés lors de l'étape 2 vont être utilisés comme terme de recherche afin de créer une base de données spécifiques aux nouvelles technologies analysées

1.1.2.1 Définition du corpus scientifique et technique

Pour délimiter un périmètre d'analyse adéquat, il importe dans un premier temps, de définir, un champ lexical de toutes les technologies développées dans le cadre du secteur émergent que l'on souhaite étudier. Cette étape, primordiale, va permettre de lister de façon exhaustive les techniques ou procédés innovants nouvellement créés. Par sa nature empirique, cette étape est amenée à être complétée au fil du temps avec l'accroissement des connaissances scientifiques et techniques sur les technologies étudiées. Les nanotechnologies sont, une nouvelle fois, un secteur probant pour illustrer l'approche itérative de cette étape, avec l'augmentation des connaissances technologiques. En effet, l'étude de la littérature économique sur les nanotechnologies permet de constater que le corpus scientifique et technique sur cette technologie a évolué suivant trois approches (Schultz & Joutz, 2010):

- **Approche 1** : les premières études se sont basées sur une recherche textuelle en se focalisant notamment sur les termes utilisant le préfixe « nano ». (Braun et al., 1997). Cette première approche exhaustive permet de s'appuyer sur un large choix de mots-clés, mais il est tout de même nécessaire, au préalable, de retirer certains mots-clés non pertinents. Dans le cas des nanotechnologies, on peut citer des mots tels que « nanosecondes » et « nanolitres » qui ont été exclus des termes de recherche car ils ne sont pas liés aux nanotechnologies.
- **Approche 2** : avec l'accroissement des connaissances sur les nanotechnologies, de nouvelles études ont été menées afin d'élargir le champ lexical au-delà des termes

utilisant uniquement le préfixe nano (Porter et al., 2008). Des termes communément utilisés dans l'univers des nanotechnologies ont été pris en compte tels que « monocouche », « point quantique », « TEM », et « AFM ». Ces termes sont liés à la mise en application des nanotechnologies et font notamment référence à l'échelle nanométrique.

- **Approche 3** : en 2004, l'USPTO a établi une classe spécifique pour les brevets issus des nanotechnologies. Cette classification a servi de base pour de nouvelles études (Schultz & Joutz, 2010). Cette approche a pour avantage de s'appuyer sur les connaissances des examinateurs de brevets dans les technologies étudiées, et évite ainsi tout jugement subjectif de la part du chercheur.

1.1.2.2 Validation par des experts métiers

Une fois la définition du corpus scientifique et technique établie, l'étape suivante permet de confronter les termes et mots-clés référencés aux avis des experts-métiers des technologies étudiées. Afin d'être pertinent et d'éviter tout biais de subjectivité, le panel d'expert interrogé doit être le plus hétérogène possible. Au cours de leur travaux sur les termes de recherche à prendre en compte pour étudier les nanotechnologies, Porter et al., (2008) ont confronté leurs mots-clés référencés auprès de 45 scientifiques spécialistes des nanotechnologies lors d'une enquête entre février et avril 2006. Parmi ces 45 experts, 19 d'entre eux ont fourni des réponses détaillées, 13 venant de milieux académiques et 6 de milieux non académiques (industries et institution gouvernementale).

1.1.2.3 Création de la base de données à analyser

A partir des mots-clés validés par les experts métiers dans l'étape 2, il est possible d'extraire les publications ou les brevets spécifiques à une technologie grâce aux différentes sources de données. Pour les publications, les sources usuelles sont « *Web of Science* », ou « *Scopus* » pour une approche pluridisciplinaire. Certaines sources sont spécifiques à des domaines d'activité, on peut, par exemple, citer « *PubMed* » qui est spécialisé dans les publications sur les secteurs de la biologie et de la médecine. Concernant les brevets, les sources de données sont principalement des offices de brevets nationaux ou régionaux. Pour l'Europe, la principale source de données est l'office européen des brevets (« OEB ») créé en 1977 et qui regroupe 38 états membres au 1^{er} octobre 2019. Pour les Etats-Unis, c'est le bureau américain des brevets et des marques de commerce (United States Patent and Trademark Office - « USPTO ») qui

gère les demandes de brevets, c'est donc la source de données de brevets la plus pertinente pour ce territoire. Enfin au niveau international, il existe l'organisation mondiale de la propriété intellectuelle (World Intellectual Property Organization – « WIPO »). Cette institution des Nations Unies est spécialisée dans la gestion de la propriété intellectuelle, elle a été créée en 1967 et compte 191 états membres en 2020. Cette organisation est également une source intéressante de brevets qui ne se limite pas à une zone territoriale comme l'office européen ou américain des brevets et des marques de commerce.

L'étude de technologies contemporaines telles que les nanotechnologies ou les biotechnologies ont permis de mettre en évidence une méthodologie spécifique permettant de caractériser des technologies émergentes. Cette méthodologie est basée sur l'analyse des connaissances scientifiques et techniques au travers de l'analyse des publications et des brevets.

Le paragraphe suivant expose l'application de la méthodologie scientométrique à l'étude des technologies de traitement Big Data avec dans un premier temps, une revue de la littérature existante sur le sujet, puis dans un deuxième temps, nous détaillerons la méthodologie retenue pour caractériser ces nouvelles technologies de traitement de données.

1.2 Définition du périmètre d'analyse des technologies de traitement de données Big Data

1.2.1 Revue de la littérature

Les études sur les technologies de traitement de données « Big data » très relativement récentes. Une première analyse a été initiée en 2014 par le UK Intellectual Property Office Informatics Team (*Big Data - A Patent Overview*, 2014). Ce rapport s'appuie sur une recherche des brevets « Big Data » grâce à des mots-clés appliqués à la base de données de brevets « *The Thomson Reuters World Patent Index* ».

Si on s'intéresse à la stratégie de recherche de cette étude, on constate que le lexique technologique retenu est exhaustif, il contient, en effet, des mots-clés spécifiques aux technologies Big data (Hadoop, Yarn, Aster) mais également des techniques de traitement de données (« data mining », « neural network », « artificial intelligence ») et des mots-clés génériques liés à l'analyse de données (« open data », « data warehouse », « software »). Grâce à cette stratégie de recherche, 22 000 brevets ont pu être extraits et analysés entre 2004 et 2013.

Cette première étude a permis de mettre en évidence une accélération des dépôts de brevets sur les technologies de traitement de données « Big Data » d'environ 48% entre 2011 et 2012. De plus, il a été établi que les dépôts de brevets sont localisés aux Etats-Unis (un peu plus de 50% des dépôts), suivi de la Chine (environ 30%). Enfin, la principale entreprise déposant de brevets sur ces technologies est la société américaine IBM (International Business Machine) avec notamment un pic de dépôts en 2012 et 2013. Cette étude met également en lumière que la production de ces nouvelles technologies aux Etats-Unis se concentre sur les entreprises IBM, Oracle, Hewlett-Packard, Microsoft et SAP ; ce sont des sociétés américaines, à l'exception de SAP, unique entreprise européenne de ce Top 5.

Si on regarde les résultats en Europe et plus particulièrement au Royaume Unis, pays d'origine de l'étude, il apparaît que les dépôts de brevets britanniques en matière de technologies dites « Big Data » représentent 1.2% des dépôts mondiaux. En termes de localisation, la majorité des inventeurs britanniques sur ces technologies travaillent pour des entreprises britanniques. Enfin, on constate que l'activité de brevetage sur ces nouvelles technologies de traitement de données ne cesse d'augmenter entre 2006 et 2013 au Royaume Unis, une croissance également constatée dans plusieurs pays Européens notamment en France et Allemagne.

L'article de Singh et al., (2015) complète cette première étude anglo-saxonne en étudiant les publications en rapport avec les technologies de traitement de données Big Data en exploitant les bases de données « *Web of Knowledge* » (« *WOK* ») et « *Scopus* » sur la période 2010 et 2014. En utilisant uniquement le terme de recherche « *Big Data* », les auteurs ont pu collecter 1415 et 6810 publications (articles, livres, enquêtes...) spécifiques à ces nouvelles technologies de traitement de données dans « *WOK* » et « *Scopus* » respectivement. Leur analyse a permis de mettre en évidence un fort dynamisme de la recherche sur les technologies Big Data entre 2010 et 2014 tant en termes de croissance du volume des publications et le nombre d'auteurs qu'ils analysent comme un indicateur du dynamisme des collaborations.

Si on s'intéresse plus particulièrement aux revues qui ont publié les documents collectés, on constate que ce sont trois revues spécialisées dans l'informatique et l'électronique qui sont les plus contributrices (Computer, Plos One, Future Generation Computer Systems the International Journal of Grid Computing and Escience). Néanmoins, on note que les deux revues arrivant à la suite de ce Top 3, sont d'une part, une revue spécifique au domaine de la santé (Health Affairs) et d'autre part, une revue scientifique généraliste (Nature). On remarque

également que les universités sont les institutions qui publient la majorité des articles sur les technologies de traitement de données Big Data sur cette période. En effet, le Top 5 pour les publications collectées dans WOK est composé de : Harvard university, University of London, Massachusetts Institute of Technology, Stanford University, University of California Los Angeles. La seule entreprise présente dans le Top 15, est la société IBM, déjà présente dans la première étude sur les technologies Big Data publiée par UK Intellectual Property Office Informatics Team. On note également que sur les 15 premiers contributeurs, l'université de Londres est la seule université non américaine.

Concernant les disciplines qui font l'objet de ces publications, on constate, assez logiquement, que l'informatique et l'électronique représentent plus de 65% des publications. Ces domaines techniques génèrent donc environ deux tiers des publications, le dernier tiers, soit une part non négligeable, est initié par des secteurs « autres » avec notamment la biologie (14%), la médecine (9%) et le management (8%).

Ce dernier point montre une tendance pluridisciplinaire dans l'utilisation des technologies de traitement de données Big Data, une tendance également observée dans d'autres secteurs émergents comme les nanotechnologies et les biotechnologies.

Les travaux de Sheng et al., (2017) nous éclairent plus particulièrement sur l'utilisation des technologies de traitement de données Big Data dans un domaine qui n'est pas historiquement un domaine d'application de ces technologies, à savoir le management. La méthodologie employée pour cette étude est globalement similaire aux travaux précédemment étudiés. Les auteurs ont effectué une recherche de mots-clés générique (« *Big data* », « *Big data analytics* », « *data analytics* », « *web analytics* » ...) sur différentes sources de données de publications telles que « *Business Source Complete* », « *Google Scholar* », « *Scopus* », « *Informa* », « *Sage* », « *Wiley* » et « *JSTOR* ». Il est intéressant de noter qu'afin de cibler des articles pertinents, les auteurs ont utilisé des associations de mots-clés dans leur méthodologie de recherche (par exemple « *Big data* » et « *text analysis* »). A travers cette stratégie de recherche, les auteurs ont pu collecter 284 articles traitant de l'application des technologies de traitement de données Big Data dans le domaine management sur la période 2005-2017.

Sur ces thématiques liées au management, les auteurs ont pu mettre en évidence que le nombre d'articles croit fortement sur les périodes très récentes. En effet, les articles publiés entre 2015 et 2017 représentent 42% de l'échantillon étudié, ce qui confirme que la diffusion de ces

nouvelles technologies de traitement de données en dehors de la sphère des technologies de l'information et de la communication est contemporaine. En termes de répartition, il a été constaté que la majorité des articles collectés traite des sujets marketing (55%), puis des thèmes de l'organisation (17%) et des techniques de management (11%). Enfin, les auteurs ont remarqué que les articles portant sur le marketing traitent en grande majorité de l'analyse du comportement des consommateurs grâce à l'exploitation des données d'internet mobile qui ont comme principale caractéristique d'être des données géolocalisables produites en temps réels.

Les premières études sur les technologies de traitement de données Big Data ont mis en évidence plusieurs caractéristiques intéressantes. Tout d'abord, ces technologies sont majoritairement présentes dans leur domaine d'activité historique : les technologies de l'information et de la communication. Cependant des études récentes ont pu mettre en évidence une diffusion de ces nouvelles technologies dans de nouveaux secteurs d'activité tels que le management. Ensuite, il apparaît que ces technologies sont produites essentiellement sur le territoire américain. En effet, d'après l'étude britannique de 2014, 50% des dépôts de brevets sont localisés aux Etats-Unis, et la principale société déposante de brevets est IBM, une société américaine. Enfin, une première étude chinoise a montré que le dynamisme de la recherche et les collaborations sur ces nouvelles technologies de traitement de données étaient élevés.

La section suivante présente la méthodologie appliquée pour construire une base de données des technologies du traitement des « Big Data » à partir des brevets de l'USPTO et construite autour de la technologie « MapReduce » de Google. Cette analyse doit permettre de définir l'écosystème autour des technologies du traitement des Big Data. Cette base sera exploitée dans les chapitres 2 et 3.

1.2.2 Présentation de la méthodologie utilisée

Les premières études sur les technologies de traitement de données Big Data se sont largement appuyées sur des analyses de types scientométriques afin de mettre en évidence les premiers éléments caractéristiques de ces nouvelles technologies. A l'instar des secteurs des biotechnologies et des nanotechnologies, des méthodologies basées sur des recherches lexicographiques des technologies les plus couramment utilisées pour traiter les données Big Data ont été utilisées comme première base d'analyse. Parmi l'ensemble de ces technologies, deux technologies sont systématiquement citées comme les technologies fondatrices des

traitements de données de type « Big Data », ce sont les technologies « MapReduce » et « Hadoop ». Ces deux technologies ont posé les bases des traitements utilisant un processus de parallélisation des données afin de pouvoir traiter efficacement une grande quantité (Dittrich et al., 2014).

1.2.2.1 Présentation de l'algorithme de traitement « MapReduce »

L'algorithme de traitement MapReduce a été exposé pour la première fois en 2004 par deux ingénieurs de la société Google : Jeffrey Dean et Sanjay Ghemawat dans un document s'intitulant : « *MapReduce : Simplified Data Processing on Large Clusters* » (2004)⁶. Ce document met en évidence le principe de parallélisation du traitement des données à travers deux grandes opérations spécifiques ; l'opération « Map » et l'opération « Reduce ».

Pour comprendre le fonctionnement de l'algorithme, prenons l'exemple de la recherche de mots dans un texte :

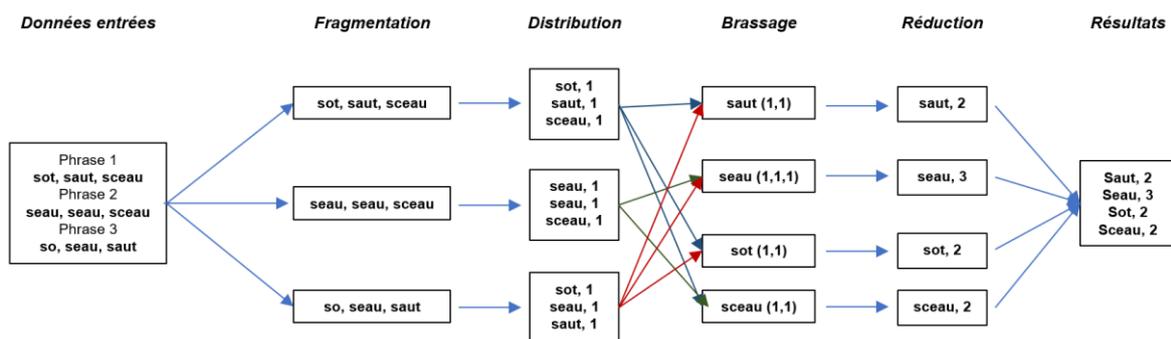


Figure 1.1 principe de l'algorithme MapReduce (Charles Xavier, 2023)

L'algorithme fonctionne en 4 étapes distinctes :

1. Fragmentation : les données analysées sont fragmentées en lot
2. Distribution : chaque lot est distribué afin de lister les valeurs importantes
3. Brassage : les valeurs importantes sont triées
4. Réduction : chaque valeur importante obtenue dans l'étape de « distribution » fait l'objet d'une agrégation

L'algorithme MapReduce permet de « paralléliser » le traitement de données. Grâce à ce principe, cet algorithme possède les propriétés suivantes (Dean & Ghemawat, 2004) :

⁶ <https://static.googleusercontent.com/media/research.google.com/fr//archive/mapreduce-osdi04.pdf>

- **Simplicité de programmation** : toutes les principales fonctionnalités sont directement implémentées dans l'algorithme (la parallélisation, la tolérance aux pannes, l'optimisation...)
- **Extensibilité** : il est possible de multiplier le nombre de machines afin d'accélérer les traitements.
- **Flexibilité** : de nombreuses problématiques de traitement de données peuvent être résolues via l'utilisation de l'algorithme (datamining, machine learning, webcrawling...).

Ces propriétés ont rapidement fait de l'algorithme MapReduce une des technologies fondatrices des technologies de traitement des données Big Data. Entre 2006 et 2015, on a pu dénombrer 6481 documents publiés à travers différents canaux (conférences, journaux, livres, brevets...) qui font référence à l'algorithme MapReduce (Hashem et al., 2016). Ces publications sont majoritairement chinoises (33%) ou américaines (28%) et se concentrent principalement sur le domaine de l'informatique (60%).

La technologie MapReduce a fait l'objet d'un brevet qui a été déposé le 18 juin 2004 par ses deux inventeurs (Jeffrey Dean et Sanjay Ghemawat) à l'office des brevets américain (United States Patent and Trademark Office) sous le titre « *System and method for efficient large-scale data processing* » (numéro de publication : 7,650,331).

1.2.2.2 Méthodologie de création des périmètres

Afin de créer des périmètres d'analyses spécifiques aux technologies de traitement des données de type « Big Data », nous allons dans un premier temps étudier la trajectoire du brevet MapReduce comme décrit dans le Figure 1.2.

Nous allons ainsi isoler :

- Les brevets citant directement MapReduce au nombre de 185 au 23/04/2020
- Les brevets que MapReduce cite sont au nombre de 9

- Une analyse de ces brevets permet d’avoir une première vision de la trajectoire technologique du brevet MapReduce, à travers l’étude de la classification coopérative des brevets (CPC⁷) comme indiqué dans la Figure 1.2.

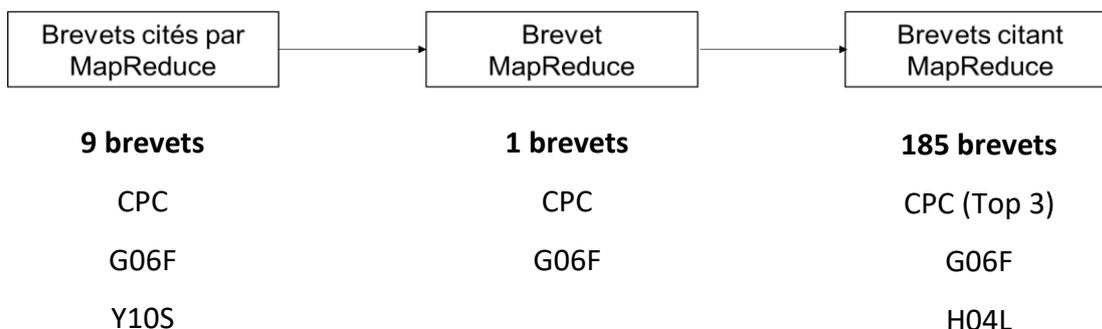


Figure 1.2 Trajectoire du brevet MapReduce : nombre de brevets et classification CPC⁸

L’analyse des brevets citant le brevet MapReduce va nous permettre de mettre en évidence tous les « *follow-on innovations* » de la technologies MapReduce, c’est-à-dire les brevets qui se sont appuyés sur cette technologie pour se développer. Nous allons procéder en une succession d’itérations afin de pouvoir extraire l’ensemble de ces follow-on innovations de la technologie MapReduce comme indiqué dans la Figure 1.3 :

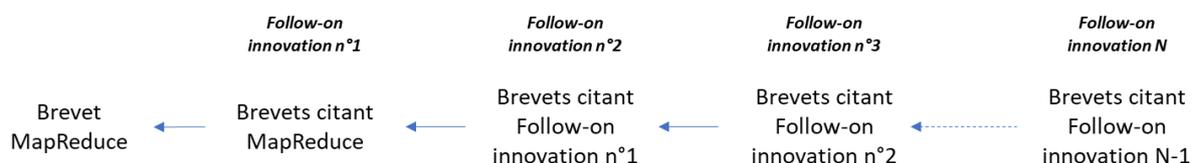


Figure 1.3. Principe d’extraction des « follow-on innovations » du brevet MapReduce

Ce « fil technologique » nous permet d’identifier tous les brevets citant directement ou indirectement la technologie « MapReduce » à différents niveaux : « brevets citant directement le brevet MapReduce » ; « brevets citant les follow-on innovations n°1 » ; etc.... L’ensemble

⁷ « Cooperative Patent Classification » est une extension de la classification internationale des brevets (CIB) et maintenue conjointement par les offices européen (EBO) et américain des brevets (USPTO). Elle est divisée en 9 sections, et sous-divisés en classes, sous-classes, groupes et sous-groupes. <https://www.epo.org/en/searching-for-patents/helpful-resources/first-time-here/classification/cpc>

⁸ G06F : traitement électrique de données numériques ; H04L : transmission d’information numérique ; G06N : systèmes de calculateurs basés sur des modèles de calcul spécifiques ; Y10S : concepts techniques couverts par d’anciens recueils de références croisées de l’USPC

de ces brevets vont constituer notre premier périmètre d'analyse avec des brevets affiliés à la technologie MapReduce et par conséquent à une technologie de traitement de données Big Data.

Dans un deuxième temps, nous allons nous intéresser aux brevets que MapReduce cite. Ces brevets font donc référence à des technologies antérieures à MapReduce. En répertoriant les différentes classes technologiques selon la classification CPC, nous allons pouvoir définir un deuxième périmètre d'analyse avec des brevets ayant des technologies de traitement de données antérieures à MapReduce.

Nos deux premiers périmètres sont constitués de brevets ayant une filiation avec la technologie MapReduce grâce à l'étude des citations. Afin de pouvoir réaliser une analyse plus exhaustive, il est nécessaire de compléter ces deux premiers périmètres par un périmètre plus générique construit à partir de brevets associés à des technologies traitement de données au sens large. A travers ce périmètre, il sera alors possible d'identifier des technologies de traitement de données plus récentes que la technologie MapReduce, et potentiellement concurrentes à celle-ci.

Pour constituer ce périmètre, nous allons utiliser une méthode fondée sur une recherche lexicographique de mots-clés spécifiques, cette méthode est utilisée usuellement dans les analyses de brevets comme nous avons pu le voir dans la revue de littérature.

Le premier champ lexicographique utilisé, dans un premier temps, est celui retenu par l'étude UK Intellectual Property Office Informatics Team (*Big Data - A Patent Overview*, 2014). Cette étude avait pour objectif de faire une première analyse des technologies de traitement Big Data grâce à une analyse des brevets collectés sur la base de données « *The Thomson Reuters World Patent Index* ».

Les mots-clés pour cette recherche lexicographique étaient les suivants :

“big data”, Hadoop®, Yarn, Aster®, Datameer®, FICO® Blaze, Vertica®, Platfora®, Splunk®, MapReduce, “open data”, “data warehous”, informatic*, “data mine?”, “data mining”, simulate*, model*, analy*, “artificial intelligence”, “neural network*”, “distributed *”, (cluster*, cloud*, grid?) [within 3 words of] (based, comput*, server?, process*, software, application), croudsourc*, “crowd sourc*”, “massively parallel process*”, “massively parallel, software”, “massively parallel database?”, “distributed process*”, “distributed server?”, “distributed quer*”, “distributed database?”, “massive data*

Après une première analyse, il s'est avéré que ce champ lexicographique possédait certains mots-clés trop génériques (simulate*, model*, analy*,"distributed*") qui identifiaient des brevets non spécifiques aux technologies de traitement de données.

En s'appuyant sur ces premiers travaux, le champ lexicographique suivant a été adopté :

Algorithms, Apache Cassandra, Apache Hive, Apache Pig, Apache Software, Apache Software Foundation, Apache Spark, Apache®, artificial intelligence, Aster data, big data, Cloud, Cloudera Impala, "crowd sourc", "data analy*", "data min*", "data predictive", "data process*", "data quer*", "data structures", "data warehous*, database distributed database, distributed process, "distributed quer*", distributed server, Hadoop, HDFS, Hortonworks Data Platform, MapReduce, massive data, massively parallel process, MongoDB, neural network, NoSQL, open data, parallel process, Yarn*

Plusieurs apports ont été effectués sur ce champ lexicographique par rapport aux travaux établis par the UK Intellectual Property Office Informatics Team en 2014. Le premier a été d'ajouter le mot-clé « *Algorithms* », qui est un mot-clé très fréquemment associé aux technologies de traitement de données (Hashem et al., 2016). Le deuxième est d'avoir systématiquement associé le mot « *Apache* » (en référence à la fondation Apache qui exploite les principales technologies de traitement de données Big Data) à toutes les technologies Big Data usuelles (« *Hive* », « *Pig* »...) afin de bien cibler des brevets faisant références à des technologies de traitement de données. Enfin le dernier apport a été d'associer des mots-clés génériques (*analyse, process, predictive...*) avec le terme « *data* » afin d'éviter de retenir des brevets non spécifiques aux technologies de traitement de données. Ainsi, tous les brevets contenant dans leur résumé les mots-clés retenus ci-dessus vont constituer le troisième périmètre d'analyse, ce sont des brevets faisant référence à des technologies de traitement de données au sens large.

Pour analyser les technologies de traitement de données, nous allons donc nous appuyer sur les **trois périmètres d'analyses suivants** (Figure 1.4).

Ces trois périmètres vont être appliqués à l'ensemble des brevets USPTO qui est l'office historique de dépôt du brevet de la technologie MapReduce.

La base de données utilisée par la suite est extraite du site internet www.patentview.org, site officiellement affilié à l'USPTO et qui inclut 7 252 172 brevets déposés entre avril 1945 et décembre 2017.

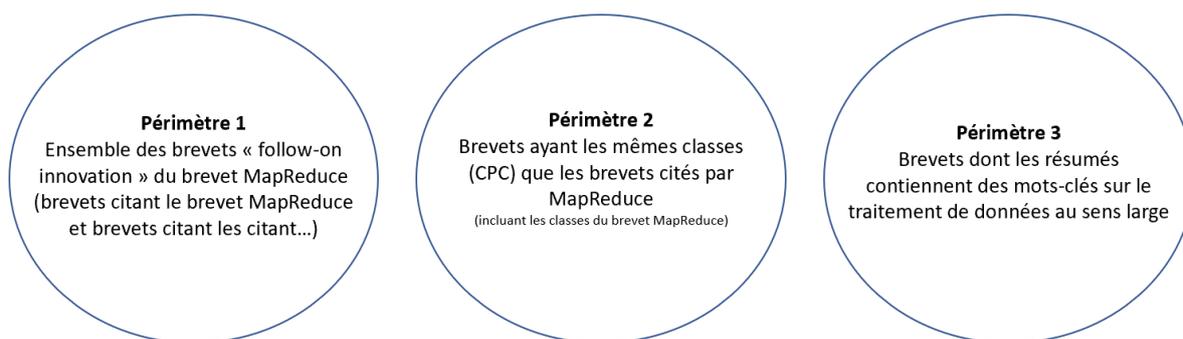


Figure 1.4. Définition des trois périmètres d'analyse

1.2.2.3 Présentation des trois périmètres d'analyse

Les trois périmètres sont constitués de 108 968 brevets sur une période s'étalant de 1994 à 2017. Afin de conserver une homogénéité dans les technologies étudiées, les brevets antérieurs à 1994, donc antérieurs de dix ans au dépôt de la technologie MapReduce n'ont pas été retenus dans les analyses effectuées.

Le nombre total de brevets retenus pour cette analyse représente 1.9 % du total de brevets déposés à l'UPSTO entre 1994 et 2017.

Ces 108 968 brevets sont structurés de la manière suivante :

- **Périmètre 1** : les brevets sur des technologies de traitement de données ayant une filiation avec la technologie MapReduce représentent 2 446 brevets sur la période 2004 à 2017, soit 2 % du total des brevets étudiés
- **Périmètre 2** : les brevets sur des technologies de traitement de données ayant une filiation avec des technologies référencées pour le développement de la technologie MapReduce représentent 38 061 brevets sur la période 1994 à 2017, soit 35 % du total des brevets étudiés
- **Périmètre 3** : les brevets sur des technologies de traitement de données au sens large représentent 68 461 brevets sur la période 1994 à 2017, soit 63 % du total des brevets étudiés

On constate clairement que les périmètres 2 et 3 sont surreprésentés avec respectivement 34% et 63% du total des brevets étudiés. Ce résultat est cohérent compte tenu de la méthode de construction de ces périmètres, basée d'une part sur les classes de CPC (périmètre 2) et d'autre part, sur une recherche lexicographique (périmètre 3).

Le périmètre 1 qui ne représente que 3% du total des brevets étudiés est un périmètre ciblé sur la filiation de la technologie MapReduce.

De plus, on constate qu'un certain nombre de brevets sont présents sur plus d'un périmètre, c'est-à-dire qu'ils combinent plusieurs caractéristiques d'un périmètre. Ces brevets présents sont définis comme des brevets « communs », ils sont au nombre de 9 132 soit 8.1 % du total des trois périmètres étudiés. Leur distribution est la suivante :

- Brevets communs entre les périmètres 1 et 2 : 130 brevets
- Brevets communs entre les périmètres 1 et 3 : 190 brevets
- Brevets communs entre les périmètres 2 et 3 : 8 649 brevets
- Brevets communs entre les périmètres 1, 2 et 3 : 59 brevets

La Figure 1.6 synthétise l'ensemble des résultats des trois périmètres ainsi que les brevets « communs » :

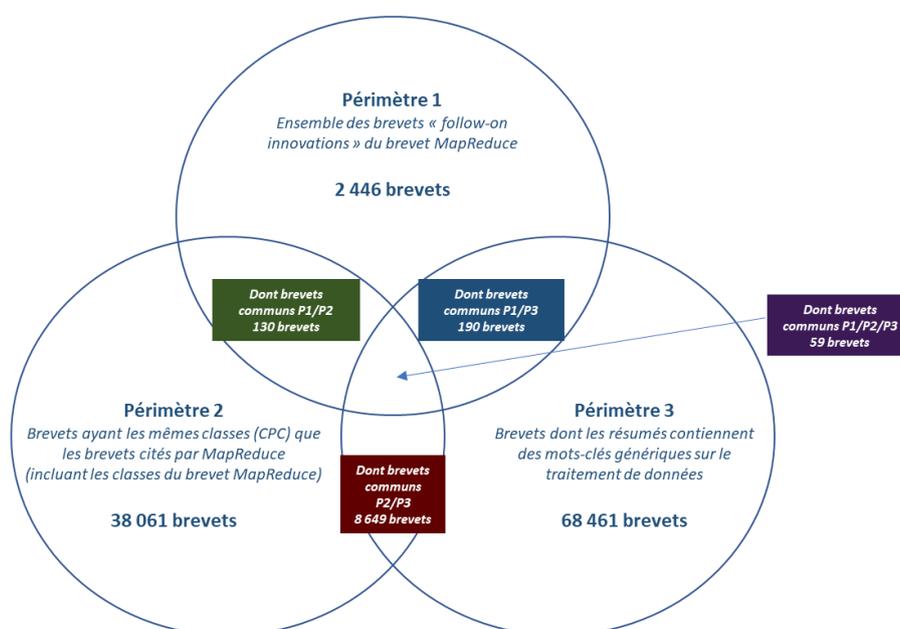


Figure 1.5. Synthèse des trois périmètres d'analyse

108 968 brevets étudiés sur la période 1994 – 2017

Source : données USPTO collectées sur patentview.org

Le tableau 1.1 dénombre la totalité des brevets « communs » par périmètre

Tableau 1.1 : nombre de brevets « communs » par périmètre

	Périmètre 1	Périmètre 2	Périmètre 3
Nombre de brevets "communs"	379	8 838	8 998
Nombre total de brevet sur le périmètre	2 446	38 061	68 461
Taux de brevets "communs" sur le périmètre	15.5%	23.2%	13.0%

Les périmètres 2 et 3 représentant 98% de la totalité des brevets étudiés, il est assez logique de retrouver la majorité des brevets « communs » sur ces deux périmètres (8 649 brevets présents sur le périmètre 2 et 3). Néanmoins, il est intéressant de constater que le périmètre 1 qui ne représente que 2 % de la totalité des brevets analysés a un taux de brevet « communs » supérieur au périmètre 3, respectivement 15.5 % contre 13 %.

Tableau 1.2 Structure du périmètre 1 avec brevets « communs »

Niveau de « follow-on innovations »	Structure du périmètre 1	Brevets « communs » P1 et P2	Brevets « communs » P1 et P3	Brevets « communs » P1, P2 et P3
Brevet citant directement				
MapReduce (Brevets N1)	8%	19%	13%	27%
Brevets citant N1 (brevets N2)	39%	25%	43%	36%
Brevets citant N2 (brevets N3)	37%	52%	38%	31%
Brevets citant N3 (Brevets N4)	7%	5%	6%	3%
Brevets citant N4 (Brevets N5)	2%			2%
Brevets citant N5 (Brevets N6)	2%			
Brevets citant N6 (Brevets N7)	1%			
Brevets citant N7 (Brevets N8)	1%			
Brevets citant N8 (Brevets N9)	2%			
Brevets citant N9 (Brevets N10)	2%			
Total brevets	2 446	130	190	59

Cette forte présence des brevets « communs » dans le périmètre 1 nous amène à nous interroger sur les caractéristiques de ces brevets « communs » et notamment à quel niveau de « follow-on innovations » ces brevets se situent. En effet, le périmètre 1 identifie tous les brevets ayant une filiation avec la technologie « MapReduce » à différents niveaux de citation (voir Figure 1.4). Plus un brevet a un lien de citation proche de la technologie MapReduce et plus ce brevet est technologiquement proche de celui-ci. Par conséquent, si les brevets « communs » du périmètre 1 se situent au niveau 1 de citation, il sera alors logique de retrouver un certain nombre de ces brevets dans le périmètre 2 et 3. En effet, ces périmètres sont des périmètres « génériques » sur l'ensemble des technologies de traitement de données.

L'analyse des brevets « communs » du périmètre 1 est détaillée dans le tableau 1.2. Le premier constat à souligner concerne la structure du périmètre 1 qui se concentre majoritairement sur les brevets :

- N2 : brevets citant directement le brevet MapReduce
- N3 : brevets citant les brevets N2

Ces brevets N2 et N3 représentent 76 % de la totalité des brevets du périmètre 1. Au-delà de quatre niveaux (brevet N5), la représentativité des brevets dans la structure du périmètre 1 chute à moins de 3% par niveau de citation. Il apparaît que les brevets les plus cités sont donc les brevets qui sont « proches » de la technologie « MapReduce ».

Si on s'intéresse aux brevets communs, on constate que les brevets présents sur les périmètres 1 et 2 sont majoritairement des brevets N3 (52 %) tandis que les brevets communs aux périmètres 1 et 3 sont des brevets N2 (43 %), et il en est de même pour les brevets présents sur les trois périmètres (36 % pour les brevets de type N2).

A l'image de la structure du périmètre 1, les brevets communs sont donc des brevets « follow-on innovations » proches de la technologies MapReduce. La surreprésentation des brevets communs sur ce périmètre s'explique donc par leur proximité forte avec la technologie MapReduce, une des technologies fondatrices des technologies de traitement Big Data.

La section suivante présente une analyse approfondie des trois périmètres retenus en nous intéressant systématiquement à un critère spécifique.

1.3 Analyse des trois périmètres

1.3.1 Etude de la temporalité

Dans cette partie, nous allons caractériser l'évolution annuelle du nombre de brevets à travers les trois périmètres. Une analyse pour chaque périmètre sera effectuée afin de déterminer si des tendances technologiques se distinguent.

La Figure 1.6 détaille l'évolution du nombre de brevets sur l'ensemble des périmètres sur la période 1994 – 2017 (les années suivantes étant incomplètes au moment de la collecte).

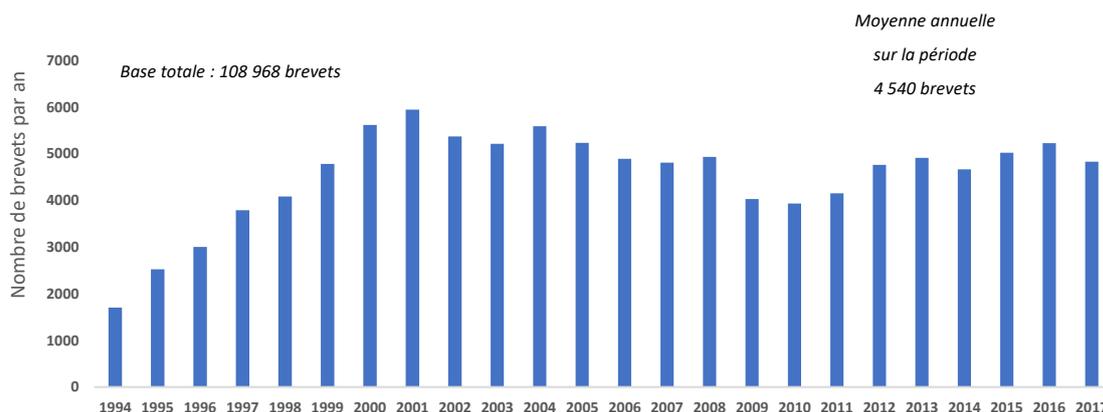


Figure 1.6. Nombre de brevets par an déposés sur l'office USPTO – ensemble des trois périmètres

Sur la période analysée, le nombre moyen de brevets déposé est de 4 540 brevets par an. Trois périodes peuvent être distinguées :

- **1994 - 2001** : phase de croissance forte des brevets sur les technologies de traitement de données
- **2002 - 2009** : phase de stagnation puis de baisse à partir 2005
- **2010 - 2017** : phase de retour de la croissance, plus mesurée que sur la période 1994-2001

Figure 1.7 permettent d'affiner l'analyse avec un détail par périmètre.

Les premières tendances mises en évidence avec l'analyse générale sont confirmées avec le détail par périmètre et nous permet d'être plus précis dans notre étude. En effet, la croissance observée sur la période 1994-2001 est concentrée sur le périmètre 2 (brevets ayant les mêmes classes CPC que les brevets cités par MapReduce) avec un taux de croissance moyen sur la période de +31 % par an contre +14 % pour le périmètre 3 (brevets dont les résumés contiennent des mots-clés sur le traitement de données au sens large).

Plus globalement, on constate que le périmètre 3 est en croissance continue sur la période analysée (1994 à 2017) même si cette croissance tend à se ralentir sur la dernière décennie. Cependant, on note clairement un ralentissement du nombre de brevets déposés sur le périmètre 2 à partir de 2004. En effet, sur ce périmètre le nombre moyen de brevets déposés entre 2002 et 2009 était de 2024 brevets par an, et ce nombre chute à 955 brevets sur la période 2010-2017 soit une baisse de -53 %.

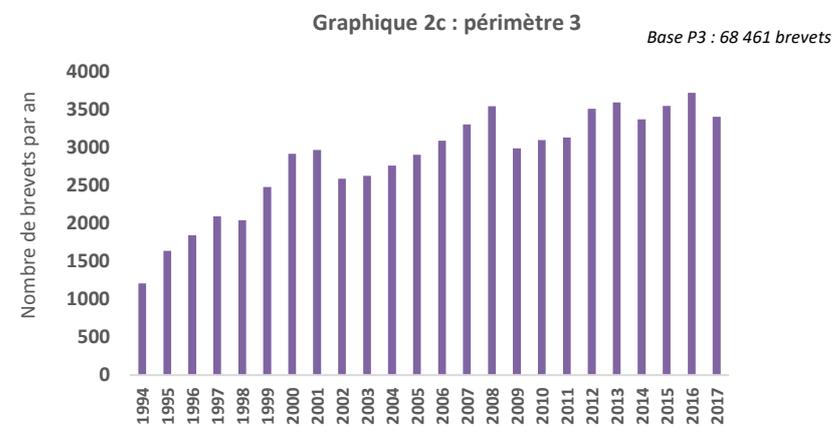
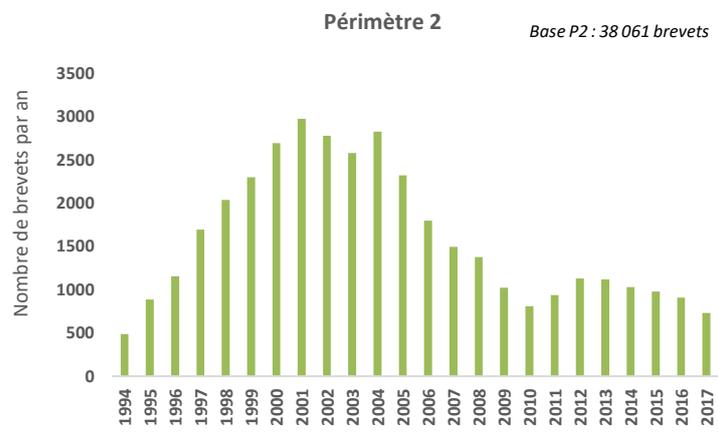
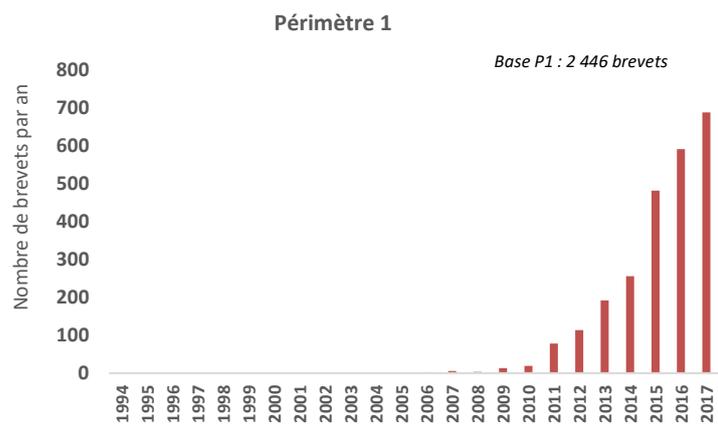


Figure 1.7 Périmètres

Concernant le périmètre 1, constitué des brevets « follow-on innovations » de la technologie MapReduce, il connaît une croissance forte à partir de l'année 2010. Il est à noter que l'année 2010 est l'année où le brevet MapReduce a été accordé à ses déposants (l'année de dépôt du brevet étant 2004).

La Figure 1.8 s'intéresse plus particulièrement à la période 2010 à 2017 avec le taux d'accroissement par an du nombre de brevet déposé sur les trois périmètres.

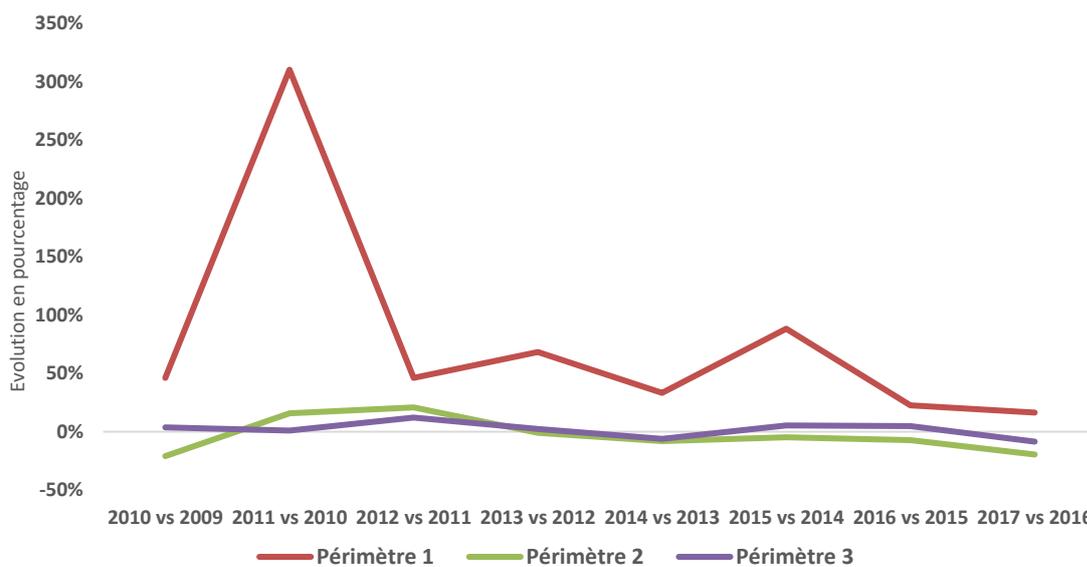


Figure 1.8. Taux d'accroissement du nombre de brevets déposés par an à l'office USPTO

Cette Figure 1.8 montre clairement que la dynamique d'accroissement du nombre de brevets déposés par an se situe sur le périmètre 1 constitué des « follow-on innovations » de la technologie MapReduce. Sur cette période analysée, le taux d'accroissement moyen est de +79 % pour le périmètre 1, contre -3.1 % et +1.8 % respectivement sur le périmètre 2 et 3.

Cette analyse de la temporalité permet donc de mettre en évidence trois périodes d'analyses distinctes :

- **1994-2001 : période « pré-Big Data »** : période où la demande en technologie de traitement de données est forte notamment avec l'émergence des nouvelles technologies de l'information et de la communication (NTIC) dans le courant des années 1990. Ce point est caractérisé par la forte croissance du nombre de brevets déposés sur le périmètre 2, à savoir les brevets ayant les classes CPC : G06F : traitement électrique de données numériques et

Y10S : regroupement de techniques diverses cité par MapReduce. Ces classes sont des classes dont les fonctions sont productives.

- **2002 – 2009 : développement des technologies Big Data** : période où l'offre en technologie de traitement de données reste forte mais l'émergence d'internet conduit à de nouveaux besoins plus spécifiques dans le traitement d'une volumétrie de données toujours plus importante. La première technologie Big Data identifiée est la technologie MapReduce qui fait l'objet d'un dépôt de brevet par la société Google en 2004. Ces nouveaux besoins spécifiques dans le traitement d'une importante volumétrie de données entraînent une stagnation des dépôts de brevets sur les périmètres 2 et 3, périmètres où les technologies de traitements sont des technologies majoritairement génériques telles que le langage SQL, technologies qui ne sont plus adaptées pour le traitement des données Big Data. On note d'ailleurs une forte baisse des dépôts de brevet sur le périmètre 2 à partir de 2005, ce qui indique que ces technologies n'étaient plus appropriées pour les nouveaux besoins dans le traitement des données « en masse ».

- **2010 – 2017 : Mise en application de l'offre de technologies Big Data** : période caractérisée par une offre technologique forte centrée sur des technologies permettant le traitement des données Big Data. Ce point est caractérisé par un taux de croissance du nombre de brevets du périmètre 1 élevé sur cette période : +79 % en moyenne sur la période comme indiqué plus haut.

1.3.2 Étude des principaux déposants

Dans cette section, nous allons nous intéresser aux principaux déposants de brevets sur des technologies de traitement de données. A l'instar des précédentes analyses, nous commençons par une analyse sur l'ensemble des périmètres avant de présenter chaque périmètre séparément. Sur l'ensemble des trois périmètres, on dénombre 20 170 organisations qui ont déposé 114 191 brevets (ce chiffre incluant les co-brevets), soit en moyenne 5.6 brevets par organisation.

La Figure 1.9 présente le Top 50 des entreprises déposantes sur la période 1994 – 2020. Le haut de la Figure met en évidence le Top 20 des déposants qui représentent 40 707 brevets soit environ 35 % de la totalité des brevets analysés sur les trois périmètres.

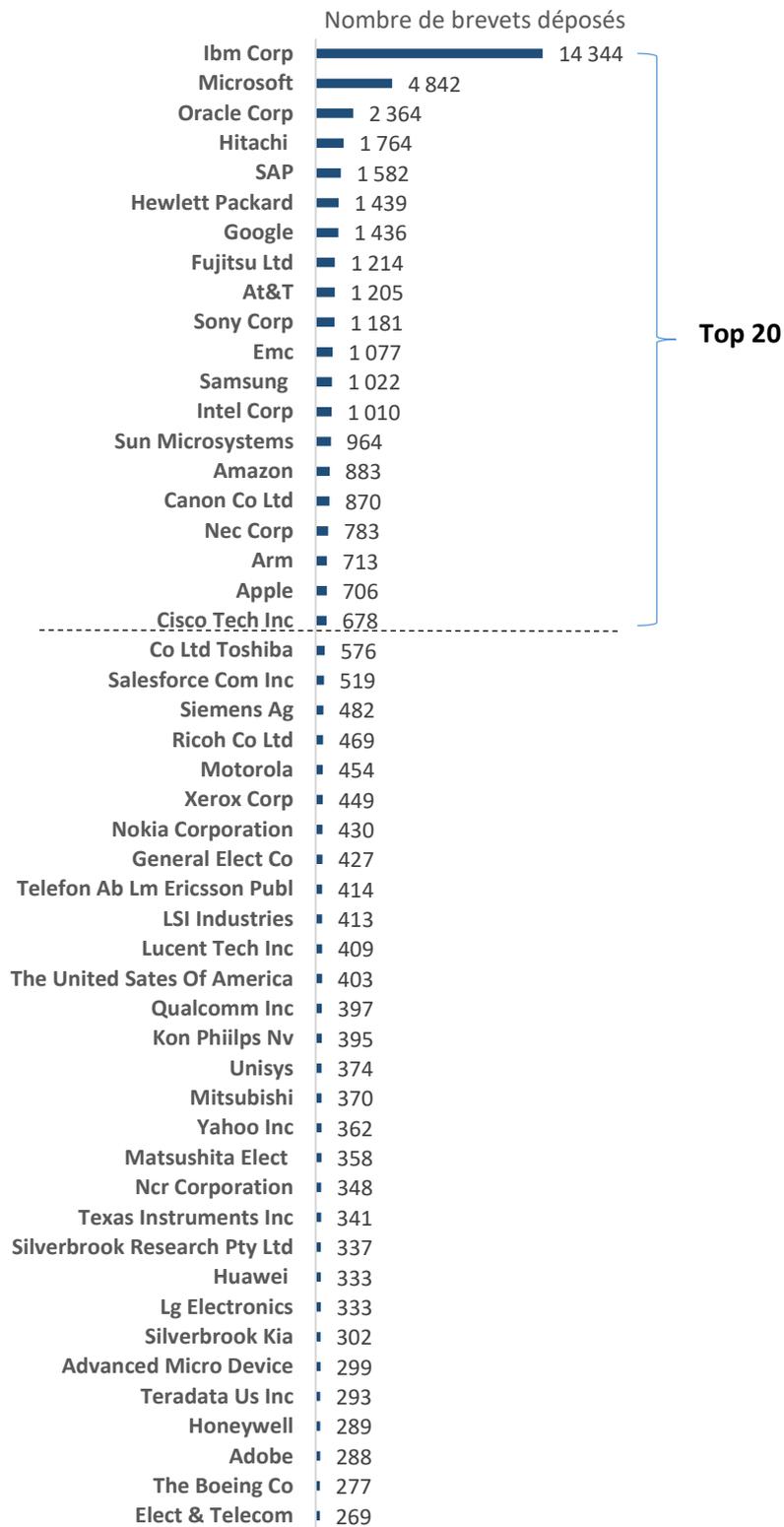


Figure 1.9. Top 50 des déposants de brevets sur la période 1994-2020

Ensemble des trois périmètres

Il apparaît que moins de 1 % des déposants produisent 35 % des brevets déposés, on est donc sur un secteur fortement concentré. La société IBM corporation domine très largement ce Top 20 avec 14 535 brevets déposés sur la période analysée, soit 12.4 % de la totalité des brevets. Le Top 3 est complété par les entreprises Microsoft (4 882 brevets) et Oracle Corp (2 403 brevets). Si on analyse plus globalement ce Top 20, on constate qu'il est constitué uniquement de déposants évoluant dans l'univers des secteurs de l'informatique et des télécommunications.

Le Top 50 représente 52 260 brevets, soit 44 % de la totalité des brevets présents sur l'ensemble des trois périmètres. Si on compare le Top 50 au Top 20, on remarque que le nombre d'organisations progresse de +150 % alors que le nombre de brevets déposés ne s'accroît que de +28 %. De plus, les déposants restent majoritairement des sociétés des secteurs de l'informatique et des télécommunications, seul le gouvernement américain (« The United States Of America ») et l'entreprise Boeing (secteur de l'aéronautique) n'appartiennent pas à ces secteurs. Ce résultat confirme que les brevets déposés sur les technologies de traitement de données sont concentrés sur un groupe restreint d'entreprises du secteur de l'informatique et des télécommunications.

Dans les analyses suivantes, nous allons donc nous concentrer sur l'analyse du Top 20 des déposants avec une approche temporelle en s'appuyant sur les trois périodes identifiées plus haut (1994-2001 : période « pré-Big Data » ; 2002-2009 : développement des Big Data et 2010-2017 : Applications des Big Data).

Tableau 1.3 Top 20, Top 50 et de l'ensemble des trois périmètres

	Top 20	Top 50	Totalité des trois périmètres
# déposants	20	50	20 170
# brevets	40 077	52 260	113 862
# brevets moyens par déposants	2 003.9	1 045.2	5.6

L'étude des déposants de brevets à travers ces trois périodes de huit ans permet de mieux comprendre la structure de ce Top 20. L'analyse temporelle du Top 20 est présentée dans le tableau 1.3. L'analyse de ce tableau permet de mettre en évidence des tendances différentes suivant les périodes et les entreprises analysées. Si on se concentre sur le Top 3, on constate que ces trois déposants ont progressé dans leurs dépôts de brevets sur la période 2002-2007 avec un accroissement moyen d'environ +51% par rapport à la période précédente. Cependant

sur la période suivante, 2010-2017, ce chiffre est en forte diminution, avec une baisse de -35 %. Globalement, on constate cette même tendance sur l'ensemble du Top 20 avec une forte progression du nombre de brevets déposés sur la période 2002-2007 ; +43 % en moyenne pour l'ensemble du Top 20 puis une baisse significative sur la période suivante 2010-2017 : -19 % en moyenne pour l'ensemble du Top 20.

Tableau 1.4 : nombre de brevets déposés par le Top 20 des déposants

Top 20 déposants	Année de création du déposant	Périodes d'analyse			Evolution	
		1994-2001	2002-2009	2010-2017	2002-2009 vs 1994-2001	2010-2017 vs 2002-2009
IBM Corp	1911	4 563	5 630	4 151	+23.4%	-26.3%
Microsoft	1976	1 074	2 628	1 140	+144.7%	-56.6%
Oracle Corp	1977	544	1 067	753	+96.1%	-29.4%
Hitachi	1910	621	880	263	+41.7%	-70.1%
SAP	1972	25	677	880	+2608.0%	+30.0%
Hewlett Packard	1939	555	620	264	+11.7%	-57.4%
Google	1998	18	309	1 109	+1616.7%	+258.9%
Fujitsu Ltd	1935	584	291	339	-50.2%	+16.5%
At&T	1983	360	598	247	+66.1%	-58.7%
Sony Corp	1946	380	539	262	+41.8%	-51.4%
Emc	1979	197	307	573	+55.8%	+86.6%
Samsung	1938	161	388	473	+141.0%	+21.9%
Intel Corp	1968	317	297	396	-6.3%	+33.3%
Sun Microsystems	1982	647	317	0	-51.0%	-100.0%
Amazon	1994	38	96	749	+152.6%	+680.2%
Canon Co Ltd	1933	327	347	196	+6.1%	-43.5%
Nec Corp	1899	380	202	201	-46.8%	-0.5%
Arm	1990	60	311	342	+418.3%	+10.0%
Apple	1976	143	254	309	+77.6%	+21.7%
Cisco Tech Inc	1984	172	257	249	+49.4%	-3.1%
Ensemble Top 20		11 166	16 015	12 896	+43.4%	-19.5%

Néanmoins sur cette période 2010-2017, deux entreprises, Google et Amazon, se distinguent avec une forte progression de leurs dépôts de brevets sur cette période. En effet, leurs taux d'accroissement de dépôt de brevet sont de +258.9 % et 680.2 % respectivement.

Une analyse avec le détail par périmètre nous permettra de mieux comprendre ces différences à travers les trois périodes analysées.

Le tableau 1.5 présente une vue d'ensemble des trois périmètres à travers les périodes retenues. Ce tableau confirme les résultats mis en évidence dans l'analyse temporelle. En effet,

le nombre de brevets déposés par le Top 20 sur la période 2010-2017 baisse fortement sur le périmètre 2 (-56.8%) et progresse faiblement sur le périmètre 3 (+7.6%). Sur cette période le dynamisme est porté par le périmètre 1, à savoir les brevets follow-on innovations de la technologie « MapReduce ». Le nombre de brevets sur ce périmètre est multiplié par dix sur cette période.

Tableau 1.5: nombre de brevets déposés de l'ensemble du Top 20 des entreprises

Ensemble Top 20	Périmètre 1	Périmètre 2	Périmètre 3	Total
1994-2001	0	6 339	4 827	11 166
2002-2009	11	8 559	7 445	16 015
2010-2017	1 186	3 698	8 012	12 896
Evolution 2002-2009 vs 1994-2001		+35.0%	+54.2%	+43.4%
Evolution 2010-2017 vs 2002-2009	+10681.8%	-56.8%	+7.6%	-19.5%

Si on s'intéresse plus particulièrement aux entreprises mises en évidence dans l'analyse globale (IBM Corp, Microsoft, Oracle Corp, Google et Amazon) sur cette période spécifique 2010-2017, on constate bien une émergence des brevets du périmètre 1 dans leur portefeuille de brevets mais elle est très différenciée suivant les entreprises, comme le montre le tableau 1.6.

Tableau 1.6 : nombre de brevets déposés sur la période 2010-2017 – détail par périmètre

Période 2010-2017	Périmètre 1	Périmètre 2	Périmètre 3	Nombre total de brevets
IBM Corp	5.6%	35.2%	59.2%	4 151
Microsoft	11.9%	31.9%	56.1%	1 140
Oracle Corp	12.4%	36.4%	51.3%	753
Google	17.0%	24.8%	58.2%	1 109
Amazon	28.2%	28.3%	43.5%	749
Ensemble Top 20	9.2%	28.7%	62.1%	12 896

Ce tableau indique que l'entreprise IBM, leader en termes de dépôts de brevets sur ces technologies a introduit été la société qui a le moins introduit de brevets « follow-on innovations » de la technologies « MapReduce » dans son portefeuille de brevets sur la période 2010-2017. Ce constat peut s'expliquer par le fait que la société IBM a été à l'origine du langage SQL (Structured Query Language), une des premières technologies de traitement des données au début des années 1970 (Chamberlin, 2012), on peut donc penser que cette technologie est encore fortement exploitée par cette entreprise. De plus, si on compare les sociétés Microsoft et Google qui ont déposé sur cette même période un nombre de brevets comparable (environ 1 100 brevets), on constate que la proportion des brevets du périmètre 1 est supérieure de 5

points pour Google par rapport à Microsoft. La technologie « MapReduce » étant une technologie initiée par Google, il est très logique que la proportion des brevets du périmètre 1 soit élevée pour cette compagnie. Néanmoins, on constate que cette part de brevets appartenant au périmètre 1 est encore plus forte pour la société Amazon ; 28.2 % du total des brevets déposés sur la période 2010-2017. Ce pourcentage est comparable à celui du périmètre 2 qui compte pourtant un nombre bien plus élevé de brevets dans son ensemble. Cette proportion de brevets du périmètre 1 pour la société Amazon peut également être mise en parallèle avec celle de la société Oracle Corp. Pour cette société, les brevets du périmètre 1 ne représentent que 12.4 % pour un nombre total de brevets déposés sur la période analysée équivalent à ceux de la société Amazon (environ 750 brevets). Un pourcentage qui peut s'expliquer par le fait que la société Oracle a été une des premières entreprises à exploiter le langage SQL pour le traitement des données (Chamberlin, 2012) et on peut donc supposer, à l'instar de la société IBM, que cette société continue à exploiter cette technologie de traitement de données.

Cette analyse des trois périmètres d'analyse nous indique une stratégie de dépôt de brevets qui n'est pas homogène à travers les principales organisations déposantes de brevets liées à des technologies de traitement de données. D'une part, nous avons les organisations « historiques » du secteur de l'informatique et des télécommunications représentées par le Top 3 : IBM corp, Microsoft, et Oracle Corp, sociétés qui ont été créées avant les années 1980. Ces organisations ont peu introduit de technologies de traitement de données de type « Big Data » représenté par les brevets « follow-on innovations » de la technologie MapReduce. En effet, les brevets du périmètre 1 représentent moins de 10% de la totalité des brevets déposés sur la période 2010-2017. D'autre part, nous avons des organisations récentes du secteur de l'informatique et des télécommunications à l'image des sociétés Google et Amazon, ces deux sociétés ont été fondées respectivement en 1998 et 1994. Ces entreprises ont introduit fortement des technologies de traitement de données de type « Big Data » dans leur portefeuille de brevets ; les brevets du périmètre 1 représentant plus de 20% de la totalité des brevets déposés sur la période 2010-2017.

1.3.3 Localisation des déposants

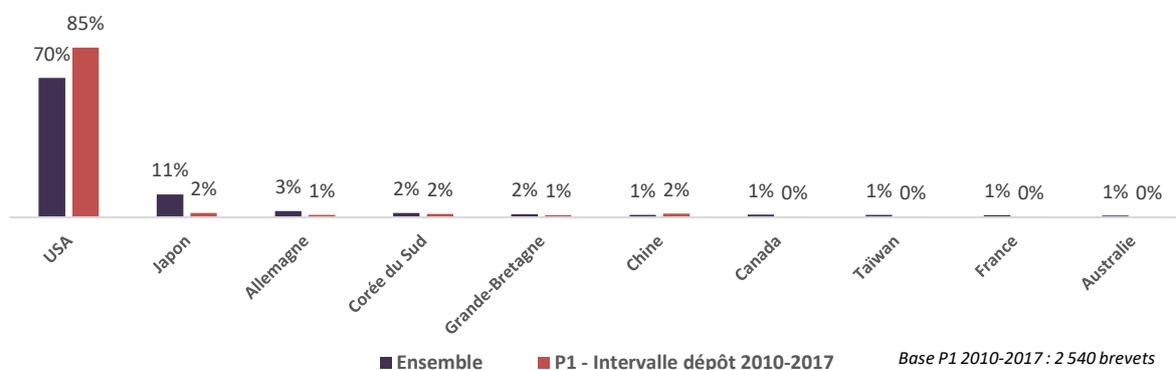
L'ensemble des déposants des brevets étudiés sur les trois périmètres d'analyse sont localisés dans 78 pays différents. Cependant on constate que la localisation des déposants est fortement concentrée autour de quelques pays, comme l'indique le tableau 1.7. Le Top 10 des localisations

des déposants représente 91% de la totalité des brevets étudiés à travers les trois périmètres. Les Etats-Unis arrivent largement en tête de ce Top 10 avec pratiquement 70% des brevets analysés qui ont des déposants localisés dans ce pays. Le continent asiatique concentre 16% des brevets analysés, avec notamment 11% pour le Japon. L'Europe est en retrait de ce Top 10 avec 6% des localisations des déposants, principalement en Allemagne, Grande Bretagne et France.

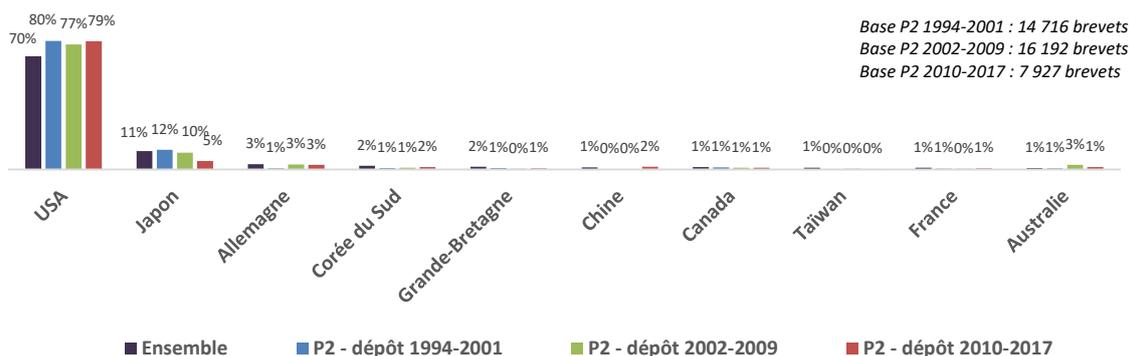
Tableau 1.7 : Top 10 localisation des déposants – ensemble des trois périmètres

Localisation des déposants	Pourcentage de brevets
USA	68%
Japon	11%
Allemagne	3%
Corée du Sud	2%
Grande-Bretagne	2%
Chine	1%
Canada	1%
Taiïwan	1%
France	1%
Australie	1%
Ensemble des brevets - 3 périmètres	110 637

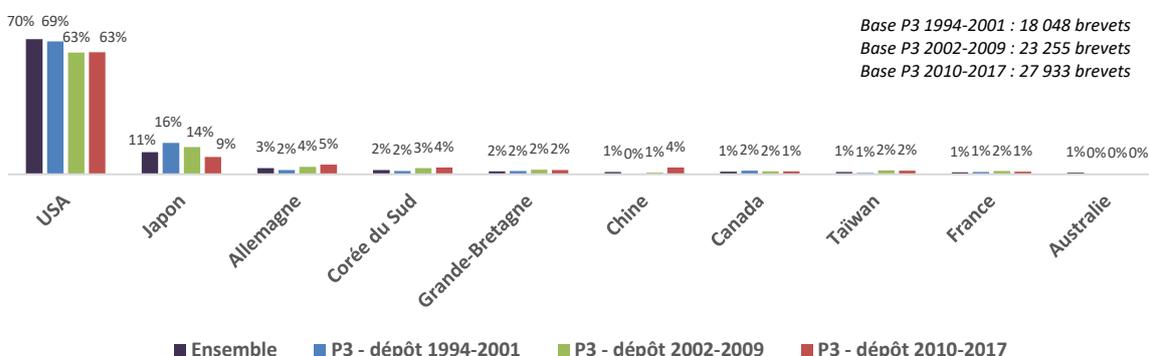
Figure 1.10 Périmètres par période.



Top 10 localisation des déposants – périmètre 1



Top 10 localisation des déposants – périmètre 2



Top 10 localisation des déposants – périmètre 3

Sur le périmètre 1 (MapReduce et Follow-on innovations), la localisation des déposants est encore plus concentrée aux Etats-Unis que sur l'ensemble des périmètres (85% sur le périmètre 1 contre 70% sur l'ensemble des périmètres). Les autres pays du Top 10 représentent moins de 1% en moyenne des localisations des brevets du périmètre 1. Ce résultat est logique, MapReduce étant une technologie d'origine américaine développée par deux ingénieurs de la société Google, il n'est pas étonnant de voir cette technologie se diffuser principalement dans des sociétés américaines.

Sur le périmètre 2 (même CPC que MapReduce), la localisation des déposants est également fortement concentrée aux Etats-Unis (environ 80% entre 1994 et 2017). On constate une certaine stabilité entre les trois périodes temporelles analysées pour les différentes localisations du Top 10, à l'exception du Japon, où le nombre de brevets est divisé par deux sur ce périmètre entre les périodes 2002-2009 et 2010-2017. On retrouve le même phénomène sur le périmètre 3, les déposants localisés au Japon ont moins de breveté de technologies de traitement de

données sur la période 2010-2017 (moins 5 points par rapport à la période 2002-2009). A l'inverse le dépôt de brevet de déposants localisés en Chine a progressé de trois points sur la période 2010-2017. Sur ce périmètre 3, les déposants d'origine américains restent largement majoritaires, environ 65%, on note néanmoins que ce chiffre est inférieur de plus de dix points aux autres périmètres.

En conclusion, les brevets sur les technologies de traitement de données sont majoritairement déposés par des sociétés d'origine américaines, environ 75% sur l'ensemble des trois périmètres. Les déposants d'origine américains sont encore plus présents sur les brevets ayant une filiation avec la technologie MapReduce (périmètre 1 et 2) avec une concentration supérieure à 80%.

1.3.4 Co-brevets

Sur l'ensemble des trois périmètres d'analyse, les co-brevets, c'est-à-dire les brevets déposés par plus d'un déposant sont au nombre de 2321, soit 2% de la totalité des brevets étudiée. Ce faible pourcentage indique donc que la majorité des brevets sont déposés par une unique organisation.

Si on étudie plus en détail ces co-brevets, on constate que plus de 95% d'entre eux ont deux déposants, comme nous l'indique le tableau 1.8 :

Tableau 1.8: nombre de déposants par co-brevet

Nombre de déposants par co-brevet	Co-brevets
2	95.43%
3	3.92%
4	0.34%
5	0.22%
6	0.04%
7	0.04%
Ensemble co-brevets	2 321

Le tableau 1.9 présente la répartition des co-brevets ayant deux déposants sur les trois périodes temporelles retenues.

Tableau 1.9 : co-brevets ayant deux déposants – période de dépôt

Période de dépôt	Ensemble des co-brevets ayant deux déposants
1994-2001	28%
2002-2009	21%
2010-2017	51%
Total	2 215

On constate que la majorité des co-brevets ont été déposés sur la période 2010-2017. On peut donc en déduire que les organisations ont initié une plus forte collaboration sur les technologies de traitement de données sur une période récente. Si on s'intéresse plus particulièrement à ces organisations, on dénombre 1 545 paires de déposants sur l'ensemble des co-brevets ayant deux déposants. Le tableau 1.10 répertorie le Top 10 des paires de co-dépôts sur l'ensemble des trois périmètres, ainsi que leur localisation. Ce Top 10 représente 7% de la totalité des co-brevets ayant deux déposants. En élargissant ce Top 10 à un Top 20, ce taux de représentativité augmente seulement de 4 points, on a donc une faible concentration des co-brevets.

Tableau 1.10: Top 10 des paires de co-dépôts

Nom des co-dépôts	Nombre des co-brevets	Localisation des déposants
Avago Tech General Ip Singapore Pte Ltd / LSI Industries	1.3%	Singapour / Etats-Unis
Freescale Semiconductor Inc / Nxp Usa Inc	1.0%	Etats-Unis
Tsinghua Univ / Nuctech Co Ltd	0.8%	Chine
IBM Corp / Motorola	0.8%	Etats-Unis
Cleversafe Inc / Ibm Corp	0.8%	Etats-Unis
Microsoft / LinkedIn Corp	0.8%	Etats-Unis
Intel Corp / Lueh Guei Yuan	0.5%	Etats-Unis / Chine
Mediaone Group Inc / U S West Inc	0.5%	Etats-Unis
Freescale Semiconductor Inc / Scott Jeffrey W	0.5%	Etats-Unis
Baan Dev Bv / Sun Microsystems	0.5%	Pays-Bas / Etats-Unis
Ensemble des co-brevets avec deux déposants	2 215	

L'étude de la localisation nous montre que la majorité des déposants de ce Top 10 est d'origine américaine. Si on étudie plus précisément ce Top 10, six paires de co-dépôts sont composés uniquement d'organisations américaines, trois sont mixtes avec au moins un déposant américain et une seule paire est composée uniquement d'organisations chinoises. La vision de ce Top 10 confirme les observations constatées dans la section précédente, à savoir une large domination des organisations américaines dans la conception des technologies liées aux traitements des données.

A l'image des sections précédentes, une vision du Top 10 à travers les trois périmètres d'analyse est présentée dans le tableau 1.11.

Tableau 1.11: répartition des brevets du Top10 des co-déposants

Déposants	Localisation des déposants	Périmètre 1	Périmètre 2	Périmètre 3	Ensemble 3 périmètres
Avago Tech General Ip Singapore Pte Ltd / LSI Industries	Singapour / Etats-Unis			100%	100%
Freescale Semiconductor Inc / Nxp Usa Inc	Etats-Unis			100%	100%
Tsinghua Univ / Nuctech Co Ltd	Chine		6%	94%	100%
Ibm Corp / Motorola	Etats-Unis			100%	100%
Cleversafe Inc / IBM Corp	Etats-Unis		94%	6%	100%
Microsoft / LinkedIn Corp	Etats-Unis	59%	18%	24%	100%
Intel Corp / Lueh Guei Yuan	Etats-Unis / Chine		92%	8%	100%
Mediaone Group Inc / U S West Inc	Etats-Unis		9%	91%	100%
Freescale Semiconductor Inc / Scott Jeffrey W	Etats-Unis			100%	100%
Baan Dev Bv / Sun Microsystems	Pays-Bas / Etats-Unis		100%		100%
Moyenne sur le Top 10		6%	32%	62%	100%

Base co-brevets 2 déposants : 2 215

En moyenne, 62% des brevets du Top 10 sont des brevets appartenant au périmètre 3. On en déduit que la majorité des collaborations représentées par ce Top 10 concerne des brevets sur les traitements de données au sens large. On constate d'ailleurs qu'une seule paire d'organisation a déposé des brevets appartenant au périmètre 1, c'est la paire Microsoft/LinkedIn. Leur collaboration s'est majoritairement faite sur des brevets du périmètre 1 (59% des co-brevets déposées par ces deux organisations).

1.3.5 Etude des champs d'applications.

Cette section s'intéresse aux champs d'application des brevets des trois périmètres et s'appuie pour ce faire sur la nomenclature WIPO⁹ qui permet d'attribuer un brevet à un secteur ou un domaine technologique sur la base des codes CIB. Elle se décompose en 5 secteurs d'activité (Électrotechnique, Instruments, Chimie, Mécanique et Autres domaines) et 35 champs d'activité. Le tableau 1.12 indique la répartition des brevets de l'ensemble des trois périmètres analysés à travers cette nomenclature « WIPO ».

⁹ https://www.wipo.int/export/sites/www/ipstats/en/docs/wipo_ipc_technology.pdf

Tableau 1.12 : répartition des brevets à travers la nomenclature WIPO – (3 périmètres)

Nomenclature WIPO	Pourcentage de brevets sur les secteurs/champs
1_Électrotechnique	84.98%
1_Machines, appareils et énergie électriques	0.67%
2_Techniques audiovisuelles	2.15%
3_Télécommunications	4.14%
4_Communication numérique	14.83%
5_Techniques de communication de base	1.04%
6_Informatique	50.83%
7_Méthodes de traitement des données à des fins de gestion	11.00%
8_Semiconducteurs	0.31%
2_Instruments	10.48%
9_Optique	0.62%
10_Techniques de mesure	3.72%
11_Analyse de matériels biologiques	-
12_Dispositifs de commande	4.43%
13_Technologie médicale	1.70%
3_Chimie	0.91%
14_Chimie fine organique	0.04%
15_Biotechnologie	0.20%
16_Produits pharmaceutiques	0.03%
17_Chimie macromoléculaire, polymères	0.01%
18_Chimie alimentaire	0.01%
19_Chimie de base	0.11%
20_Matériaux, métallurgie	0.03%
21_Technique de surface, revêtement	0.04%
22_Technologie des microstructures, nanotechnologie	0.09%
23_Génie chimique	0.19%
24_Écotechnologie	0.15%
4_Mécanique	2.47%
25_Manutention	0.34%
26_Machines-outils	0.15%
27_Moteurs, pompes, turbines	0.25%
28_Machines à fabriquer du papier et des textiles	0.36%
29_Autres machines spéciales	0.31%
30_Procédés et appareils thermiques	0.16%
31_Éléments mécaniques	0.10%
32_Transport	0.81%
5_Autres domaines	1.16%
33_Mobilier, jeux	0.51%
34_Autres biens de consommation	0.39%
35_Génie civil	0.25%
Nombre d'occurrences*	141 795

* un brevet pouvant être présent sur plusieurs secteurs/champs

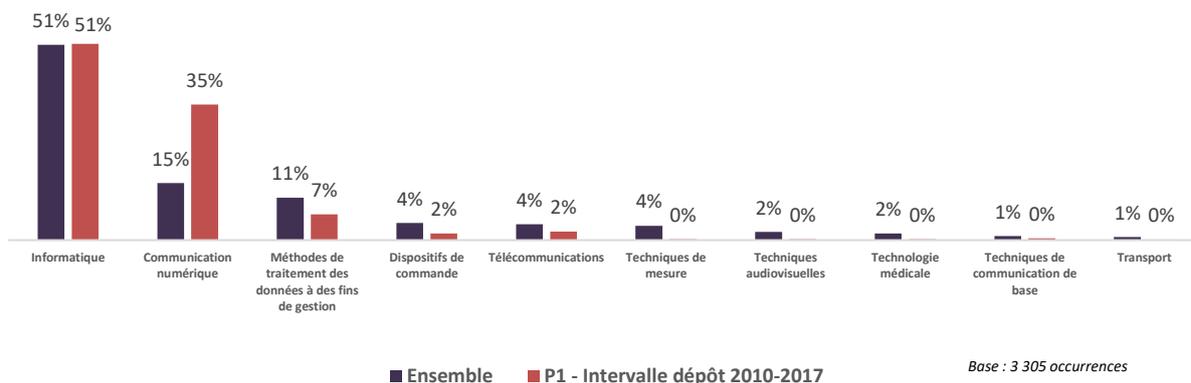
Très logiquement, le secteur « Électrotechnique » concentre la majorité des brevets sur les technologies de traitements de données (85%) suivi du « Secteur Instrument » (10%). Cette concentration est encore plus marquée si l'on considère le détail par champs d'application. En effet, la moitié des brevets analysés sont catégorisés sur le champ d'application

« Informatique » (51%), les champs « Télécommunication » et « Méthodes de traitement des données à des fins de gestion » arrivent respectivement en deuxième (15%) et troisième (11%) position. Cette première analyse sur l'ensemble des trois périmètres indique clairement que les technologies de traitement de données sont majoritairement concentrées sur le secteur des technologies de l'information et de la communication.

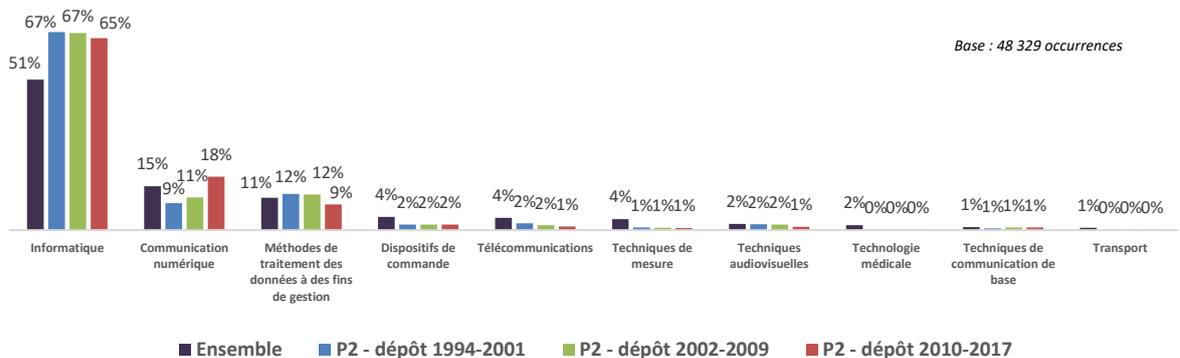
Pour compléter cette première analyse, il est intéressant d'avoir cette vision sectorielle pour chacun des trois périmètres afin de vérifier si cette concentration est également vérifiée à ce niveau de détail.

Les figures 1.11 suivantes présentent le Top 10 des champs d'activités à travers les trois périmètres d'analyse.

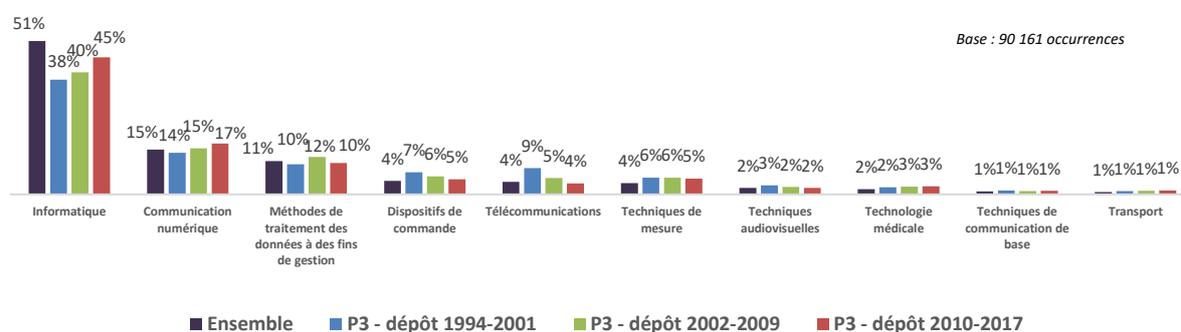
Figure 1.11. Top 10 par champs d'activités



Top 10 champs d'activités – périmètre 1



Top 10 champs d'activités – périmètre 2



Top 10 champs d'activités – périmètre 3

Sur le périmètre 1 (brevets follow-on innovation du brevet MapReduce) on constate que 86% des brevets sont présents sur les champs d'application « Informatique » et « Communication numérique » alors que ces champs ne représentent que 66% sur l'ensemble des brevets. La concentration mise en évidence précédemment est donc encore plus forte sur ce périmètre 1 sur ces deux champs d'applications. Le champ d'application « Informatique » représente 51% de la totalité des brevets du périmètre 1, un chiffre comparable à l'analyse globale. Néanmoins, on constate une sur-représentation du champs « Communication numérique » qui concentre 35% de la totalité des brevets du périmètre 1 contre 11% sur l'ensemble des brevets. Cette concentration des brevets sur le champ « Communication numérique » sur le périmètre 1 est également singulière si on la compare aux deux autres périmètres. En effet sur le périmètre 2 et 3, ce champ représente respectivement 18% et 17% sur la période 2010-2017.

L'analyse du périmètre 2 permet également de constater une forte concentration des brevets sur les champs d'activités « Informatique » et « Communication numérique » (80% en moyenne de l'ensemble des brevets). Ce chiffre est légèrement inférieur au périmètre 1 mais les brevets sont plus fortement concentrés sur le champ « Informatique » (67% en moyenne contre 51% pour les brevets du périmètre 1). Cette concentration reste globalement stable sur les trois périodes temporelles, passant de 69% entre 1994-2001 à 65% sur 2010-2017. C'est le champs « Communication numérique » qui progresse le plus significativement sur ce périmètre. En effet, ce champ représentait 9% sur la période 1994-2001 et atteint 18% sur la période 2010-2017, soit une progression de 9 points sur la période temporelle étudiée.

Enfin le périmètre 3 est le périmètre où la concentration des brevets est significativement la moins élevée, les champs « Informatique » et « Communication numérique » ne représentent que 57% contre 86% et 80% pour le périmètre 1 et 2 respectivement. Le champ d'application « Informatique » reste le champ où le dépôt de brevet est le plus important : 41% en moyenne mais cette concentration de brevet est bien en deçà des périmètres 1 et 2, respectivement 51% et 67%. Néanmoins, on constate que c'est sur ce champ « Informatique » que la progression des dépôts de brevet est la plus forte dans le temps, elle était de 38% sur la période 1994-2001 et atteint 45% entre 2010-2017.

Il est intéressant de compléter cette première approche sectorielle avec une analyse des CPC (Coopérative Patent Classification) qui va nous permettre de calculer l'étendue technologique du brevet, le *Patent Scope*, un indicateur qui permet de quantifier la dispersion des brevets à travers les différentes classes.

Le tableau 1.13 représente le Top 10 des CPC sur l'ensemble des périmètres pour la période la plus récente (2010-2017).

Tableau 1.13: répartition des brevets sur le Top 10 des classes et sous-classes de CPC (ensemble des trois périmètres sur la période 2010-2017)

Libellés des CPC	Répartition des brevets
G06 : Calcul, comptage	51%
G06F : Traitement électrique de données numériques	33%
G06Q : Technologies de l'information et de la communication, spécialement adaptées à des fins administratives	8%
G06N : Agencements informatiques fondés sur des modèles de calcul spécifiques	3%
G06K : Lecture de données graphiques	3%
G06T : Traitement ou génération de données d'image	3%
H04 : Technique de la communication électrique	21%
H04L : Transmission d'information numérique	12%
H04N : Transmission d'images	4%
H04W : Réseaux de télécommunications sans fil	3%
H04M : Communications téléphoniques	1%
Y10 : Regroupe les nouveaux développements technologiques	3%
Y10S : concepts techniques couverts par d'anciens recueils de références croisées de l'USPC	3%

Nombre d'occurrence* 71 145

* un brevet pouvant être présent sur plusieurs secteurs/champs

Le tableau présente les principales classes et sous-classes de CPC sur l'ensemble de nos périmètres d'analyse. A l'instar de l'analyse sur la nomenclature « WIPO », on retrouve une forte concentration de brevets sur les technologies de traitement de données sur la classe « Calcul, comptage », une classe dont les applications sont liées au domaine de l'informatique.

Cette classe est présente sur la majorité des brevets, 51%, dont 33% sont répertoriés uniquement sur le sous-groupe « Traitement électrique de données numériques ».

« Technique de la communication électrique » est la deuxième classe de CPC où sont présents une part importante des brevets liés à des technologies de traitement (21%). Cette classe est liée au domaine de la communication, avec notamment la sous-classe « Transmission d'information numérique » qui concentre 12% des brevets.

Pour étudier la dispersion au sein de nos périmètres d'analyse, il est intéressant de calculer l'indicateur PatentScope. Il est calculé de la manière suivante (Squicciarini et al., 2013) :

$$PatentScope_n = n_p ; n \in \{CPC_1^4; \dots, CPC_i^4; CPC_j^4; \dots; CPC_n^4\} \& CPC_i^4 \neq CPC_j^4$$

où n_p désigne le nombre de sous-classes distinctes de CPC à 4 chiffres répertoriées dans le brevet p.

Cet indicateur permet de connaître le « taux » de brevet présent sur une classe ou sous-classe de CPC. En utilisant l'indicateur normalisé en fonction de la valeur maximum des brevets au sein d'une même cohorte (même année de dépôt et technologie) (Squicciarini et al. 2013), un Patentscope égale à 1 signifie que la totalité des brevets analysés est catégorisée sur cette classe ou sous-classe de CPC et inversement s'il est égal à 0, aucun brevet n'est catégorisé sur la classe ou sous-classe de CPC considérée.

La Figure 1.12 présente l'indicateur Patentscope pour chaque périmètre d'analyse sur le Top 10 des sous-classes de CPC pour la période 2010-2017

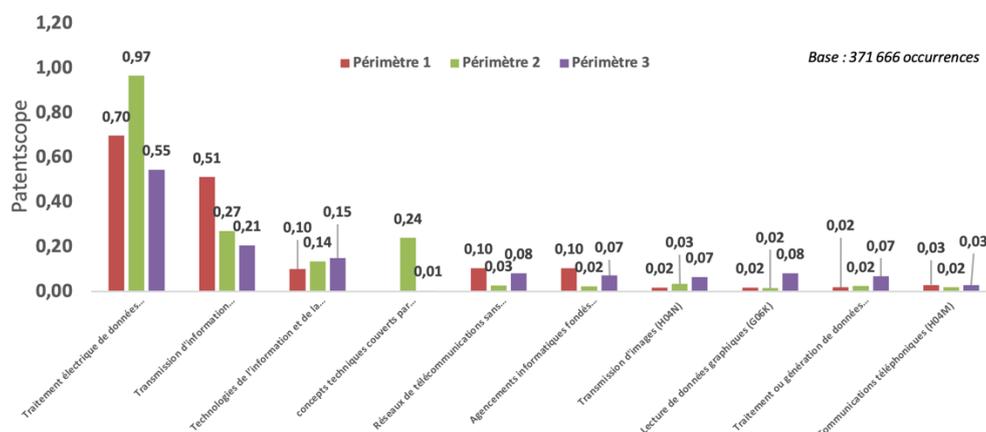


Figure 1.12. Indicateur PatentScope sur le Top 10 des sous-classes de CPC – trois périmètres

L'étude de l'indicateur PatentScope au travers des différents périmètres d'analyse confirme la forte concentration des brevets sur les deux sous-classes : « traitement électrique de données numériques », (G06F) et « transmission d'information numérique » (H04L). Les deux premiers périmètres sont d'ailleurs très fortement concentrés sur la sous-classe « traitement électrique de données numériques » avec des PatentScope de 0.70 et 0.97 pour le périmètre 1 et 2 respectivement. Il est intéressant de noter que la valeur de l'indicateur PatentScope est proche 1 pour ces deux périmètres, ce qui signifie que la majorité des brevets déposés entre 2010 et 2017 (la quasi-totalité pour le périmètre 2) se situe dans cette catégorie pour ces deux périmètres. Concernant le périmètre 3, l'indicateur PatentScope est moins élevé sur cette classe (0.55) puisqu'en moyenne un brevet sur deux sur ce périmètre se situe dans cette sous-classe « traitement électrique de données numériques ». De plus, on constate une dispersion légèrement plus forte des brevets de ce périmètre par rapport aux brevets des deux autres périmètres sur la seconde partie du Top 10 des sous-classes de CPC.

L'étude des champs d'applications au travers de la nomenclature « WIPO » et de la classification des brevets met en évidence que l'utilisation des technologies de traitement de données se fait majoritairement dans le domaine des technologies de l'information et de la communication avec toutefois des différences notables. Tout d'abord, les brevets ayant une filiation avec la technologie MapReduce (brevets du périmètre 1) ont une proportion plus forte à être utilisés dans des applications ayant des débouchés dans la communication numérique. Ensuite, les brevets dont les technologies ont influencé la création de la technologie MapReduce (brevets du périmètre 2) sont des brevets qui ont une plus forte application dans le domaine de l'informatique. Enfin, les brevets sur les technologies de traitement de données au sens large (brevets du périmètre 3) sont légèrement moins concentrés sur les champs d'applications mentionnés plus haut et ils sont présents dans des domaines d'applications plus variés que les brevets des deux autres périmètres.

1.4 Conclusion

Ce premier chapitre avait pour objectif de caractériser l'émergence des technologies de traitement de données Big Data. Afin de réaliser cette étude, la technologie MapReduce, une des technologies majeures dans le secteur des technologies de traitement Big Data, a été utilisée

comme référence technologique afin de définir nos périmètres d'analyse. A travers l'étude du brevet MapReduce déposé par l'entreprise Google en 2004, deux périmètres d'analyse ont pu être identifiés, le périmètre 1 constitué de brevets ayant un lien de citation direct ou indirect avec le brevet MapReduce et le périmètre 2 constitué des brevets ayant les mêmes classes technologiques que les brevets cités par le brevet MapReduce. Pour compléter ces deux premiers périmètres, un troisième périmètre a été créé grâce à une recherche lexicographique sur des mots-clés spécifiques aux technologies de traitements de données.

L'étude de ces trois périmètres, et notamment du périmètre 1, a permis de mettre en évidence un certain nombre de caractéristiques du secteur des technologies Big Data. Le premier point concerne la dimension temporelle, après une période de développement des technologies (2002-2009), le nombre de brevets sur les technologies de traitement Big Data a fortement progressé à partir des années 2010 (environ +80% en moyenne de dépôt de brevets sur le périmètre 1). Le second point concerne des acteurs américains ; il apparaît que ces technologies sont majoritairement développées aux Etats-Unis (85% des brevets du périmètre 1 sont déposés par des organisations américaines). A partir des années 2010, période de fortes activités sur le secteur des technologies Big Data, ce sont des entreprises relativement récentes, telles que Google ou Amazon, qui ont été les plus dynamiques sur ces technologies. Enfin les technologies de traitement Big Data se concentrent majoritairement sur les champs de l'informatique et de la communication numérique (plus de 80% des brevets du périmètre 1 appartiennent à ces deux champs d'application selon la classification WIPO).

Comme mentionné précédemment, la technologie MapReduce développée par l'entreprise Google en 2004 a servi de point de départ aux travaux réalisés dans ce premier chapitre. L'étude de la technologie MapReduce met également en évidence deux points singuliers liés à cette technologie. Tout d'abord, MapReduce a servi de base pour le développement du Framework Hadoop qui est également considéré comme une des technologies majeures des technologies de traitement Big Data (Ibrahim et al., 2009). La technologie Hadoop a été transmise par son inventeur Doug Cutting (un ingénieur travaillant chez l'entreprise Yahoo!) à la fondation communautaire Apache en 2008 afin de devenir le projet open-source « *Apache Hadoop* ». A partir de cette date, les membres de la communauté Apache pouvaient donc accéder à la technologie MapReduce par l'intermédiaire de ce projet open-source. De plus, la technologie MapReduce a bénéficié d'une accessibilité directe par l'intermédiaire du programme « *Open*

*Patent Non-Assertion Pledge*¹⁰ » initié par l'entreprise Google en 2013. Ce programme permet aux entreprises tierces, de toutes natures, d'utiliser la technologie MapReduce de manière libre à partir du moment où elles respectaient les conditions énoncées par le dispositif. Il apparaît que la technologie MapReduce a profité d'une forte accessibilité à travers ces deux mécanismes d'innovation ouverte, il est donc nécessaire de s'interroger de l'impact de ces deux mécanismes sur l'innovation au sein du secteur des technologies Big Data.

1.5 Références

- Ajiferuke, I., Burell, Q., & Tague, J. (1988). Collaborative coefficient: A single measure of the degree of collaboration in research. *Scientometrics*, 14(5–6), 421–433.
<https://doi.org/10.1007/BF02017100>
- Banerjee, P., Gupta, B. M., & Garg, K. chandra. (2000). Patent statistics as indicators of competition an analysis of patenting in biotechnology. *Scientometrics*, 47(1), 95–116.
<https://doi.org/10.1023/A:1005669810018>
- Boisseau, P., & Loubaton, B. (2011). Nanomedicine, nanotechnology in medicine. *In Comptes Rendus Physique (Vol. 12, Issue 7, pp. 620–636)*.
<https://doi.org/10.1016/j.crhy.2011.06.001>
- Braun, T., Schubert, A., & Zsindely, S. (1997). Nanoscience and nanotechnology on the balance. *Scientometrics*, 38(2), 321–325. <https://doi.org/10.1007/BF02457417>
- Chamberlin, D. D. (2012). Early History of SQL. *IEEE Annals of the History of Computing*, 34(4), 78–82. <https://doi.org/10.1109/MAHC.2012.61>
- Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. *OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, pp. 137-150* ; <https://research.google/pubs/mapreduce-simplified-data-processing-on-large-clusters/>
- Dittrich, J., Quiané, J.-A., & Quiané-Ruiz, Q. (2014). Efficient Big Data Processing in Hadoop MapReduce. <http://infosys.cs.uni-saarland.de>
- Douzou P, Durand G, & Siclet G. (2001). Les biotechnologies. *Presses universitaires de France*.

¹⁰ [Open Patent Non-Assertion Pledge - Google](https://www.google.com/patents/opnpledge/pledge/).
<https://www.google.com/patents/opnpledge/pledge/>

- Hashem, I. A. T., Anuar, N. B., Gani, A., Yaqoob, I., Xia, F., & Khan, S. U. (2016). MapReduce: Review and open challenges. *Scientometrics*, *109*(1), 389–422. <https://doi.org/10.1007/s11192-016-1945-y>
- Hullmann, A. (2007). Measuring and assessing the development of nanotechnology. *Scientometrics*, *70*(3), 739–758. <https://doi.org/10.1007/s11192-007-0310-6>
- Ibrahim, S., Jin, H., Lu, L., Qi, L., Wu, S., & Shi, X. (2009). Evaluating MapReduce on virtual machines: The Hadoop case. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5931 LNCS*, 519–528. https://doi.org/10.1007/978-3-642-10665-1_47
- Launois P, Benisty, H., Berthoz, A., Corriu, R., Dupas, C., Fert, A., Grangier, P., Henry, C., Joachim, C., Joblin, C., Kapsa, P., Ledoux-Rak, I., Marzin, J., Nozières, J., Pannetier, B., Pérez, A., Stievenard, D., Vieu, C., Yvon, J., & Zyss, J. (2004). Nanosciences et nanotechnologies. <https://www.cnrs.fr/comitenational/doc/rapport/2004/lesateliers/023-046-Chap2-Nanosciences.pdf>
- Manjunatha S.B, Biradar D. P, & Aladakatti Y. R. (2016). Nanotechnology and its applications in agriculture:A review. <https://www.researchgate.net/publication/303665019>
- Meyer, M., Persson, O. (1998) Nanotechnology-interdisciplinarity, patterns of collaboration and differences in application. *Scientometrics* *42*, 195–205 <https://doi.org/10.1007/BF02458355>
- Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. (2008). Refining search terms for nanotechnology. In *Journal of Nanoparticle Research* (Vol. 10, Issue 5, pp. 715–728). <https://doi.org/10.1007/s11051-007-9266-y>
- Porter, M. E. (1985). Technology and competitive advantage. *Journal of Business Strategy*, *5*(3), 60–78. <https://doi.org/10.1108/eb039075>
- Schultz, L. I., & Joutz, F. L. (2010). Methods for identifying emerging General Purpose Technologies: A case study of nanotechnologies. *Scientometrics*, *85*(1), 155–170. <https://doi.org/10.1007/s11192-010-0244-2>
- Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economics*, *191*, 97–112. <https://doi.org/10.1016/j.ijpe.2017.06.006>
- Singh, V. K., Banshal, S. K., Singhal, K., & Uddin, A. (2015). Scientometric mapping of research on ‘Big Data.’ *Scientometrics*, *105*(2), 727–741. <https://doi.org/10.1007/s11192-015-1729-9>

- Squicciarini M, Dernis H, & Criscuolo C. (2013). Measuring Patent Quality: Indicators of Technological and Economic Value. <https://doi.org/10.1787/5k4522wkw1r8-en>
- Stryjewska, A., Kiepusa, K., Librowski, T., & Lochyński, S. (2013). Biotechnology and genetic engineering in the new drug development. Part I. DNA technology and recombinant proteins. *Pharmacological Reports*, 65(5), 1075–1085. [https://doi.org/10.1016/S1734-1140\(13\)71466-X](https://doi.org/10.1016/S1734-1140(13)71466-X)
- Wenda, S., Illner, S., Mell, A., & Kragl, U. (2011). Industrial biotechnology—the future of green chemistry? *Green Chemistry*, 13(11), 3007. <https://doi.org/10.1039/c1gc15579b>
- Wieczorek A. (2003). Use of biotechnology in agriculture -- benefits and risks. *Honolulu (HI): University of Hawaii*. 6 p. (Biotechnology; BIO-3)
- UK Intellectual Property Office Informatics Team. (2014). Big Data - A patent overview

CHAPITRE 2. INNOVATION OUVERTE ET COLLABORATION DANS LE SECTEUR DES TECHNOLOGIES DE TRAITEMENT DES DONNEES « BIG DATA » : PROPRIETE INTELLECTUELLE ET ENJEUX STRATEGIQUES

2.1 Introduction

Les différentes analyses du chapitre 1 ont mis en évidence un fort dynamisme dans les technologies de traitement de données. Ces analyses ont permis d'étudier environ 109 000 brevets en lien avec des technologies de traitement de données déposées sur la période 1994-2017. Ces brevets ont été déposés par plus de 20 000 organisations, dont des acteurs majeurs du secteur des technologies de l'information et de la communication tels que IBM, Microsoft, Google ou encore Amazon. Si on s'intéresse plus spécifiquement aux technologies de traitement des données « Big Data », on constate que certaines organisations ont plus fortement investi sur ces technologies au début des années 2000. Ce fut le cas de Google avec le dépôt de brevet sur la technologie « MapReduce » en 2004. Au même titre que « MapReduce », la technologie « Hadoop » est également considérée comme une des technologies « socles » des technologies de traitement des données « Big Data » (Dhyani & Barthwal, 2014). La technologie « Hadoop » est une infrastructure de développement créée en 2005 par l'ingénieur Doug Cutting en se basant sur la technologie « MapReduce ». En 2006, Doug Cutting rejoindra l'organisation Yahoo avec ses travaux sur la technologie « Hadoop » et deux ans plus tard, Yahoo décidera de proposer la technologie « Hadoop » en libre accès sous la forme d'un projet open source. A partir de cette période, de nombreuses technologies basées sur l'environnement « Hadoop/MapReduce » verront le jour (Khezzr & Navimipour, 2017), on peut notamment citer les technologies « Pig » développées par Yahoo en 2006 ou encore « Hive » développé par Facebook en 2010. Même si ces nouvelles technologies de traitement de données « Big Data » sont initiées par des organisations privées, on constate qu'elles ont toutes intégré des projets regroupés au sein de la fondation Apache (Apache Software foundation ou « ASF »). La fondation Apache est une organisation à but non lucratif créée en juin 1999 aux Etats-Unis. Elle est à l'origine de nombreux logiciels et projets open source dont le très connu serveur web « Apache http ».

La Figure 2.1. résume la chronologie des événements pour les principales technologies de traitement de données « Big Data » :

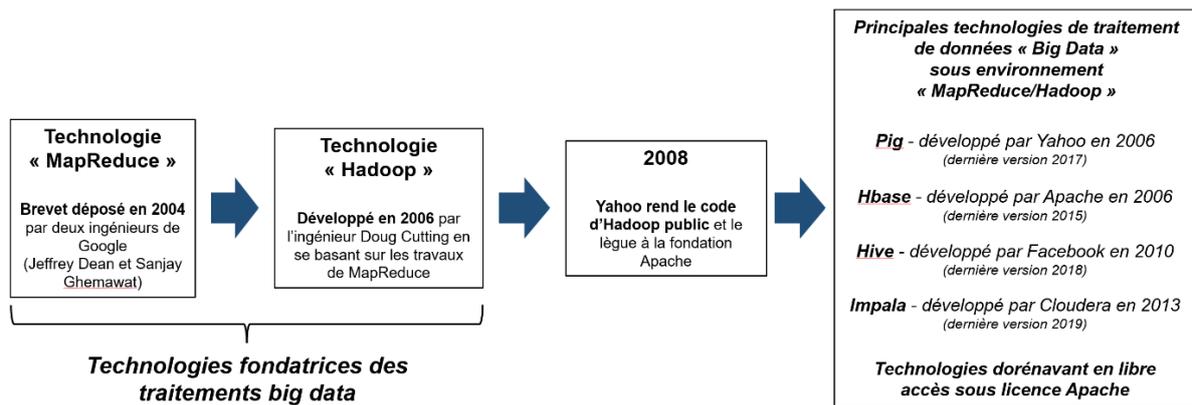


Figure 2.1 Historique de l'écosystème des technologies de traitement des données « Big data » ; source : Charles Xavier (2022)

La fondation Apache regroupe dorénavant en son sein les technologies majeures de traitement de données « Big Data », elle est également à l'origine de nombreux projets visant l'amélioration de ces technologies. Par ailleurs, cette fondation est sponsorisée par de nombreuses organisations du secteur de l'informatique et des télécommunications. Il apparaît que toutes ces organisations se concurrencent au sein du secteur des technologies de l'information et de la communication mais néanmoins elles collaborent au sein de la fondation Apache sur des projets permettant l'amélioration des technologies de traitement « Big Data ». Il est légitime de s'interroger sur deux points :

- Pourquoi des organisations concurrentes collaborent dans le cadre d'une organisation à but non lucratif sur des technologies qui sont essentielles pour leur développement ?
- Quel est l'impact de cette coopération sur la performance à l'innovation des organisations développant des technologies de traitement « Big Data » ?

Pour ce deuxième point, il sera intéressant de distinguer, d'une part, les grands groupes technologiques souvent sponsors de la fondation Apache et d'autre part, les petites et moyennes entreprises, ainsi que les organisations comme les universités.

Par ailleurs, l'étude de la temporalité du chapitre I a également pu mettre en évidence un fort dynamisme depuis 2010 des brevets ayant une filiation avec la technologie MapReduce, qualifiées de « follow-on » innovations, avec un taux d'accroissement du nombre de brevets déposés par an proche de +80%. Ce fort taux d'accroissement est à mettre en parallèle avec une décision forte de la société Google concernant la gestion de la propriété intellectuelle de la technologie MapReduce. En effet, en 2013 Google annonce la mise en place du programme

« Open Patent Non-Assertion (OPN) Pledge » qui donne libre accès à la technologie MapReduce à travers dix brevets de la firme. Avec ce programme, Google « s'engage à ne pas poursuivre un utilisateur, un distributeur ou un développeur de logiciels open source sur des brevets en particulier, à moins d'être attaqué en premier »¹¹. En choisissant de mettre sous forme de « Patent Pledge » les brevets de la technologie MapReduce, Google choisit donc d'ouvrir sa propriété intellectuelle sur les technologies de traitement de données « Big Data ». Ce choix fort, à contre-courant du mécanisme classique du dépôt de brevet permettant d'obtenir un monopole temporaire, a forcément entraîné des conséquences sur la diffusion de la technologie MapReduce et donc la coopération entre les organisations.

Pour étudier les collaborations dans le secteur des technologies « Big Data », il est également important de considérer l'intégration d'une des principales technologies « Big Data » sous forme de « Patent Pledge » en complément des collaborations potentielles au sein de la fondation Apache. A travers l'étude des « Patents Pledge », on va s'intéresser à l'impact de cet accès libre à la technologie MapReduce sur l'écosystème des technologies de traitement de données « Big Data ». En effet, comme indiqué précédemment, une grande majorité des technologies de traitement « Big Data » développées au cours de la dernière décennie s'appuie sur l'utilisation de la technologie MapReduce en complément de la technologie Hadoop (voir schéma 1).

Il est donc légitime de se demander si :

1. cet accès libre a favorisé l'émergence d'un standard de traitement des données « Big Data » basé sur ces deux technologies
2. l'émergence de ce standard a eu un impact sur les autres technologies de traitement de données ?

L'établissement d'un standard est un objectif sectoriel fort qui peut fortement inciter des organisations concurrentes à collaborer sur un même secteur (Shapiro & Varian, 1999). L'analyse des « Patents Pledge » va permettre d'avoir des premiers éléments de compréhension sur ces différentes questions.

Avant de s'intéresser spécifiquement aux mécanismes de collaboration au sein du secteur des technologies de traitement des données « Big Data », il est nécessaire de comprendre, dans un

¹¹ <https://opensource.googleblog.com/2013/03/taking-stand-on-open-source-and-patents.html>

premier temps, les incitations qui poussent des organisations concurrentes à collaborer entre elles au sein d'un même secteur d'activité

2.2 Les incitations à collaborer avec ses concurrents

Historiquement, la littérature économique a mis en avant que la mise en concurrence d'entreprises favorise le dynamisme d'un secteur d'activité. En effet, les entreprises doivent développer un avantage concurrentiel par une stratégie d'innovation forte afin de créer une réelle plus-value face aux autres acteurs du secteur (Porter, 1991). Cette vision met en avant l'impact positif de la concurrence sur l'innovation. Cependant à la fin du siècle dernier, une nouvelle vision met en avant que les organisations sont amenées à collaborer tout en étant en concurrence, cette vision a été portée par plusieurs auteurs. Nalebuff & Brandenburger (1997) ont mis en évidence, grâce à des travaux s'appuyant sur la théorie des jeux, que la simultanéité de comportement coopératif et concurrentiel était bénéfique pour les organisations. Lado et al., (1997), grâce à une méthodologie similaire, ont montré que la coopération et la concurrence n'étaient pas deux phénomènes opposés mais plutôt deux phénomènes indépendants. Enfin Bengtsson & Kock, (1999) ont utilisé la théorie des réseaux pour caractériser la coopétition entre les organisations. Pour eux, ce phénomène qualifié de « coopétition » dépend de deux facteurs essentiels :

- La position relative des acteurs dans le secteur
- Le besoin en ressources extérieurs

Si ces deux facteurs sont importants alors les organisations sont plus enclines à adopter une stratégie de coopétition. En adoptant cette stratégie, les organisations peuvent donc bénéficier des avantages d'une stratégie basée sur la compétition mais également des avantages d'une stratégie basée sur la coopération. En bénéficiant de ce double avantage, une stratégie de coopétition est donc supérieure à des stratégies individuelles de concurrence et de coopération (Dagnino et al., 2007).

Pour appuyer ces références littéraires, on peut prendre l'exemple de l'étude du marché des semi-conducteurs à la fin du siècle dernier (Lim et al., 2010). Les auteurs ont étudié le comportement de trois acteurs majeurs de ce secteur : IBM, Intel et AMD.

Au début des années 1970, IBM est le leader historique du secteur, un leadership qu'on retrouve également dans leur production scientifique tant au niveau des articles scientifiques et que des

dépôts de brevets. Intel et AMD sont des acteurs moindres sur ce secteur, et leur production scientifique est comparable aux autres acteurs du secteur. La production scientifique est un élément essentiel sur le marché des semi-conducteurs car c'est un secteur où la technologie est complexe et qui nécessite de capitaliser sur les technologies existantes, d'où l'importance des brevets pour ce secteur. Au cours des années 1970, IBM et Intel ont commencé à collaborer, Intel fournisseur des processeurs à IBM pour la production de leur Personal Computer (PC). Cette relation de coopération est dominée par IBM sur toute la décennie des années 80, IBM étant le plus important client d'Intel, il entre au capital de celui-ci et impose certains choix stratégiques comme une collaboration technologique avec le concurrent direct AMD. Au cours de cette même décennie, Intel, grâce à la collaboration avec IBM, se renforce sur le marché des semi-conducteurs et investit fortement en recherche et développement. Au début des années 1990, Intel est devenu un des leaders du marché des semi-conducteurs, et il devient de moins en moins dépendant d'IBM, en collaborant notamment avec d'autres acteurs majeurs tels que l'entreprise Compaq, concurrent direct d'IBM sur le marché des PC. IBM décide alors de s'adresser à d'autres fournisseurs de processeurs et débute alors une phase de concurrence entre les deux organisations, concurrence qui durera toute la décennie. Au début des années 2000, avec l'avènement d'internet, Intel et IBM décide de collaborer de nouveau sur le marché des serveurs, en pleine expansion avec notamment la production de serveurs spécifiques, « blade server », une collaboration d'intérêt entre les deux entreprises avec l'émergence de nouveaux acteurs liés à Internet. En trois décennies, 1980, 1990 et 2000, Intel et IBM ont donc successivement collaboré, puis se sont concurrencés puis enfin ont de nouveau collaboré, montrant ainsi que le comportement de coopération pouvait alterner dans le temps.

Les auteurs ont également mis en évidence que la production scientifique des organisations était également fonction de ces différentes phases de collaboration et de compétition. En effet, au début des années 1990, le nombre de brevet d'Intel sur les classes spécifiques aux semi-conducteurs ne représentait que 20 % des brevets d'IBM, en 1999 ce pourcentage atteint 50 % et 70 % en 2004. Ces chiffres montrent bien le « rattrapage » d'Intel sur IBM en termes de production de brevets au cours des années 1990, phase de concurrence importante entre ces deux entreprises.

Avant de s'intéresser à l'impact de l'adoption d'une telle stratégie pour un secteur d'activité, il est d'abord nécessaire de présenter les différentes modalités de coopération entre organisations concurrentes.

2.2.1 Les modalités de coopération entre concurrents

Il existe deux principales modalités de coopération entre des organisations concurrentes : d'une part, les accords de coopération bilatérale entre deux concurrents par la voie de transactions sur leur propre propriété intellectuelle et d'autre part, les accords de coopération multilatérale au travers d'une organisation tiers facilitant cette collaboration : les « pools ».

2.2.1.1 Accord de coopération bilatérale : les licences d'exploitation des brevets

Une des modalités les plus courantes de coopération bilatérale entre organisations concurrentes est l'exploitation de la propriété intellectuelle grâce à la mise en place de contrat de licences sur les brevets détenus par chacune des organisations. En effet, une organisation détentrice de brevets peut accorder un droit d'exploitation de ceux-ci grâce à la mise en place de licences. Il existe trois types de licences d'exploitation (Cohendet et al., 2006) :

- Licence simple : licence accordant le droit d'exploitation d'un brevet à un tiers en contrepartie d'une rémunération.
- Licence exclusive : licence accordant le droit d'exploitation exclusif d'un brevet à un tiers en contrepartie d'une rémunération, généralement plus élevée que la licence simple. Cette licence exclusive est également fonction d'un territoire spécifique (région, pays, continent). Il est important de noter que cette licence n'équivaut pas à un accord de cession, puisqu'elle ne transfère aucun droit de propriété.
- Licences croisées : mécanisme par lequel le titulaire d'un premier brevet accorde une licence de son brevet au titulaire d'un second brevet, le titulaire du second brevet accordant lui aussi une licence de son propre brevet au titulaire du premier brevet. Dans ce type de licence, il y a généralement aucune rémunération spécifique, la valeur des brevets de chaque organisation faisant office de compensation financière.

Les licences croisées sont le mécanisme le plus souvent retenu quand deux organisations souhaitent coopérer de façon bilatérale, notamment quand les deux organisations sont concurrentes sur un secteur d'activité. Chen & Chen, (2011) ont mis en évidence ce type de coopération entre deux organisations concurrentes sur le secteur de la technologie LED (Light Emitting Diode) entre les entreprises Nichia et Osram. Nichia, société japonaise, est l'entreprise historique du secteur. Elle est la première à généraliser la production de masse de la technologie LED blanche en 1996, une technologie innovante permettant notamment une plus grande efficacité énergétique, des coûts de maintenance réduits, et une durée de vie plus longue. Osram,

une société allemande, également un acteur majeur de ce secteur, a généralisé la production de masse de la technologie à partir des années 2000. L'étude de leur brevet montre également cet écart temporel dans la maîtrise des technologies nécessaires à la production de matériel LED, en effet le premier dépôt de brevet sur l'office américain USPTO date de 1987 pour Nichia alors qu'il date de 1992 pour Osram. Après une longue période de litiges juridiques concernant leurs brevets respectifs, Nichia et Osram ont conclu un règlement et signé un accord de licences croisées en 2002 couvrant l'exploitation de leurs brevets sur les technologies LED et des technologies d'emballage associées. Il est logique de s'interroger sur l'intérêt d'un tel accord.

Chen et Chen ont donc entrepris d'analyser les brevets des deux organisations. En s'intéressant plus particulièrement aux classes des brevets, ils ont mis en évidence des différences notables entre ces deux sociétés concurrentes. En effet, Nichia s'est concentré sur les classes « Lampes électriques et appareils à décharge » et « Compositions » tandis qu'Osram s'est focalisé sur les « Procédés de revêtement » et « Compositions » également. Grâce à cette analyse de brevets, Chen et Chen ont donc conclu que les deux organisations pouvaient effectivement se concurrencer sur la classe de brevets « compositions », les deux sociétés étant fortes sur cette classe, cependant une coopération était possible sur les classes « Lampes électriques et appareils à décharge » et « Procédés de revêtement » compte tenu de leur spécialisation respective sur ces deux classes. L'accord de licences croisées entre Nichia et Osram a donc une vision de complémentarité technologique, l'objectif principal étant de prendre un avantage technologique sur les autres concurrents du secteur des technologies LED.

2.2.1.2 Accord de coopération multilatéral : les « pools » de brevets

On peut définir les « pools » de brevets comme « *des organisations formelles ou informelles où des entreprises à but lucratif partagent les droits de leurs brevets entre eux et avec des tiers* » (Lerner et al., 2007). Des « pools » de brevets ont été créés dans de nombreux secteurs d'activités. On peut notamment citer le « pool » sur les technologies propres à l'aviation créé en 1917 aux Etats-Unis, le « pool » formé par les grands industriels américains de la radio en 1924 et plus récemment, en 1997, le « pool » sur la technologie du MPEG2 (technologies de compression de l'image et du son utilisées notamment dans la création de DVD) créé par des acteurs majeurs du secteur de l'électronique (Clark, 2000).

C'est au milieu du XIXème qu'un des premiers « pools » a vu le jour aux Etats-Unis, il portait sur la machine à coudre (Lampe & Moser, 2010). Avant la mise en place du pool, il était

nécessaire de payer une redevance de 25 dollars à l'inventeur Elias Howe pour chaque machine à coudre vendue. En effet, Howe détenait un des brevets essentiels à la fabrication des machines à coudre. Ce brevet a été remis en cause par les principaux acteurs du secteur et il en a découlé un litige juridique qu'on a nommé « *Sewing Machine War* ». Pour remédier à cette situation, les principales entreprises du secteur et l'inventeur Howe ont décidé de mettre en commun leurs brevets sur la machine à coudre. Ils ont ainsi créé le pool de brevet « *Sewing Machine Combination* » en 1856.

Son fonctionnement était le suivant :

- Au moins 24 fabricants devaient être licenciés (une requête spécifique de Howe).
- Les fondateurs du pool partageaient à parts égales les bénéfices.
- L'inventeur Howe recevait une redevance de 5 dollars par machine vendue aux Etats-Unis, 1 dollar pour les machines exportées.

Il est important de noter que les prix des machines à coudre n'étaient pas fixés, les fondateurs du pool restaient en concurrence sur ce marché qui était ouvert à la concurrence.

La force d'un pool est donc de concentrer en son sein des brevets essentiels, qui sont des brevets nécessaires à la reproduction d'un bien, un service ou une technologie (Ayerbe & Azzam, 2015). Par ailleurs, on constate également que la formation d'un pool crée une forme de dichotomie au sein d'un secteur d'activité avec deux groupes d'entreprises ; on a, d'une part, les entreprises faisant partie du pool (les fondateurs et membres) et, d'autre part, les entreprises à l'extérieur du pool, ne faisant pas partie de l'organisation (non-membres). Les analyses du premier patent pool « *Sewing Machine Combination* » a mis en évidence une baisse de l'incitation à innover après la formation du pool pour les deux groupes d'entreprises (Lampe & Moser, 2010). De plus, il a également été constaté une augmentation des litiges entre les membres et les non-membres du pool concernant les droits de propriétés intellectuelles.

Dans le cadre de son étude du pool des technologies MPEG2, Vakili, (2012) a défini un cadre analytique listant quatre hypothèses concernant les effets de la formation d'un « pool » sur un secteur d'activité :

1. **Baisse du coût d'accès à la technologie et avantages aux fondateurs historiques du pool** : en regroupant au sein d'une même organisation des brevets dits essentiels sur une technologie donnée, les pools permettent de réduire les coûts de transaction (diminution des coûts de recherche, de négociation et de litiges). Cette baisse des coûts de transaction associée à une diffusion facilitée des technologies renforce l'incitation à créer de

nouvelles technologies propres à chaque organisation présente dans le secteur considéré. Il est néanmoins important de constater que les fondateurs historiques du « pool » sont les premiers à profiter de ces avantages. Ils vont d'une part, capitaliser sur les aspects financiers que procure le « pool » avec la monétisation des différentes technologies et, d'autre part, posséder une capacité plus forte à innover que les acteurs hors du pool du fait de l'accès aisé aux différentes technologies.

2. **Augmentation de la concurrence** : après la formation d'un pool, la concurrence des organisations au sein d'un secteur d'activité peut s'accroître à travers deux mécanismes. Le premier mécanisme porte sur les fondateurs du « pool ». En effet, grâce à la facilité d'accès aux technologies au travers du pool, les fondateurs voient leur capacité à innover se renforcer, ils peuvent donc se concurrencer plus fortement sur le marché. De plus, les redevances perçues et la baisse des coûts de litiges, suite à la formation du « pool », permettent d'accroître la capacité financière des membres historiques. Le deuxième mécanisme est plus spécifique aux acteurs hors du « pool ». En effet, le fait de pouvoir accéder aux technologies grâce au « pool » permet de baisser les barrières à l'entrée sur le secteur considéré. Il en résulte une concurrence plus forte des nouveaux entrants pour les fondateurs historiques. Finalement, la formation d'un pool conduit à une augmentation de la concurrence entre les acteurs fondateurs du pool mais également avec les acteurs hors du pool.
3. **Accroissement des risques de litiges** : la formation d'un pool a pour but de réduire les litiges entre les acteurs du « pool » et les acteurs hors du « pool ». Dans les faits, il apparaît que les acteurs hors du pool sont plus enclin à faire l'objet de litiges par les acteurs du « pool ». En effet, un « pool » ayant pour principe de regrouper en son sein un grand nombre de technologies, les membres d'un « pool » vont agir comme une seule entité afin de faire respecter leur droit de propriété intellectuelle. Si on se réfère à l'exemple de la formation du « pool » « *Sewing Machine Combination* », il a été constaté que le nombre de litiges avec les organisations hors du « pool » a significativement augmenté (Lampe & Moser, 2010).
4. **Modification de l'orientation des investissements en recherche et développement** : la création d'un « pool » a pour principal objectif de promouvoir les différentes technologies grâce à un accès simplifié. En facilitant l'accès à une technologie, on va mécaniquement augmenter le nombre d'utilisateurs de cette même technologie et il est admis que plus le nombre d'utilisateurs d'une technologie est important, plus la valeur d'usage de cette technologie s'accroît, c'est un effet de réseau que l'on détaillera plus

longuement dans la section suivante. Cet accroissement de valeur va inciter les organisations du « pool » mais également celles extérieures au « pool » à innover en orientant leurs dépenses en recherche et développement afin de profiter des avantages des effets de réseau. Cette orientation induite va conduire les acteurs à utiliser une technologie commune qui va être considérée comme un standard pour le secteur. Cependant, l'avènement de ce standard ne va pas inciter les acteurs du « pool » ni ceux à l'extérieur du « pool » à proposer de nouvelles technologies, le taux d'innovation dans le secteur sera donc enclin à baisser.

Ce cadre analytique indique que l'accès aux technologies grâce à la formation d'un pool de brevets entraîne des mécanismes d'incitation à l'innovation complexes entre les acteurs d'un même secteur. D'une part, ils peuvent être positifs grâce à la baisse des coûts d'accès aux technologies et aux avantages des premiers entrants (i.e. les fondateurs du pool). D'autres part, ils peuvent être négatifs du fait de l'accroissement de la concurrence et des litiges entre les acteurs du « pool » et ceux hors du « pool ». De plus, il apparaît que la formation d'un pool pourrait avoir un impact sur l'orientation technologique d'un secteur. Ce point conduit à la standardisation des technologies du « pool » ayant pour effet de diminuer l'incitation à innover des acteurs du secteur considéré.

Vakili a testé ce cadre analytique à travers l'étude de la formation du pool des technologies MPEG2. Le pool sur les technologies MPEG2 a été créé en 1997 avec initialement 27 brevets provenant de 9 organisations (8 entreprises et une université). Le pool s'est rapidement agrandi avec 226 brevets en 1998 et en 2011, il possédait 1000 brevets provenant de 27 organisations et 1518 licenciés. Son étude met en évidence que la formation du « pool MPEG2 » a conduit à une baisse de l'incitation à innover, un résultat similaire au résultat obtenu lors de l'analyse du pool « *Sewing Machine Combination* ». Cette baisse de l'incitation à innover est principalement due à l'orientation des dépenses en recherche et développement vers les technologies mise en avant à travers le pool (hypothèse 4 du cadre analytique). En effet, la promotion des technologies du pool va conduire à une augmentation du nombre d'utilisateurs de ces technologies. Avec l'accroissement du nombre d'utilisateurs, les acteurs du pool vont rendre compatible leurs prochaines technologies (« *follow-on technologies* ») avec les technologies du pool afin de profiter du bassin d'utilisateurs existants. En rendant compatible les technologies, on oriente mécaniquement les dépenses en recherche et développement et par conséquent la capacité d'innovation.

La formation d'un pool n'est donc pas anodine pour un secteur d'activité mais elle va également engendrer une simultanéité de comportements de coopération et de compétition entre les différents acteurs du pool. Cette simultanéité de comportements de coopération et de compétition va se matérialiser, d'une part, lors de la phase de constitution du pool, et d'autre part, lors de son fonctionnement (Ayerbe & Azzam, 2015). Dans la phase de constitution du pool, les organisations participant au pool vont être amenées à coopérer davantage entre eux par l'intégration de leurs brevets respectifs au sein du pool. Ces brevets sont majoritairement des brevets dits « essentiels » pour le secteur d'activité et leur accès « libre » à travers le pool permet d'accélérer la collaboration entre les organisations. Cependant lors de cette phase de constitution du pool, les différentes organisations vont également vouloir influencer l'orientation technologique du pool en fonction de leurs portefeuilles de brevet respectifs. Les participants vont ainsi souhaiter intégrer le maximum de brevets issus de leur portefeuille pour démontrer leur supériorité technologique face aux autres acteurs du pool (Lévêque, 2007). On appelle ce phénomène une « course aux brevets » (Hall & Ziedonis, 2001) : si on prend l'exemple du pool MPEG2 à sa constitution, les brevets des entreprises Philips et Sony représentaient 69 % de l'ensemble des brevets du pool qui était composé de 9 fondateurs.

2.2.2 Coopétition et émergence d'un standard technologique

Les différents accords détaillés précédemment mettent en évidence que l'enjeu derrière la coopération au sein d'un secteur d'activité, et cela même si les organisations se concurrencent, est l'accès aux technologies essentielles. L'accès ou le non-accès à ces brevets dits essentiels va donc être vital pour les différents acteurs d'un secteur. En effet, de nombreuses études scientifiques ont montré que le processus de création d'une invention est souvent cumulatif, c'est-à-dire qu'il concentre des technologies existantes. Dans leur revue de la littérature, Hall & Harhoff, (2012) ont mis en évidence que suivant l'invention concernée, l'utilisation des brevets pouvait différer. On distingue, d'une part, l'invention « séquentielle », le brevet d'une invention de première génération va permettre la commercialisation d'une ou plusieurs inventions de deuxième génération ; et d'autre part, l'invention « fragmentée », plusieurs brevets d'invention de première génération vont contribuer au dépôt d'un seul brevet d'une invention de deuxième génération.

Dans le cas de l'invention « séquentielle », l'ordre de création est un critère important car un seul brevet va permettre la création d'une ou plusieurs technologies. Le détenteur du brevet va

profiter de son droit de propriété pour imposer des licences d'exploitation aux autres acteurs du secteur d'activité. Cette situation conduit souvent à des litiges juridiques afin d'encadrer ce coût de licence et ainsi éviter un phénomène de « hold-up » qui permettrait au détenteur du brevet essentiel de pratiquer des tarifs élevés (Contreras, 2013) .

Dans le cas de l'invention « fragmentée », plusieurs brevets vont être nécessaires à la fabrication d'un seul brevet et on considère que les brevets sont disponibles en même temps sans ordre précis. Dans ces conditions, les organisations sont confrontées au « *patent thicket* » que Shapiro (2001) décrit comme « *un dense réseau de droits de propriété intellectuelle qui se cumulent et au travers duquel une entreprise doit se frayer un chemin pour parvenir à commercialiser une nouvelle technologie* ». De ce fait, ils devront donc acquérir l'ensemble des licences des différents brevets afin de pouvoir déposer leur propre brevet. Il en résulte une augmentation des coûts de transaction et les organisations ne sont plus incitées à innover du fait de l'augmentation globale des coûts (coûts de licences et coûts de R&D propres). Heller (1998) a exposé ce problème en le désignant par la « tragédie des anti-communs ». Ce problème est un problème récurrent dans le secteur des TIC (technologies de l'information et de la communication) où les technologies sont souvent la résultante de combinaisons de technologies, comme la technologies MPEG2 où 27 brevets initiaux ont été nécessaires à cette technologie.

En résumé, les brevets essentiels sont des brevets non substituables, sans eux il est impossible de reproduire la technologie ciblée. Dans le cadre du pool de brevets MPEG2, les neuf membres fondateurs du pool ont fait appel à un organisme extérieur pour déterminer quels étaient les brevets essentiels à la technologies MPEG2 et qui devait obligatoirement intégrer le pool. Ce choix des brevets essentiels a d'ailleurs été validé par le département de justice américain¹² au moment de justifier la légitimité du pool.

L'accès aux brevets essentiels, que ce soit dans le cadre d'une invention « séquentielle » ou « fragmentée » est donc un enjeu considérable pour tous les acteurs d'un secteur d'activité. C'est d'ailleurs dans ce cadre que les différentes modalités de coopération, mentionnées plus haut, se mettent en place et il est possible de les synthétiser de la manière suivante :

¹² [Justice Department Approves Proposal for Joint Licensing of Patents Essential for Meeting Video Technology Standard Used in Electronics and Broadcast Industries ;
https://www.justice.gov/archive/atr/public/press_releases/1997/1173.htm](https://www.justice.gov/archive/atr/public/press_releases/1997/1173.htm)

- Invention séquentielle : accès à un brevet essentiel, mise en place de licences entre les acteurs, simples ou croisées suivant la technologie.
- Invention fragmentée : accès à plusieurs brevets essentiels, mise en place d'un « pool » technologiques.

Ces différentes modalités de coopération, en facilitant l'accès aux brevets essentiels, vont permettre d'accélérer les développements technologiques au sein du secteur d'activité considéré. En effet, les différents acteurs sont incités à innover car ils ne sont plus freinés dans le développement de leurs propres technologies pour des raisons d'inaccessibilité aux brevets majeurs du secteur. Du fait d'un accès aisé, ces brevets vont également voir leur pouvoir renforcer sur le secteur d'activité considéré car plus une technologie est utilisée par un grand nombre d'utilisateurs et plus sa valeur économique s'accroît, on nomme ce principe « effet de réseau » (Shapiro & Varian, 1999). Les effets de réseau ont largement été étudiés dans la littérature économique et de nombreux secteurs d'activité ont profité de ce type d'effet pour se développer rapidement, on peut notamment citer les exemples suivants :

- Secteur bancaire avec le système de paiement par carte bancaire
- Secteur des télécommunications avec le système « Global System for Mobile Communications » (GSM)
- Secteur de l'informatique avec l'architecture unifié « x86 » des semi-conducteurs

Les effets de réseau, basés sur le principe qu'une technologie est utilisée par un grand nombre d'utilisateurs, vont fortement impacter l'orientation technologique du secteur d'activité bénéficiant de cet effet du fait de trois caractéristiques majeures (Contreras, 2015b)

- **Interopérabilité** : les acteurs du secteur vont développer des technologies qui vont être compatibles avec la technologie bénéficiant de l'effet de réseau afin de toucher le plus grand nombre d'utilisateurs.
- **Economie d'échelle** : en touchant un plus grand nombre d'utilisateurs, les organisations vont pouvoir faire des économies d'échelle dans la production de leur technologie.
- **Coûts de changement élevé et effets de « cadenas »** : en basant leur technologie sur une technologie bénéficiant d'un effet de réseau, les organisations vont avoir un coût de changement technologique élevé du fait des deux premières caractéristiques « interopérabilité » et « économie d'échelle », ils seront donc peu enclins à modifier leur technologie, c'est l'effet de « cadenas ».

Au travers de ces trois caractéristiques des effets de réseaux, l'accès aux brevets essentiels va conduire mécaniquement à l'émergence d'un standard technologique. On peut synthétiser le processus de la manière suivante (Figure 2.2) :

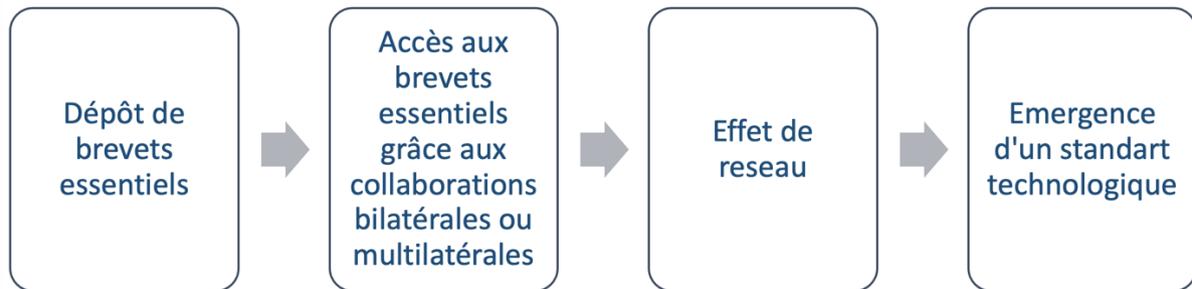


Figure 2.2 Processus d'émergence d'un standard technologique

Source : Charles Xavier (2023)

L'émergence d'un standard technologique justifie que des acteurs d'un même secteur, et cela même s'ils sont concurrents, collaborent entre eux. En effet, comme mentionné plus haut, les organisations souhaitent éviter un effet de « hold-up » technologique, à savoir qu'un seul acteur du secteur accapare le standard technologique et bénéficie ainsi des avantages des effets de réseau qui en découlent. Ce constat est particulièrement fort dans le secteur des TIC où de nombreux standards existent. On peut notamment citer les technologies « *Universal Serial Bus* » (USB), les connectivités Wifi et Bluetooth. Toutes ces technologies ont fait l'objet d'accords entre les acteurs du secteur, notamment à travers la création de pool de brevets. Quand de nombreux standards cohabitent au sein d'un secteur, on parle alors d'un « écosystème business » (Gueguen, 2009) et dans ce contexte, les comportements de coopération sont nécessaires afin de rendre compatible l'ensemble des standards au sein de l'écosystème.

On peut citer l'exemple des terminaux mobiles intelligents qui étaient fabriqués par de nombreuses entreprises et utilisaient des systèmes d'exploitation différents selon les modèles, comme le montre la Figure 2.3 :

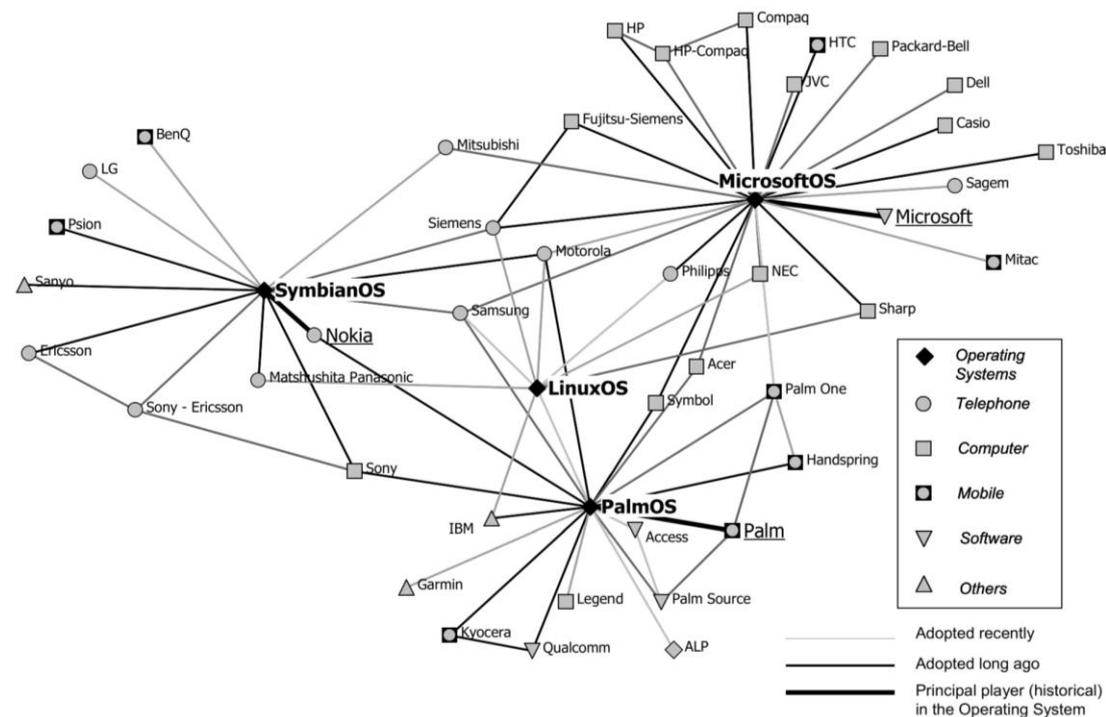


Figure 2.3 Terminaux mobiles intelligents - utilisation des principaux systèmes d'exploitation 1998–2006 (source : Gueguen, 2009)

Entre 1998 et 2006, 4 systèmes d'exploitation ont cohabité sur le marché des terminaux mobiles intelligents : PalmOS, SymbianOS, LinuxOS et MicrosoftOS. Certains constructeurs ont fait le choix d'utiliser un seul des 4 systèmes comme Toshiba ou Dell qui utilisaient exclusivement le système développé par Microsoft. D'autres acteurs ont fait le choix d'utiliser plusieurs systèmes, comme Motorola ou Samsung qui équipaient, selon les modèles, leurs terminaux avec un système d'exploitation Symbian, Palm ou Microsoft. Il est intéressant de noter que Symbian est le système promu par le constructeur Nokia et PalmOS, celui promu par Palm, deux concurrents directs de Motorola et Samsung. Motorola et Samsung ont cependant choisi de rendre leurs appareils compatibles avec les systèmes développés par leurs concurrents afin d'intégrer leur écosystème. Cet exemple montre bien que quand plusieurs technologies émergent sur un secteur, les acteurs vont choisir un comportement de coopération afin de rendre les différentes technologies compatibles entre elles.

En synthèse, il apparaît donc que les acteurs d'un secteur vont adopter un comportement de coopération afin d'accéder aux technologies essentielles de ce secteur, et des effets de réseau vont alors se développer et conduire à l'émergence de standards technologiques. Ce

comportement de coopération va perdurer dans le temps afin d'assurer l'interopérabilité des technologies pour tirer avantage des standards établis et des avantages qui en découlent.

Dans le cadre des brevets sur les technologies de traitement Big Data, la technologie MapReduce, qui a fait l'objet d'un dépôt de brevet par l'entreprise Google en 2004, est considéré comme une technologie fondatrice de ce secteur. Ce brevet a été accordé en 2010 par l'office américain des brevets (USPTO) et cette validation a grandement inquiété¹³ les acteurs du secteur du fait que cette technologie est le cœur de la structure Hadoop, structure en accès libre depuis 2006 au travers de la fondation Apache. La combinaison MapReduce/Hadoop est le socle de nombreuses technologies de traitement de données Big data (voir schéma 1 plus haut). Afin de rassurer le secteur, l'entreprise Google a souhaité que la technologie MapReduce soit toujours accessible et a donc décidé, dès 2013, de mettre sous forme de « *Patent Pledge* » 10 brevets affiliés à la technologie MapReduce¹⁴. A travers ce mécanisme, Google a souhaité permettre que les acteurs du secteur ne soient pas contraints par le brevet MapReduce et puissent continuer à utiliser cette technologie soit à travers la structure Hadoop portée par la fondation Apache soit directement en accèdent librement au brevet.

La section suivante va donc s'intéresser plus spécifiquement aux collaborations dans le secteur des technologies des traitements Big data, que ça soit à travers la fondation Apache ou un accès simplifié aux brevets par le mécanisme des Patent Pledge.

2.3 Innovation ouverte et collaborations au sein des technologies « Big Data »

A travers l'étude du brevet MapReduce, on constate deux types de mécanisme d'innovation ouverte qui incitent la collaboration entre les acteurs du secteur, d'une part, la fondation Apache et d'autre part, l'accès libre aux brevets grâce aux patent pledge. Comme indiqué plus haut, la technologie MapReduce est une des composantes essentielles de la technologie Hadoop, qui est un projet central de la fondation Apache. Le projet Hadoop est d'ailleurs le projet avec le plus grand nombre de contributeurs (242 contributeurs) parmi les 350 projets que gère actuellement

¹³ [Google's MapReduce patent: what does it mean for Hadoop? | Ars Technica ; https://arstechnica.com/information-technology/2010/01/googles-mapreduce-patent-what-does-it-mean-for-hadoop/](https://arstechnica.com/information-technology/2010/01/googles-mapreduce-patent-what-does-it-mean-for-hadoop/)

¹⁴ [Google Pledges Not To Sue Open Source Developers, Users And Distributors Over 10 MapReduce Patents \(Unless First Attacked\) | TechCrunch ; https://techcrunch.com/2013/03/28/google-pledges-not-to-sue-open-source-developers-users-and-distributors-over-10-mapreduce-patents-unless-first-attacked/?guccounter=1](https://techcrunch.com/2013/03/28/google-pledges-not-to-sue-open-source-developers-users-and-distributors-over-10-mapreduce-patents-unless-first-attacked/?guccounter=1)

la fondation. De nombreuses organisations et ingénieurs collaborent donc sur les technologies MapReduce/Hadoop au travers de la fondation Apache. De plus, le mécanisme des Patent Pledge permet d'accéder au brevet d'une technologie de façon libre à partir du moment où les conditions d'utilisation de la technologie stipulées par le propriétaire du brevet sont respectées. Comme mentionné dans la section précédente, quelques années après l'obtention de son brevet, l'entreprise Google a mis sous forme de Patent Pledge la technologie MapReduce (ainsi que 6 brevets affiliés) afin de ne pas bloquer les utilisateurs de cette technologie et cela notamment dans le cadre du développement de la technologie Hadoop. Les prochaines sections vont s'intéresser à ces deux mécanismes de collaboration.

2.3.1 Innovation ouverte « communautaire » : la fondation Apache

2.3.1.1 Présentation de la fondation Apache

La fondation Apache est un organisme à but non lucratif créé en 1999 dont l'objectif est de favoriser le développement de solutions technologiques « open source » afin de les rendre accessible au plus grand nombre. La fondation a été créée à l'initiative de huit développeurs à l'origine du serveur « Apache http » en 1995, un des serveurs web les plus anciens. Cette fondation avait pour but initial de maintenir à jour le serveur « Apache http » grâce à la communauté d'utilisateurs qui s'était développée sur les premières années du projet. Le développement de cette communauté a conduit à l'émergence de nouveaux projets indépendants du projet initial mais toujours en lien avec le secteur des technologies de l'information et de la communication. On peut notamment citer certains projets récents tels que « Airflow », une plate-forme de gestion de flux de travail à l'initiative de l'entreprise Airbnb en 2014, « Ivy » un outil de gestion des dépendances développé par la fondation elle-même en 2010 ou encore « Tomcat », un serveur d'applications développé également par la fondation en 2006. En ce qui concerne les applications pour le secteur des technologies de traitement « Big Data », les projets les plus connus, en dehors de « Hadoop » cité plus haut, sont « Cassandra » et « Hive » développés par l'organisation Facebook (Méta) en 2008 et 2011 respectivement. Ces deux applications permettent une meilleure gestion d'une grande quantité de données. En effet, d'une part, « Hive » est une infrastructure de données simplifiant l'analyse et le requêtage grâce à un langage proche du langage SQL et d'autre part, « Cassandra » permet de gérer des quantités importantes de données sur un grand nombre de serveurs en limitant les défaillances.

En dehors de ces deux applications spécifiques aux traitements de données de masse, on peut également citer l'application « *Kafka* », une plateforme de traitement *stream* qui a été initialement développée par l'organisation LinkedIn puis léguée à la fondation Apache en 2011. Il est intéressant de noter que nombre de ces projets cités sont à l'initiative d'entreprises privées (Yahoo, Airbnb, Facebook, ou encore LinkedIn) et plus récemment en 2016, Google a souhaité transmettre à la fondation Apache son application « *Dataflow* », une technologie de traitement de données, successeur déclaré de la technologie « MapReduce ». Actuellement on dénombre 320 projets actifs au sein de la fondation Apache portés par plus de 8000 contributeurs.

De nombreuses organisations privées du secteur des technologies de l'information et de la communication sont donc très actives au sein de la communauté Apache. La question est alors de savoir quel est l'intérêt pour ces organisations de rejoindre une communauté collaborative telle que la fondation Apache ? Avant de s'intéresser spécifiquement à cette question, il est tout d'abord nécessaire de comprendre le fonctionnement de la fondation Apache.

2.3.1.2 Le fonctionnement de la fondation Apache

Pour rejoindre les projets portés par la fondation Apache, il est nécessaire de suivre dans un premier temps une phase d'incubation nommée « projet incubateur » (Yang et al., 2022) . Cette phase, obligatoire, a été formalisée dès 2002 et permet de garantir l'intérêt technologique de tous les projets externes désirant rejoindre la fondation. En effet, cette phase intronise le cadre de gouvernance propre à la fondation que doivent suivre les différents projets, elle se nomme « la voie Apache » (« *The Apache Way* » en version originale).

De l'aveu même de la fondation Apache, il est difficile de définir précisément « *la voie Apache* » mais la fondation liste sept critères fondamentaux à respecter (source : www.apache.org) :

1. **Autorité acquise** : tous les individus ont la possibilité de participer, mais leur influence est basée sur le mérite acquis publiquement en lien avec leur contribution à la communauté. Ce mérite est propre à chaque individu, il ne dépend pas de l'employeur de l'individu, et n'est pas transférable.
2. **Communauté de pairs** : ce sont les individus qui participent à la fondation Apache et non les organisations. La structure à plat de la fondation stipule que les rôles sont égaux quel que soit le titre, les votes ont le même poids et les contributions sont faites sur la base du volontariat et cela même si les individus sont rémunérés pour leur travaux sur les projets.

3. **Communications ouvertes** : la fondation Apache exige que toutes les communications liées à un code et à la prise de décision soient accessibles au public afin d'assurer une collaboration asynchrone.
4. **Prise de décision par consensus** : les projets Apache sont supervisés par une équipe auto-sélectionnée de bénévoles actifs qui contribuent à leurs projets respectifs. Les projets sont auto-gouvernés avec une forte tendance à rechercher un consensus pour maintenir l'élan et la productivité. Bien qu'il ne soit pas possible d'établir un consensus total à tout moment, la tenue d'un vote ou une autre coordination peut être nécessaire pour aider à éliminer tout blocage.
5. **Supervision responsable** : Le modèle de gouvernance de la fondation Apache est basé sur la confiance et la supervision déléguée. Plutôt que des règles détaillées et des structures hiérarchiques, la gouvernance de la fondation est basée sur des principes, avec des projets autonomes fournissant des rapports directement au conseil d'administration. Les contributeurs Apache s'entraident en effectuant des validations par des pairs, en utilisant des mesures de sécurité obligatoires, en garantissant la conformité des licences et en protégeant la marque Apache et la communauté dans son ensemble contre les abus
6. **Indépendance** : la fondation Apache est strictement neutre vis-à-vis des fournisseurs. Aucune organisation ne peut obtenir de privilèges spéciaux ou contrôler la direction d'un projet, indépendamment de l'emploi des contributeurs travaillant sur des projets Apache.
7. **Communauté au-dessus du code** : ce critère est une maxime affirmant que le bien-être de la communauté est prioritaire par rapport à l'excellence d'un code.

Tous les projets intégrant le « projet incubateur » doivent obligatoirement adopter une gouvernance répondant à ces critères. A la fin de la phase d'incubation, si le projet répond aux exigences technologiques de la communauté il est validé, dans le cas contraire il est retiré. Il est important de noter qu'un projet validé peut également être retiré si la suite de son développement ne répond plus aux exigences de la fondation.

La Figure 2.4. décrit le cycle de vie pour les projets rejoignant la fondation Apache.

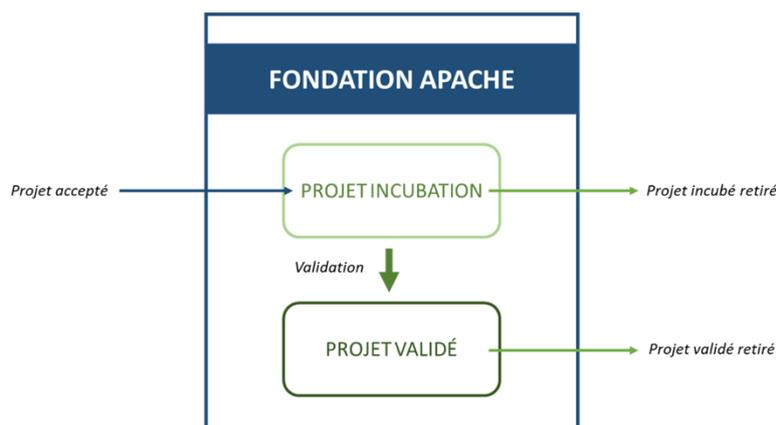


Figure 2.4 Cycle de vie d'un projet au sein de la fondation Apache

(source : Yang et al, 2022)

Si on synthétise les critères de « *la voie Apache* », on constate que le mode de gouvernance retenu par la fondation Apache est un mode de gouvernance décentralisée, chaque projet ayant sa propre autonomie et fédérant une communauté de contributeurs volontaires. Il est important de noter que le mérite revient aux individus et non aux organisations même si celles-ci rémunèrent les contributeurs pour leurs travaux. Sur ce point, la fondation souhaite affirmer son indépendance vis-à-vis des organisations privées.

Cependant, on constate que la fondation Apache possède également un programme de parrainage, ayant quatre niveaux, qui est fonction du niveau de donation :

- Niveau « Platinum » : 125 000 dollars par an
- Niveau « Gold » : 50 000 dollars par an
- Niveau « Silver » : 25 000 dollars par an
- Niveau « Bronze » : 6 000 dollars par an

En fonction du niveau de parrainage choisi, on a accès à différents privilèges au sein de la fondation Apache, notamment pour le niveau « Platinum » et « Gold » la possibilité de co-écrire les études de cas de la fondation. Parmi les organisations « Platinum » on trouve notamment Google, Microsoft, Facebook, Apple ou Amazon. Les parrains « Gold » sont, par exemple, les organisations IBM, Baidu, Salesforce ou Indeed.

La fondation Apache perçoit donc un financement externe dont les principaux donateurs sont des organisations privées qui se concurrencent au sein du secteur des technologies de l'information et de la communication. Quelles sont les raisons qui incitent ces organisations privées à financer une fondation communautaire revendiquant son entière indépendance ?

2.3.1.3 Intérêt pour les organisations privées de collaborer au sein de la fondation Apache

Pour comprendre l'intérêt des organisations privées à rejoindre un projet collaboratif au sein de la fondation Apache, il est important de comprendre le principe de la licence Apache, licence qui est appliquée à tous les projets soutenus par la fondation.

En général, les conditions d'utilisations d'une technologie sont régies par les droits de propriété intellectuelle accordés aux inventeurs. Les inventeurs ont toutes la jouissance de la technologie ; ils peuvent l'utiliser, la modifier, la vendre, la céder selon leurs volontés. Par opposition à cette vision fermée, il existe des licences dites ouvertes souvent nommés « *copyleft* » (en opposition à « *copyright* ») pour lesquelles l'inventeur d'une technologie autorise un tiers à utiliser, étudier, modifier et diffuser cette technologie, à partir du moment où cette même autorisation reste préservée. Dans le cas d'une licence initiale « *copyleft* », il est donc impossible d'imposer par la suite une utilisation fermée de la technologie (Charleux, 2020).

Dans le cas de la licence de la fondation Apache, une approche hybride a été retenue (Chang et al., 2007). En effet, une technologie sous licence Apache peut être librement utilisée, reproduite, modifiée, distribuée ou vendue. Cependant, une technologie sous licence Apache peut-être également combinée avec d'autres technologies, la nouvelle technologie ainsi créée peut alors être revendue et distribuée sous une autre licence que celle de la fondation Apache. Il faut néanmoins indiquer clairement que cette nouvelle technologie est basée sur une ou des technologies issues de projets de la fondation Apache.

En résumé, toutes les technologies sous licence de la fondation Apache peuvent donc être utilisées pour des projets commerciaux développés par un tiers, à partir du moment où celui-ci indique que la technologie initiale provient d'un projet de la fondation Apache. Cette vision hybride permet donc aux organisations privées d'exploiter à leurs propres fins les technologies issues des projets Apache. On peut citer l'exemple du serveur applicatif « *WebSphere* » de l'organisation IBM dont un des plugins a comme base le serveur Apache http. A travers cet

exemple, on voit clairement l'intérêt pour les grandes entreprises du secteur des technologies de l'information et de la communication de contribuer fortement aux activités de la fondation Apache.

Cependant, on constate également que de nombreuses petites entreprises, startups et laboratoires universitaires contribuent fortement aux projets de la fondation Apache. Pourquoi ces organisations, agiles et souvent moteur dans la création de nouvelles technologies innovantes, auraient un intérêt dans le développement de projets de la fondation Apache, au risque de voir leurs travaux être utilisés librement par des grands groupes technologiques ?

Yang et al (2022) ont étudié cette question à travers l'analyse de 476 projets Apache et ils ont mis en évidence quatre points motivant ce choix :

- **Visibilité de la technologie** : incuber une solution technologique au sein de la fondation Apache permet de donner de la visibilité à sa technologie du fait de l'importante communauté gravitant au sein de la fondation. Les ingénieurs vont ainsi promouvoir leurs technologies au travers des réseaux des différents contributeurs qui composent la fondation. On peut citer l'exemple des fondateurs du projet *DataSketches* qui ont mentionné dans leur motivation à rejoindre la fondation Apache « *Nous pensons que le fait d'avoir DataSketches en tant que projet Apache apportera une contribution immédiate, intéressante et substantielle à la communauté open source et aura une meilleure opportunité de fournir une contribution significative à la fois à la science et à l'ingénierie des algorithmes d'esquisse* ».
- **Profiter du modèle de gouvernance Apache** : en rejoignant la fondation Apache, on accepte de structurer son projet selon le modèle de gouvernance Apache basé sur la « voie Apache » qui valorise la méritocratie à travers sa communauté de contributeur. Cette structuration définie est un élément de stabilité à long terme, ce qui permet de rassurer les utilisateurs potentiels sur la pérennité du projet.
- **Amélioration de la technologie par la communauté** : la fondation Apache s'appuie sur une communauté de contributeurs. A partir du moment où un projet rejoint la communauté, l'ensemble des contributeurs peuvent l'améliorer, ce qui permet une innovation continue de la technologie. Ce point est une raison forte pour les petites entreprises à rejoindre la fondation Apache.
- **Rejoindre un écosystème de technologies confirmées** : la fondation Apache soutient plus de 300 projets et notamment des projets dont les technologies sont majeures pour leur secteur d'activité. On a déjà cité Hadoop, Pig, Cassandra pour le secteur des technologies de traitement des données Big Data. En rejoignant la fondation Apache, on rejoint également

l'écosystème de ces technologies, un élément essentiel quand on souhaite assurer l'interopérabilité de sa technologie avec ces technologies essentielles. Les créateurs de la technologie *HMS* ont ainsi indiqué : « *La fondation Apache est un hôte naturel pour HMS étant donné qu'il héberge déjà Hadoop, Pig, HBase, Cassandra et d'autres projets logiciels cloud émergents* ».

2.3.2 Innovation ouverte « unilatérale » : les patents pledge

Dès l'obtention officielle du brevet pour la technologie MapReduce en 2010, l'entreprise Google a rassuré les acteurs du secteur des technologies Big Data en indiquant qu'il n'imposerait pas de licence d'exploitation pour l'utilisation de cette technologie. L'entreprise a formalisé ce principe en 2013 en créant le dispositif « *Open Patent Non-Assertion Pledge*¹⁵ » qui intègre plus d'une centaine de brevets de l'entreprise, incluant les brevets sur la technologie MapReduce (le brevet initial et six brevets affiliés). Google indique qu'il ne déclenchera aucune procédure judiciaire envers les utilisateurs des brevets contenus dans ce dispositif à partir du moment où l'entreprise n'est pas attaquée en premier lieu. Google met en avant la vision défensive de ce dispositif et souhaite favoriser l'innovation en ouvrant l'accès à ces technologies.

Ce mécanisme d'accessibilité aux brevets d'une entreprise est connu sous le terme de « patent pledge ». Il est nécessaire de définir dans un premier temps le principe des patents pledge et notamment les conditions d'utilisation afférentes à ce type de mécanisme de propriété intellectuelle.

2.3.2.1 Le principe des patents pledge et les conditions d'utilisation

A travers le mécanisme des patents pledge, une organisation s'engage, sous certaines conditions, à accorder des licences de leurs brevets dit « pledge » à un public restreint ou illimité, moyennant une compensation monétaire « raisonnable » ou nulle, au moyen de contrats standardisés écrits ou sociaux (Ehrnsperger & Tietze, 2019). Ce premier point permet de bien distinguer la différence entre le principe des « pools » de brevets que nous avons décrit précédemment et le principe des « patents pledge ». Dans le premier cas, il s'agit d'un engagement multilatéral entre plusieurs organisations qui acceptent de mettre en commun leurs

¹⁵ [Open Patent Non-Assertion Pledge - Google \(https://www.google.com/patents/opnpledge/pledge/\)](https://www.google.com/patents/opnpledge/pledge/)

brevets, dans le second cas, c'est un engagement unilatéral et volontaire d'une organisation de mettre en accès libre ses brevets (Bronzo, 2022).

Il est important de préciser que l'engagement pris par les organisations à travers le mécanisme des patents pledge n'est pas un acte de renonciation à leurs droits de propriété intellectuelle, et à l'image de la fondation Apache, le titulaire des brevets « pledge » met systématiquement en place un cadre d'exploitation pour l'utilisation de sa propriété intellectuelle en accès libre. Ehrnsperger & Tietze, (2019) ont défini une taxonomie des conditions d'utilisation des patents pledge à travers trois éléments :

- **Accessibilité** : une organisation peut limiter l'accès à sa propriété intellectuelle ouverte à une certaine catégorie d'utilisateurs. Par exemple, Toyota autorise l'utilisation de ses brevets pledge uniquement « *aux constructeurs automobiles qui produiront et vendront des véhicules à pile à combustible, ainsi qu'aux fournisseurs de pièces détachées pour piles à combustible et aux sociétés énergétiques* ». Cette accessibilité peut également être illimitée, comme ce fut le cas en 2005 pour 500 brevets américains détenus par l'entreprise IBM.
- **Compensation** : une organisation peut demander une compensation financière directe pour l'exploitation de ses brevets « pledge », dans la majorité des cas, cette compensation est dite « raisonnable » ou nulle.
- **Conditions** : on distingue deux types de conditions :
 - **Les conditions directes** : on va limiter les conditions d'utilisation des brevets pledge à une utilisation spécifique, à une temporalité ou à un territoire. On peut citer l'exemple de l'entreprise Microsoft qui donne accès à ses brevets pledge à la communauté universitaire sous condition d'une utilisation non commerciale. On peut également prendre l'exemple du constructeur automobile Toyota qui a donné accès à ces brevets sur la pile à combustible jusqu'en 2020.
 - **Les conditions indirectes** : ce sont des conditions induites liées à l'exploitation des brevets pledge. Ces conditions portent notamment sur le respect de la propriété intellectuelle que confère initialement le brevet. Les utilisateurs des brevets pledge doivent respecter ce droit et ne pas le dénoncer juridiquement. On peut prendre l'exemple de l'entreprise Tesla qui précise dans ces conditions d'utilisations¹⁶ que « *Tesla s'engage irrévocablement à ne pas engager de poursuite contre toute partie*

¹⁶ [Additional Resources | Tesla Other Europe, https://www.tesla.com/en_eu/legal/additional-resources#patent-pledge](https://www.tesla.com/en_eu/legal/additional-resources#patent-pledge)

pour violation d'un brevet Tesla par le biais d'activités liées aux véhicules électriques ou à des équipements connexes, aussi longtemps que cette partie agit de bonne foi. Une partie « agit de bonne foi » tant que cette partie et ses sociétés liées ou affiliées n'ont pas : revendiqué, aidé d'autres personnes à faire valoir ou eu un intérêt financier dans toute revendication de (i) tout brevet ou autre droit de propriété intellectuelle contre Tesla ou (ii) tout droit de brevet contre un tiers pour son utilisation de technologies liées aux véhicules électriques ou à des équipements connexes »

Ce dernier exemple sur les conditions d'utilisation des brevets Tesla est une condition qui apparaît de façon récurrente dans les conditions de nombreuses organisations. On pourrait qualifier cette condition de **condition « défensive »** car elle oblige les utilisateurs à ne pas attaquer juridiquement le titulaire des brevets pledge dans le but de remettre en cause son droit de propriété intellectuelle. L'entreprise Google utilise une approche similaire dans son dispositif « *Open Patent Non-Assertion Pledge* » en indiquant « *Étant donné que notre engagement est une promesse de ne pas revendiquer certains brevets Google sans exiger aucun paiement de la part du destinataire de l'engagement, nous pensons qu'il est tout à fait juste que nous conditionnions l'engagement au fait que le destinataire de l'engagement (et ses sociétés affiliées) ne revendique pas de brevets ou ne profite pas de la revendication de brevets contre Google, ses sociétés affiliées ou ses produits ou services ».*

Dans les deux cas, les deux entreprises, Tesla et Google, n'engageront aucune poursuite judiciaire pour les exploitants de leurs brevets pledge ; ils donnent donc de façon **implicite** leur autorisation d'exploiter leurs brevets. Cette condition « défensive » est à mettre en parallèle avec la licence FRAND (« *Fair, Reasonable And Non Discriminatory* », qui se traduirait par « équitable, raisonnable et non discriminatoire »). Ce type de licence est fréquemment utilisée dans le cadre de l'utilisation de technologies que l'on considère comme des « standards » telles que les technologies « WIFI », « USB » ou « 4G ». Elle est également couramment utilisée dans le cadre de l'exploitation des patents pledge, on peut, par exemple, citer « *l'Open Covid Pledge*¹⁷ » qui regroupe de nombreuses organisations qui ont décidé, à travers ce dispositif, d'ouvrir une partie de leur propriété intellectuelle dans le cadre de la lutte contre la pandémie de Covid-19. A travers les conditions de cette licence FRAND, on souhaite encourager

¹⁷ [Open COVID Pledge - Open Covid Pledge, https://opencovidpledge.org](https://opencovidpledge.org)

l'utilisation d'un brevet que l'on considère comme essentiel (voir 1.2), cette licence donne de façon **explicite** l'autorisation d'exploiter un brevet.

Les différents exemples mentionnés précédemment montrent clairement que la mise sous forme de patent pledge est un choix murement réfléchi avec des conditions d'utilisations spécifiques pour chaque organisation. Ces conditions ont pour objectif d'encadrer l'exploitation des brevets pledge par un tiers, tout en réaffirmant les droits de propriété intellectuelle des organisations propriétaires de ces brevets. Cependant, on constate que de nombreux litiges se sont déclarés autour de certains brevets ouverts faisant l'objet d'une licence FRAND, on peut, par exemple, citer (source : cabinet Dhenne Avocats¹⁸) :

- Nokia vs Apple (2011) : litiges sur les brevets liés aux technologies de téléphonie mobile
- Samsung vs Ericsson (2012) : litiges sur les brevets liés aux technologies de communication mobile
- Federal Trade Commission vs Qualcomm (2017) : litiges pour tarification excessif par Qualcomm pour ses brevets sur les technologies de communication mobile et violation des lois antitrust.

Ces exemples montrent bien que l'ouverture de sa propriété intellectuelle n'est pas sans risques pour une organisation et cela même si elle a pris le soin de bien définir des conditions d'utilisations spécifiques.

Il est donc nécessaire de s'interroger sur les motivations qui poussent des organisations à rendre accessible leur propriété intellectuelle à travers le mécanisme des « patent pledge ».

2.3.2.2 Intérêt pour une organisation de recourir au patent pledge

L'accroissement du nombre de patents pledge dans différents secteurs d'activité montre l'engouement des organisations envers ce mécanisme de propriété intellectuelle. Il a été mentionné dans la section précédente que la mise sous forme des patents pledge est un acte réfléchi qui conduit à définir de manière précise les conditions d'utilisations des brevets en accès libre. Les organisations doivent avoir un réel intérêt à utiliser ce mécanisme des patents pledge et il apparaît que les motivations derrière cet acte sont nombreuses.

Contreras, (2015) a défini une taxonomie afin de répertorier les différents cas possibles :

¹⁸ [Avocat brevets | Propriété intellectuelle | Dhenne Avocats \(dhenne-avocats.fr\)](#)

▪ Cas n°1 : incitation

En ouvrant leur propriété intellectuelle, les organisations titulaires des brevets « pledge » vont inciter les organisations tierces à adopter les technologies promues par ces brevets. En effet, avec la mise sous forme de patent pledge, les barrières liées à la protection que confère un brevet sont réduites, avec notamment des coûts de licence plus faibles voire inexistantes dans la majorité des cas. Du fait de cet accès aisé aux technologies, les organisations tierces vont avoir une incitation plus forte à innover. Pour les organisations titulaires des brevets « pledge », l'augmentation du nombre d'utilisateurs de leurs technologies va induire des effets de réseaux (voir 1.2) qui peuvent conduire à l'émergence d'un standard technologique. Comme mentionné dans la section 1.2 ; la recherche d'un standard technologique est un objectif majeur pour de nombreuses organisations, notamment sur le secteur des TIC.

La volonté d'établir un standard technologique est une des raisons principales qui peut motiver les organisations à ouvrir leur propriété intellectuelle mais ce n'est pas l'unique raison, le développement d'un marché émergent peut également encourager ce type de mécanisme. En effet, le dynamisme d'un marché est fortement lié à la concurrence entre les différents acteurs de ce marché (Porter, 1991). Si un acteur possède une avance technologique, les autres acteurs ne seront pas incités à entrer sur ce marché du fait de la barrière technologique. Le marché ne se développera pas en termes d'offres, ce qui conduira à une demande limitée et donc à un marché avec de faibles gains potentiels. En abaissant les barrières technologiques grâce à l'ouverture de sa propriété intellectuelle, une organisation va inciter de nouveaux entrants et donc stimuler son marché. L'exemple de l'entreprise Tesla illustre bien ce propos : en mettant sous forme de patent pledge l'ensemble de ses brevets en 2014, Tesla a souhaité stimuler le marché des véhicules électriques (Baudry et Dumont, 2020). Suite à cette annonce, l'entreprise Toyota a également ouvert ses brevets sur la technologie des véhicules électriques, et on a pu constater sur les années suivantes une croissance importante du parc de véhicules électriques dans le monde, comme le montre la Figure 2.5.

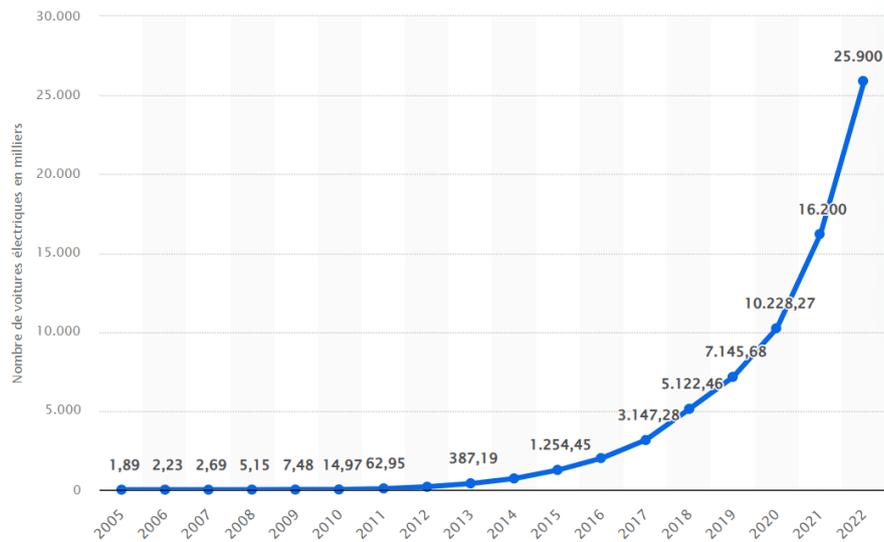


Figure 2.5. Parc de voitures à batterie électrique et de voitures hybrides rechargeables dans le monde de 2005 à 2022 (source Statista, 2023)

Il est intéressant de noter que la recherche d'un standard technologique et la volonté de développer un marché sont souvent des objectifs complémentaires. La mise en place d'un standard sur un secteur d'activité va induire une interopérabilité entre les technologies, les différents acteurs souhaitant profiter des bénéfices liés aux effets de réseaux. Cette interopérabilité entre les technologies va également permettre de baisser les barrières technologiques qui peuvent être un frein à l'entrée de nouveaux entrants, ce qui va donc permettre de dynamiser le marché. Dans l'exemple des véhicules électriques, c'est la volonté des acteurs du secteur de standardiser les batteries et bornes de recharge qui a été un des facteurs de dynamisme du marché. C'est d'ailleurs pour cette raison que les premiers brevets « pledge » initiés par l'entreprise Tesla portaient sur des technologies liées aux batteries électriques.

▪ **Cas n°2 : action collective**

Les actions entreprises par un acteur sur un secteur d'activité peut influencer les décisions des autres acteurs de ce même secteur. En ouvrant sa propriété intellectuelle, une organisation peut souhaiter induire une action collective similaire chez les autres organisations de ce secteur. L'organisation initiale n'est pas forcément à la recherche d'un standard technologique, elle cherche en premier lieu la réciprocité de son action d'ouverture de sa propriété intellectuelle. L'action collective est souvent initiée pour des domaines touchant à l'intérêt général tel que les sujets environnementaux. Contreras (Contreras, 2015a) cite l'exemple de l'entreprise IBM qui a été à l'origine du dispositif « *Eco-Patent Commons* » en 2008 avec 29 brevets de la firme. Ce dispositif avait pour but de rendre accessible des brevets favorisant la protection de

l'environnement. IBM a rapidement été rejoint dans son action par des entreprises telles que Sony ou Nokia.

▪ **Cas n°3 : Restriction volontaire**

Une organisation peut rendre accessible ses brevets afin de garantir que certaines technologies ne seront pas verrouillées par les droits que confère le dépôt de brevet. Ce fut le cas pour la technologie MapReduce dont le brevet a été déposé par l'entreprise Google en 2004 et accordé par l'office USPTO en 2010. Durant la période 2004-2010 de nombreuses entreprises ont utilisé la technologie MapReduce notamment à travers la fondation Apache et le projet Hadoop. La validation du brevet MapReduce par l'office USPTO en 2010 a grandement inquiété les différents acteurs du secteur TIC. L'entreprise Google, pour répondre à ces craintes, a volontairement restreint son droit de propriété intellectuelle sur les brevets affiliés à la technologie MapReduce en initiant le dispositif « *Open Patent Non-Assertion Pledge* ». Il est intéressant de noter que cette restriction volontaire peut également avoir comme objectif de rassurer des organismes gouvernementaux sur un éventuel abus de position dominante sur un marché. Contreras (2015) cite l'exemple des entreprises Apple, Microsoft et Google en 2012 qui ont chacun assuré au Ministère Américain de la Justice qu'ils ne déclencheraient aucune procédure judiciaire contre les organisations utilisant leurs technologies de télécommunications sans fil. Cette déclaration a été faite dans le cadre d'une opération de rachat de société et de portefeuille de brevet qui était en cours d'examen par le ministère. Il s'est avéré que le choix d'approuver ces acquisitions a été grandement influencé par la déclaration des trois entreprises. Enfin la restriction volontaire de son droit de propriété intellectuelle peut également être la résultante de pression extérieure, notamment d'organismes non gouvernementaux, ayant pour objectif de rendre accessible une technologie qui est considérée d'intérêt général. Durant la pandémie de Covid-19, de nombreux organismes ont appelé à lever les brevets sur les vaccins luttant contre le Covid-19, dès octobre 2020, le laboratoire Moderna¹⁹ mettait sous patent pledge ses brevets déposés sur la période 2018-2020.

▪ **Cas n°4 : Philanthropie**

Les organisations peuvent décider d'ouvrir leur propriété intellectuelle afin de faire avancer une cause sociale, environnementale ou d'intérêt général. Le dispositif « *Open Covid Pledge* » initié par des entreprises telles que Facebook, Amazon, IBM ou Microsoft au début de la pandémie

¹⁹ [Modernas Updated Patent Pledge \(modernatx.com\), https://investors.modernatx.com/Statements--Perspectives/Statements--Perspectives-Details/2022/Modernas-Updated-Patent-Pledge/default.aspx](https://investors.modernatx.com/Statements--Perspectives/Statements--Perspectives-Details/2022/Modernas-Updated-Patent-Pledge/default.aspx)

de Covid-19 en 2020 illustre bien ce propos. Ce dispositif indique clairement que ces entreprises se sont « *engagés à mettre gratuitement à disposition leur propriété intellectuelle afin de l'utiliser pour mettre fin à la pandémie de COVID-19 et minimiser l'impact de la maladie.* ». Derrière cette philanthropie se cache également des enjeux de relations publiques et d'image de marque. En effet, en participant à une cause d'intérêt général grâce à l'ouverture de sa propriété intellectuelle, une organisation envoie un signal positif à l'opinion publique. Il est à noter que la philanthropie des organisations est généralement concomitante avec leur intérêt personnel. En ouvrant leur propriété intellectuelle sur la technologie des véhicules électriques, Tesla et Toyota participent à une cause environnementale d'intérêt général : la réduction des gaz à effet de serre, tout en développant leur propre marché.

Au total, il apparaît que les motivations associées aux patents pledge révèlent un intérêt individuel pour les organisations souhaitant ouvrir leur propriété intellectuelle. La raison principale étant la recherche d'un standard technologique afin de bénéficier des effets de réseaux qui en découlent.

Comme nous avons pu le voir précédemment, les patent pledge sont présents dans de nombreux secteurs d'activité (TIC, automobile, laboratoire pharmaceutique...) ce qui montre un certain engouement pour ce mécanisme de propriété intellectuelle. Ce chapitre s'intéresse plus particulièrement à la technologie de traitement Big Data qu'est MapReduce, qui comme déjà indiqué précédemment est l'objet d'un patent pledge et est une des technologies phares de la fondation communautaire Apache grâce au projet Hadoop.

La section suivante a pour objet d'analyser l'impact de la technologie MapReduce sur le secteur des technologies des traitements Big Data. Il s'agit de comprendre si, à travers les deux mécanismes d'innovation ouverte que sont la fondation Apache et les patents pledge, cette technologie a permis de favoriser l'innovation au sein de ce secteur.

2.4 Stratégie d'analyse empirique

2.4.1 Modèles d'estimation et hypothèses

Dans la section précédente, nous avons pu constater que la technologie MapReduce et ses brevets associés ont fait l'objet d'une accessibilité forte grâce à deux mécanismes distincts :

- Intégration dans l'écosystème de la fondation Apache en étant une composante principale du projet Hadoop
- Accessibilité directe grâce au dispositif « *Open Patent Non-Assertion Pledge* » déployé par l'entreprise Google en 2013

L'objectif de cette partie empirique est d'évaluer si ces deux facteurs ont un impact sur le nombre de brevets déposés s'appuyant sur ces technologies qualifiées de *follow-on innovations* et si leur qualité technologique en a été affectée par la mesure du nombre de citations reçues.

Deux hypothèses²⁰ peuvent être formulées concernant cet impact :

Hypothèse 1a : Le libre accès à la technologie MapReduce et ses brevets associés devrait avoir un impact positif sur les innovations ultérieures fondées sur ces technologies.

L'impact est positif si les entreprises sont incitées à utiliser cette technologie qui n'induit pas de paiement de licence pour son utilisation, en étant sponsors de la fondation Apache ou directement par l'utilisation du Patent Pledge. Cette facilité d'accès réduit le coût de la recherche de nouvelles technologies, les négociations et les coûts de licence (Vakili, 2016).

Hypothèse 1b : Le libre accès à la technologie MapReduce et ses brevets associés devrait avoir un impact négatif sur les innovations ultérieures fondées sur ces technologies.

L'impact est négatif si les entreprises ne sont pas incitées à développer de nouvelles technologies et se limitent à exploiter les technologies développées par Google.

La stratégie empirique pour tester l'impact de la technologie MapReduce et ses brevets associés suit celle habituellement utilisée dans la littérature (Rysman & Simcoe, 2006; Singh & Agrawal,

²⁰ Il s'agit d'hypothèses semblables à celles formulées par Vakili (2016) concernant l'impact des patent pool sur le follow-on innovations.

2011; Vakili, 2016 ; Hall & Helmers, 2013 ; Contreras, Hall & Helmers, 2018 ; De Rassenfosse & Palangkaraya, 2023).

La méthodologie consiste à comparer les effets des deux mécanismes d'innovation ouverte sur deux échantillons :

- Un échantillon traité qui inclut les brevets MapReduce et associés ainsi que les « *follow-on innovations* » qui s'appuient sur la technologie MapReduce et associés, autrement dit, qui les citent.
- Un échantillon de contrôle qui a les mêmes caractéristiques que l'échantillon traité, à savoir, des brevets de la même classe technologique et déposé la même année et qui ne citent pas MapReduce et les follow-on innovations.

Il s'agit d'étudier les différences entre les brevets traités et les brevets de contrôle en termes de citations reçues.

$$CITATIONS_{i,t} = \beta_0 + \beta_P PLEDGE_i + \beta_A APACHE_i + \beta_X X_i + \beta_t + \delta_{t-appAnnée} + \epsilon_{i,t}$$

$CITATIONS_{i,t}$ représente le nombre de citations reçues par le brevet i (traité ou control) à l'année t .

Le choix du nombre de citation comme variable à analyser s'inspire de la littérature existante dans le domaine. En effet, de nombreuses études se sont appuyées sur cet indicateur pour mesurer la valeur technologique comme proxy de la valeur économique d'un brevet (Harhoff et al., 1999 ; Lanjouw & Schankerman, 2001 ; Hall et al., 2005 ; Frietsch et al. 2010) En s'appuyant sur les travaux de Hall et al., (2005), deux analyses sont effectuées, une analyse des citations conservant les autocitations des organisations et une analyse les excluant. Dans son analyse, Hall et al. (2005) indiquent que les autocitations montrent qu'une entreprise, qui se cite, occupe une position concurrentielle forte dans la technologie considérée et qu'elle sera en mesure d'internaliser une partie des retombées des connaissances créées par ses propres développements. Elle pourra donc innover sur ses propres technologies à moindre coût car sa dépendance vis-à-vis des technologies extérieures sera moindre. En conservant les autocitations, Hall et al. précisent que c'est la « *valeur privée* » du brevet qui est analysée. En excluant les autocitations, on va s'intéresser à l'impact du brevet dans son écosystème technologique, on dit alors que c'est la « *valeur sociale* » du brevet qui est étudiée.

Dans le cas des technologies de traitement des données Big Data, il était important de distinguer ces deux types de valeur car ce sont des technologies émergentes et de ce fait elles peuvent contribuer à valoriser individuellement les organisations grâce aux innovations qu'elles apportent (« *valeur privée* ») mais également participer à l'essor d'un secteur d'activité (« *valeur sociale* »).

L'estimation intègre deux variables d'intérêt $PLEDGE_i$ et $APACHE_i$ qui représentent, respectivement, des indicatrices égales à 1 si le brevet est un brevet pledge associé à la technologie MapReduce et si le déposant du brevet est un sponsors Apache. Au travers de ces deux indicatrices, nous allons capter l'impact des deux mécanismes d'innovation ouverte qui ont profité à la technologie MapReduce et à ses brevets associés.

Le vecteur X_i représente les variables qui permettent de contrôler les variations du nombre de citations reçues comme les caractéristiques des inventeurs (un plus grand nombre d'inventeurs contribue à accroître la complexité, la qualité technologique du brevet et à recevoir plus de citations), le domaine technologique comme certaines technologies peuvent recevoir plus de citations que d'autres même si le brevet de contrôle provient du même domaine. Il est également souhaitable de contrôler la provenance de la citation, à savoir, les facteurs de proximité entre le brevet et les brevets citant, autrement dit, est-ce que les inventeurs qui citent sont liés aux inventeurs qu'ils citent, est-ce qu'ils ont pu travailler pour le même déposant ou s'ils sont localisés au même endroit ?

Une estimation en différence-en-différence est utilisée pour distinguer l'impact après 2012 lorsque les brevets MapReduce ont été déclarés comme pledge. Il s'agit de voir si le nombre de citations a augmenté après 2012. Cette approche est intéressante car nous observons les citations reçues par les brevets traités et de contrôle non seulement après qu'ils aient été transformés en pledge mais aussi les années avant que la technologie MapReduce et ses brevets associés ne forment un pledge. Les citations reçues avant l'implémentation des pledge peuvent être utilisés comme benchmark. Cette approche nécessite une base de données qui inclut les observations des années non seulement après l'introduction des pledge mais aussi avant. Nous définissons la variable $Post - pledge_{i,t}$ comme une indicatrice qui prend la valeur 1 pour les observations qui s'appuient sur la technologie MapReduce devenue un pledge et pour laquelle l'année de citation est t intervenant après 2012.

L'équation s'écrit à présent :

$$\begin{aligned}
 CITATIONS_{i,t} = & \beta_0 + \beta_P PLEDGE_i + \beta_{PP} Post - pledge_{i,t} + \beta_A APACHE_i \\
 & + \beta_{PPP} Post - pledge_{i,t} \times PLEDGE_i \\
 & + \beta_{PPA} Post - pledge_{i,t} \times APPACHE_i + \beta_X X_i + \beta_t + \delta_{t-appAnnée} + \epsilon_{i,t}
 \end{aligned}$$

β_{PP} capte la différence systématique entre les brevets traités et de contrôle qui existe avant le passage en Pledge. Pour chaque brevet et chaque contrôle, on compte le nombre annuel de citations reçues après l'année du dépôt.

β_{PPP} et β_{PPA} captent la différence entre les deux groupes après le passage en Pledge.

Au total, la stratégie empirique retenue pour être synthétisée par la Figure 2.6 :

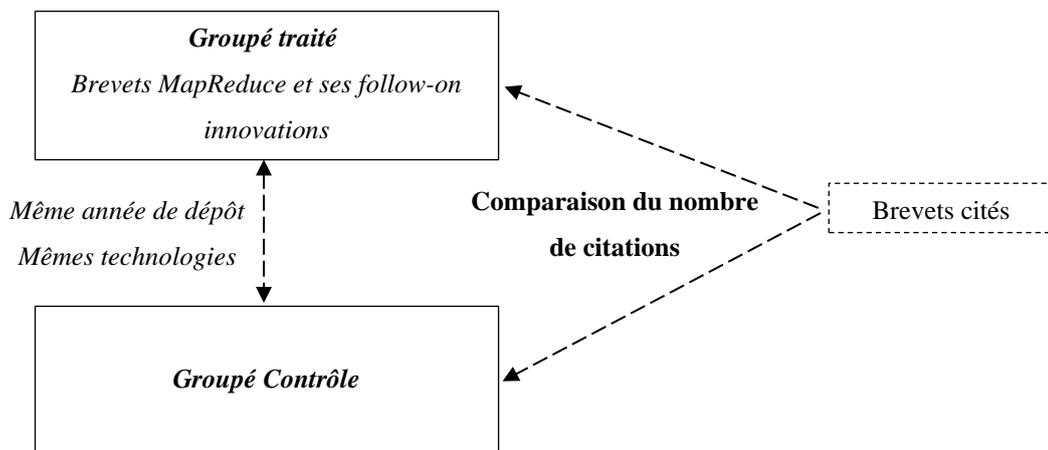


Figure 2.6 Impact de la technologie MapReduce et ses brevets associés sur les innovations ultérieures – stratégie empirique (Charles Xavier, 2023)

La section suivante va maintenant détailler la méthode de constitution des deux groupes de travail.

2.4.2 Constitution des deux groupes de travail

L'ensemble des données ont été collectées sur le site *patentview.org* qui est affilié à l'office de brevet *United States Patent and Trademark Office* (USPTO) sur la période 1994 à 2020.

Pour constituer le groupe traité, nous avons procédé à la même méthodologie que dans le chapitre 1 en collectant tous les brevets follow-on innovations du brevet MapReduce. Nous avons également collecté tous les brevets follow-on innovations des brevets que l'entreprise

Google décrit comme des brevets affiliés à la technologie MapReduce²¹. Le groupe traité se base donc sur 7 brevets qui sont décrit dans le tableau suivant :

Tableau 2.1 Liste des brevets associés à la technologie MapReduce

Numéro de publication du brevet	Date de dépôt du brevet	Titre du brevet	Commentaires
7650331	18 juin 2004	System And Method For Efficient Large-Scale Data Processing	<i>Brevet initial sur la technologie MapReduce</i>
7756919	18 juin 2004	Large Scale Data Processing In A Distributed And Parallel Processing Environment	
7590620	29 septembre 2004	System And Method For Analyzing Data Records	
8126909	31 juillet 2009	System And Method For Analyzing Data Records	
8612510	12 janvier 2010	System And Method For Large-Scale Data Processing Using An Application-Independent Framework	
8924426	15 août 2011	Joining Tables In A Mapreduce Procedure	
9405808	28 février 2012	System And Method For Analyzing Data Records	

Il est important de préciser que l'ensemble des 7 brevets associés à la technologie MapReduce sont des brevets « pledge » et donc en accès libre.

Nous avons ensuite procédé par une succession d'itérations pour collecter les follow-on innovations sur trois niveaux de citation. Ce choix de se concentrer sur les trois niveaux de citation provient des analyses effectuées dans le chapitre 1 où nous avons constaté que plus de 80% des follow-on innovations sur le brevet MapReduce initial étaient concentrés sur ces trois niveaux. Au total, le groupe traité initial est composé 2 654 brevets répartis de la manière suivante :

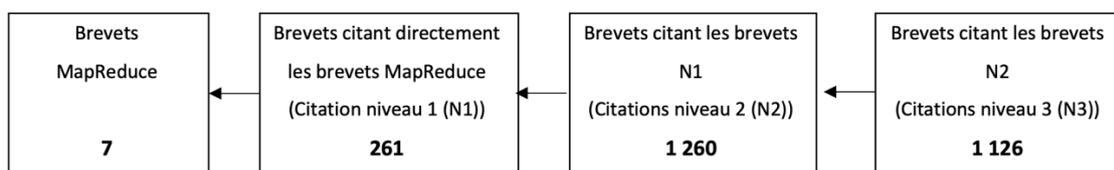


Figure 2.7 Composition du groupe traité

Pour constituer le groupe traité, nous avons fait correspondre ces brevets avec les brevets du périmètre 3 du chapitre I en prenant le soin de retirer les brevets communs avec le périmètre 1. La correspondance des brevets de contrôle s'est faite sur deux critères :

²¹ Voir source : [Pledged Patents – Open Patent Non-Assertion Pledge - Google](https://www.google.com/patents/opnpledge/patents/).

- Critère n°1 : le brevet de contrôle doit avoir la même année de priorité que le brevet traité
- Critère n°2 : le brevet de contrôle doit avoir strictement la même combinaison de classes technologiques (CPC) que le brevet traité

Sur les 2 654 brevets initiaux du groupe traité, 1 207 brevets ont au moins un brevet de contrôle ayant les critères indiqués précédemment. Le tableau 2.2 détaille les correspondances pour chaque niveau de citations des brevets traités.

Tableau 2.2 Nombre de brevets traités ayant au moins un brevet de contrôle

	Nombre de brevets traités initiaux	Nombre de brevets traités ayant au moins un brevet de contrôle	Pourcentage de complétude
Brevets MapReduce	7	7	100.0%
Brevets citant directement les brevets MapReduce (citation niveau 1)	261	172	65.9%
Brevets citant les brevets N1 (citations niveau 2)	1260	574	45.6%
Brevets citant les brevets N2 (citations niveau 3)	1126	454	40.3%
Brevets MapReduce et follow-on innovations	2 654	1 207	45.5%

Notre groupe traité final est composé de 1207 brevets sur la technologie MapReduce et ses follow-on innovations. Le groupe contrôle est composé 31 113 brevets ayant les mêmes années de dépôt et les mêmes spécifications technologiques que les brevets du groupe traité mais ils n'ont aucune filiation avec eux.

Les deux groupes et leur principales caractéristiques basées sur les critères de correspondance retenus précédemment sont décrits dans la Figure 2.8 qui présente la répartition des brevets des deux groupes par année de dépôt. L'axe des ordonnées de gauche représente le nombre de dépôts de brevets par an pour le groupe traité et celui de droite le nombre de dépôts pour le groupe contrôle. On constate une croissance des dépôts de brevets sur les technologies MapReduce et follow-on innovations à partir de 2010. Cette croissance peut s'expliquer par le fait que le brevet initial sur la technologie MapReduce a été accordé officiellement par l'office USPTO en 2010.

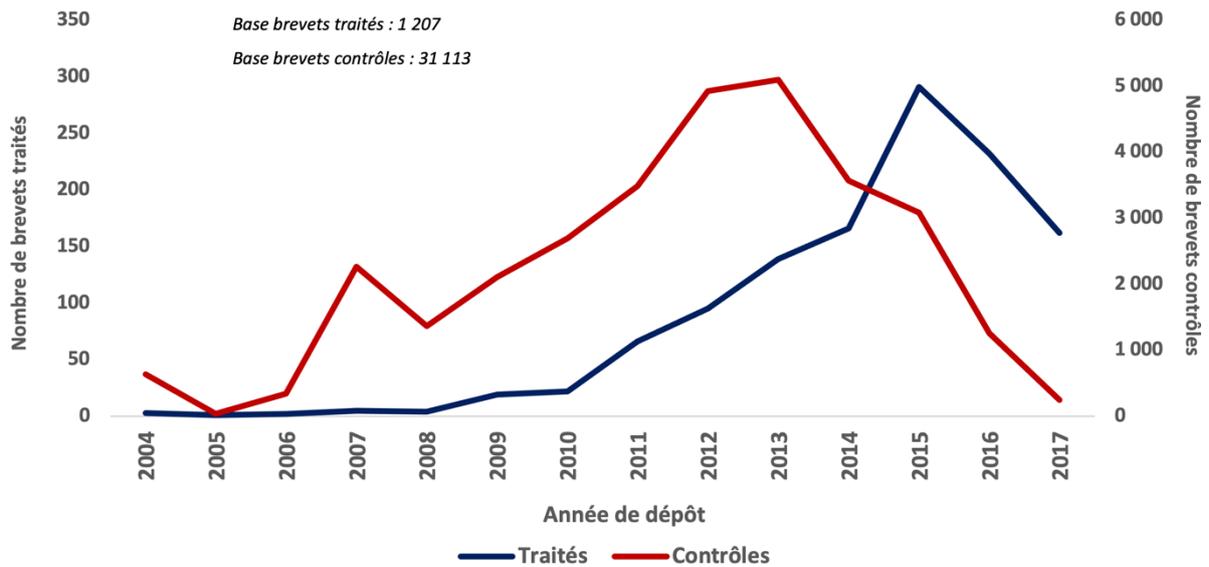


Figure 2.8 Brevets traités & contrôles - nombre de dépôts de brevets par an

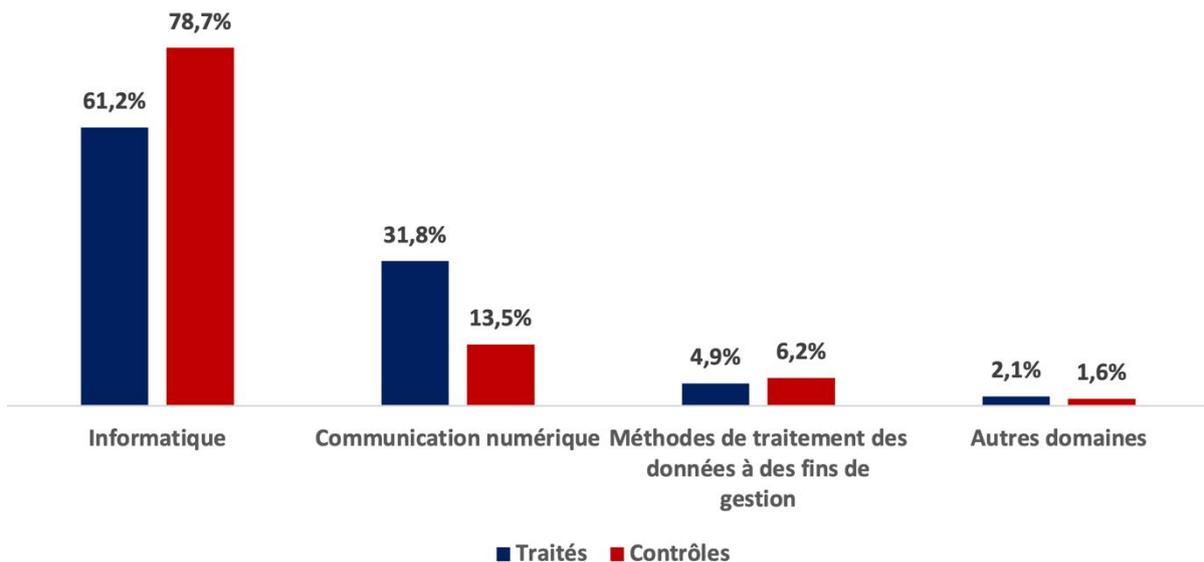


Figure 2.9 Brevets traités & contrôles - champs technologiques (WIPO)

La Figure 2.9 détaille les champs technologiques pour les brevets des deux groupes. Plus de 90% des brevets des deux groupes sont concentrés sur les champs technologiques « informatique » et « communication numérique », une concentration logique compte tenu de la nature de la technologie MapReduce.

Figure 2.10 Brevets traités et contrôlés

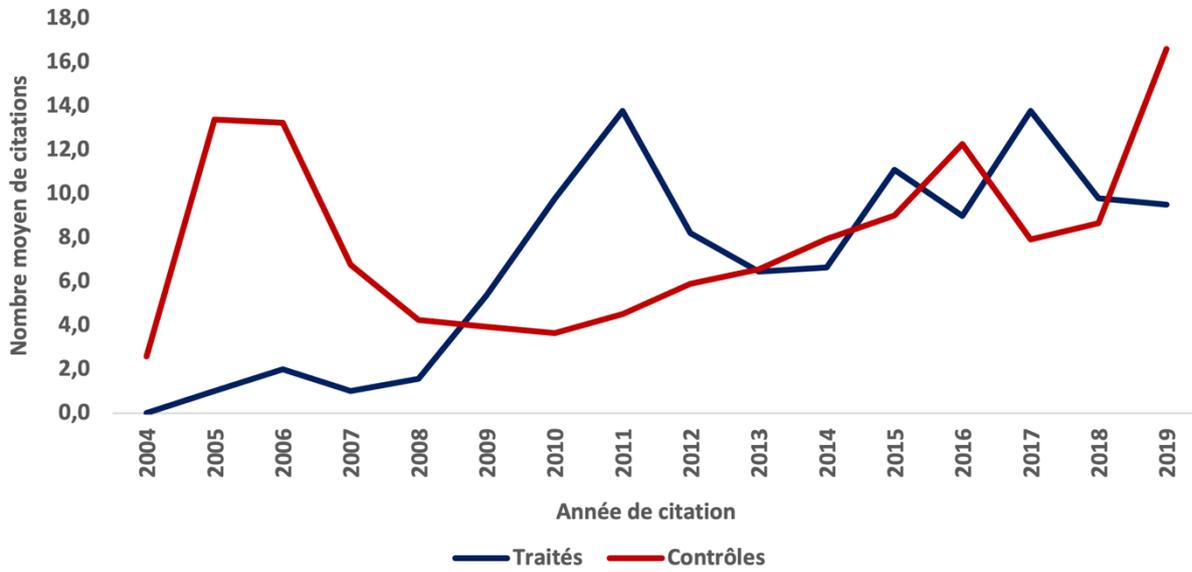


Figure 2.10a : brevets traités & contrôlés - nombre moyen de citations par an avec autocitations des organisations

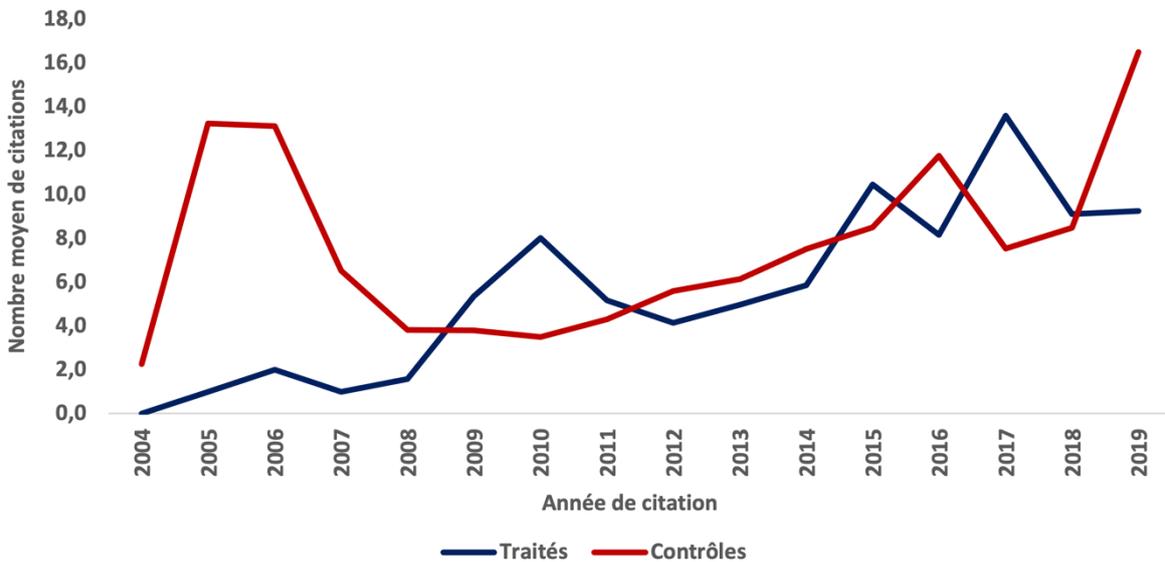


Figure 2.10b : brevets traités & contrôlés - nombre moyen de citations par an sans autocitations des organisations

Enfin les Figures 2.10 présentent le nombre moyen de citations reçues par an pour les brevets des deux groupes (avec et sans autocitations des organisations). Le nombre moyen de citations (avec ou sans autocitations) sur les technologies MapReduce et follow-on innovations (groupe traité) est en croissance constante sur la période étudiée. On constate deux périodes pour

lesquelles ces brevets sont particulièrement cités, les périodes 2010-2012 et 2015-2017. Ces deux périodes n'ont pas la même tendance si on s'intéresse plus particulièrement aux autocitations des organisations. En effet, sur la période 2010-2012, qui correspond à la période où le brevet initial MapReduce a été officiellement accordé par l'office USPTO mais où il n'est pas encore officiellement en accès libre direct par le dispositif « *Open Patent Non-Assertion Pledge* » de l'entreprise Google, le nombre moyen de citations sur les brevets MapReduce et follow-on innovations est de 10.6 en tenant compte des autocitations alors qu'il n'est que de 5.8 sans les autocitations. Il apparaît clairement que sur cette période, 2010-2012, les organisations titulaires des brevets MapReduce et follow-on innovations ont fortement utilisé ces technologies pour leur développement interne.

Si on s'intéresse maintenant à la période 2015-2017, période où le brevet initial MapReduce est en accès libre, on constate que le nombre moyen de citations en considérant les autocitations est de 11.3 et il est de 10.7 sans les autocitations. Sur cette période, les brevets MapReduce et follow-on innovations ont donc été davantage exploités par les organisations qui n'étaient pas à l'origine de ces technologies.

Pour les brevets ayant une technologie similaire mais n'ayant aucun lien avec la technologie MapReduce (groupe contrôle), le nombre moyen de citations était important sur le début de la période analysée (en moyenne 13.3 citations par brevet sur les années 2005 et 2006 avec et sans autocitations) mais ce nombre décroît rapidement à partir des années 2008 et ces technologies reçoivent significativement moins de citations sur la période 2010-2012 que les brevets MapReduce et follow-on innovations (4.5 citations en moyenne contre 8.2 avec et sans autocitations). A partir de l'année 2013, le nombre moyen de citations sur ces technologies retrouve une tendance croissante, comparable à celle des brevets MapReduce et follow-on innovations.

Nous allons maintenant détailler les résultats des estimations économétriques sur ces deux groupes de travail.

2.5 Estimations agrégées : citations reçues par brevet

Cette section va détailler deux types d'analyse empirique, d'une part, une analyse qui se situe au niveau du brevet cité et qui agrège les citations reçues et d'autre part, une analyse désagrégée avec les citations reçues par brevet et par année pour les échantillons des traités et des contrôles.

2.5.1 Variables et descriptions

L'analyse se situe au niveau du brevet cité et agrège les citations reçues. Comme les brevets les plus anciens ont le temps d'accumuler plus de citations, et selon les préceptes de la littérature, les citations sont comptées au cours des 5 années suivant la publication.

Variables dépendantes

- **Nombre de citations de brevets reçues (5 ans)** est le nombre de citations reçues par le brevet dans une fenêtre de 5 années après la publication.
- **Nombre de citations de brevets reçues sans autocitations (5 ans)** est le nombre de citations reçues par le brevet dans une fenêtre de 5 années après la publication sans les **autocitations définies au niveau des entreprises déposants**. Les autocitations ont été définies sur la base des déposants du brevet cité et citant pour tenir compte de l'impact des stratégies de dépôts de brevets par les autres acteurs du marché suivant la stratégie de Vakili, (2016).

Variables d'intérêt

- **Brevets Pledge MapReduce et follow-on innovation** représente le groupe traité et inclut tous les brevets MapReduce qui sont des brevets pledge ainsi que les follow-on innovations, c'est-à-dire toutes les citations qui se sont appuyées sur les brevets MapReduce à trois niveaux de citation
 - Niveau de citation 1 : les brevets citant directement les brevets Pledge MapReduce
 - Niveau de citation 2 : les brevets citant les brevets de niveau 1
 - Niveau de citation 3 : les brevets citant les brevets de niveau 2
- **Sponsors Apache** sont les organisations qui font partie de la fondation Apache. Ceux-ci peuvent encore être divisés en sous-groupes selon leur dotation financière à la fondation : sponsors Apache Platine, Or, Argent et Bronze. La liste des sponsors est indiquée dans le tableau 2.3.

Tableau 2.3 liste des sponsors de la fondation Apache

Sponsors Platine	Sponsors Or	Sponsors Argent	Sponsors Bronze
Amazon Web Services	Aiven	Aetna	Bestecasinobonussen.nl
Yahoo!	Anonymous	Alibaba Cloud Computing	Cafe24
Vmware	Baidu	Capital One	Cerner
Apple	Bloomberg	Didi Chuxing	Crafter CMS
Facebook	Cloudera	Line	Curity
Google	Comcast	Red Hat, Inc.	Dime Alley
Pineapple Fund	Confluent	Replicated	Dollar Hand
Microsoft	IBM	Talend	Goread.io Followers
Huawei	Indeed	Target	GridGain
	OceanX		HotWax Systems
	Salesforce		LeoVegas Indian Online Casir
	Union Investment		Libriciel
	VISA		Miro-Kredit AG
	Workday		OpenLogic by Perforce
			Paf
			Praca.pl
			PureVPN
			RX-M
			RenaissanceRe
			Sentry
			Technology Innovation Institu
			The Blog Starter
			Twitter
			Writers Per Hour

Total 56 organisations (source : fondation Apache)

Variables de contrôle :

Proximité entre inventeurs des brevets cités et citant

- **Autocitation de l'inventeur** : il s'agit d'une variable qui indique le nombre d'inventeurs parmi les inventeurs qui citent un de ses propres brevets.
- **Distance géodésique inventeurs** == 1 (respectivement == 2 et == 3) indique si au moins un inventeur du brevet citant a co-inventé avec un inventeur du brevet cité au cours des 5 précédentes années, et s'ils ont un co-inventeur en commun (== 2) ou le co-inventeur d'un co-inventeur en commun. Il s'agit d'un indicateur de proximité de réseau qui indique à quelle distance (collaboration directe, indirecte de niveau 1 ou 2) se situent les inventeurs des brevets citant et cités au cours des 5 années précédant la citation.
- **Nombre de co-inventeurs (max)** : il s'agit du nombre de co-inventeurs directs au cours des 5 dernières années (*degree centrality*) parmi les inventeurs du brevet cité. Ici on choisit le *degree* le plus élevé parmi les inventeurs du brevet.

- **Même organisation** : indique le nombre d'inventeurs des brevets citant ayant travaillé pour la même organisation que le brevet cité au cours des 5 années précédant la citation
- **Même ville** : indique le nombre d'inventeurs du brevet citant ayant été localisé dans la même ville que les inventeurs du brevet cité au cours des 5 années précédant la citation
- **Même Etat** : indique le nombre d'inventeurs des brevets citant ayant été localisé dans le même Etat aux Etats-Unis que les inventeurs du brevet cité au cours des 5 années précédant la citation.

Caractéristiques des brevets cités

- **Co-brevets (Déposants)** représentent les brevets qui ont plusieurs déposants. Ils apparaissent plusieurs fois dans les observations
- **Nombre d'inventeurs** listés sur le document du brevet est une mesure de la valeur des brevets, il indique aussi le montant des ressources investies dans l'invention (Sapsalis et al., 2006)
- **Nombre de sous-groupes CPC** caractérisant le brevet est un proxy de l'étendue technologique du brevet est un indicateur de complexité technologique
- **Taille de la famille** est un proxy de la valeur de l'invention (Lanjouw et al., 1998 ;Harhoff et al., 2003)
- **Le nombre de citations antérieures** (*Backward citations*) (Lanjouw & Schankerman, 2001)
- **Le nombre de citations hors brevets** sont les références ajoutées à un brevet, elle est considérée comme un proxy de la proximité à la science (Narin et al., 1997)
- **Le nombre de revendications** aussi appelés *claims* dans la littérature (Lanjouw & Schankerman, 2001)

Le tableau 2.4 suivant indique les principales statistiques descriptives des différentes variables.

Tableau 2.4 Statistiques descriptives				
Variables	Moyenne	Ecart-type	Minimum	Maximum
Variables dépendantes				
Nombre de citations de brevets reçues (5 ans)	3.59	10.30	0.00	330.00
Nombre de citations de brevets reçues sans autocitations (5 ans)	2.81	7.87	0.00	207.00
Variables d'intérêt				
Brevets Pledge MapReduce et follow-on innovation (Groupe traité)	0.04	0.20	0.00	1.00
Brevets Pledge MapReduce uniquement	0.00	0.02	0.00	1.00
Sponsors Apache	0.38	0.48	0.00	1.00
Sponsors Apache Platine	0.18	0.38	0.00	1.00
Sponsors Apache Or	0.19	0.40	0.00	1.00
Sponsors Apache Argent	0.01	0.09	0.00	1.00
Sponsors Apache Bronze	0.00	0.02	0.00	1.00
Niveau de citation 1	0.01	0.08	0.00	1.00
Niveau de citation 2	0.02	0.14	0.00	1.00
Niveau de citation 3	0.02	0.12	0.00	1.00
Variables de contrôle				
Proximité entre inventeurs des brevets cités et citant				
Autocitation de l'inventeur	0.55	3.46	0.00	330.00
Distance géodésique inventeurs == 1 (0/1)	0.18	0.39	0.00	1.00
Distance géodésique inventeurs == 2 (0/1)	0.13	0.34	0.00	1.00
Distance géodésique inventeurs == 3 (0/1)	0.12	0.33	0.00	1.00
Nombre de co-inventeurs (max)	9.12	14.60	0.00	144.00
Même organisation	1.32	6.47	0.00	330.00
Même ville	1.18	5.88	0.00	330.00
Même État	0.33	0.47	0.00	1.00
Caractéristiques des brevets cités				
Co-brevets (Déposants)	0.09	0.28	0.00	1.00
Nombre d'inventeurs	3.63	2.87	1.00	66.00
Nombre de sous-groupes CPC	3.39	2.76	1.00	44.00
Taille de la famille	1.90	1.85	0.00	26.00
Nombre de citations antérieures	30.89	76.92	0.00	4272.00
Nombre de citations hors brevets	9.09	23.28	0.00	1038.00
Nombre de revendications	18.61	8.17	0.00	229.00
Observations	44 106			

En moyenne les brevets (traités et contrôles) reçoivent 3.59 citations contre 2.81 sans autocitations des organisations. Si on s'intéresse aux variables d'intérêt, on constate que ce sont les sponsors platine et or qui sont les plus représentés dans les deux groupes. En termes de caractéristiques des brevets, on peut voir que ces brevets s'appuient sur un nombre de brevets conséquent étant donné que le nombre moyen de citations antérieures est de 30.89.

Le tableau 2.5 donne des tests d'égalité de moyennes entre le groupe des brevets traités et le groupe des brevets de contrôle. Les résultats du test indiquent que les brevets du groupe traité, brevets pledge MapReduce et follow-on innovations reçoivent plus de citations et que ces

brevets appartiennent proportionnellement plus à des entreprises du groupe Apache et notamment les plus influents qui appartiennent aux sponsors platine.

Tableau 2.5. Test de comparaison des moyennes entre le groupe traité et les contrôles.

	Traités		Contrôles	
	Moyenne	Moyenne	Différence	Ecart-type
Variable dépendantes				
Nombre de citations de brevets reçues (5 ans)	4.135	3.562	- 0.573**	0.248
Nombre de citations de brevets reçues sans autocitations (5 ans)	3.386	2.784	- 0.602***	0.189
Variabes d'intérêt				
Brevets Pledge MapReduce et follow-on innovation (Groupe traité)	1.000	0.000	- 1.000	0.000
Brevets Pledge MapReduce uniquement	0.007	0.000	- 0.007***	0.000
Sponsors Apache	0.430	0.374	- 0.056***	0.012
Sponsors Apache Platine	0.312	0.170	- 0.143***	0.009
Sponsors Apache Or	0.105	0.198	0.092***	0.009
Sponsors Apache Argent	0.011	0.007	- 0.004*	0.002
Sponsors Apache Bronze	0.001	0.000	- 0.001*	0.000
Niveau de citation 1	0.150	0.000	- 0.150***	0.002
Niveau de citation 2	0.475	0.000	- 0.475***	0.002
Niveau de citation 3	0.369	0.000	- 0.369***	0.002
Variabes de contrôle				
Proximité entre inventeurs des brevets cités et citants				
Autocitation de l'inventeur	0.897	0.533	- 0.364***	0.083
Distance géodésique inventeurs == 1 (0/1)	0.164	0.183	0.019**	0.009
Distance géodésique inventeurs == 2 (0/1)	0.126	0.134	0.008	0.008
Distance géodésique inventeurs == 3 (0/1)	0.115	0.125	0.009	0.008
Nombre de co-inventeurs (max)	5.869	9.257	3.389***	0.350
Même organisation	1.522	1.306	- 0.216	0.156
Même ville	1.468	1.170	- 0.298**	0.141
Même État	0.218	0.338	0.121***	0.011
Caractéristiques des brevets cités				
Co-brevets (Déposants)	0.131	0.087	- 0.044***	0.007
Nombre d'inventeurs	3.781	3.628	- 0.153**	0.069
Nombre de sous-groupes CPC	5.110	3.321	- 1.789***	0.066
Taille de la famille	2.179	1.892	- 0.287***	0.044
Nombre de citations antérieures	107.286	27.630	- 79.656***	1.810
Nombre de citations hors brevets	23.190	8.489	- 14.701***	0.555
Nombre de revendications	19.176	18.587	- 0.590***	0.196
Observations	1 805	42 301		

Note : la routine ttest de stata est ici appliquée et utilise une statistique de Student pour les tests.

Les brevets du groupe traité ont également des indicateurs de valeur technologique plus élevés que les brevets de contrôle sur l'ensemble des indicateurs qu'il s'agisse de la taille des familles, des citations antérieures et hors brevets, des revendications. Ils ont également en moyenne plus d'inventeurs et s'appuient sur des technologies plus complexes ayant plus de sous-groupes.

La prochaine section expose les différents résultats de l'estimation au niveau du brevet.

2.5.2 Résultats des estimations agrégées

Les tableaux suivants présentent les résultats de l'estimation agrégée au niveau du brevet sans le détail par année du niveau de citations des brevets follow-on innovation de MapReduce et distinction dans le type de sponsors Apache.

Le tableau 2.6 indique les résultats en tenant compte des autocitations des organisations, le tableau 2.7 les excluent. 6 estimations ont été réalisées sur le nombre de citations avec ou sans autocitations des organisations. Les deux premières estimations portent sur l'ensemble des variables d'intérêt et de contrôle et dans les 4 dernières, nous introduisons une interaction entre les brevets MapReduce et follow-on innovation et les sponsors Apache (estimations 3 et 4) et les co-brevets (estimations 5 et 6). Dans les deux types d'estimation des effets fixes « année » et « technologies » ont systématiquement été introduits afin de capter toutes les différences inobservables sur l'hétérogénéité de ces deux dimensions. Un troisième effet fixe « entreprises » est ajouté sur les estimations 2, 4 et 6.

	Nombre de citations					
	(1)	(2)	(3)	(4)	(5)	(6)
Brevets Pledge MapReduce et follow-on innovation (Groupe traité)	-0.17*** [0.04]	-0.14*** [0.05]	-0.29*** [0.05]	-0.26*** [0.05]	-0.17*** [0.04]	-0.14*** [0.04]
Brevets Pledge MapReduce uniquement	0.72*** [0.24]	0.29 [0.21]	0.56** [0.24]	0.19 [0.22]	0.72*** [0.23]	0.29 [0.21]
Sponsors Apache	-0.07*** [0.01]		-0.08*** [0.01]		-0.07*** [0.01]	
Co-brevets	-0.12*** [0.03]	-0.11*** [0.03]	-0.12*** [0.03]	-0.11*** [0.03]	-0.12*** [0.03]	-0.11*** [0.03]
Brevets Pledge MapReduce et follow-on innovation (groupe traite) x Sponsors Apache			0.30*** [0.08]	0.23** [0.09]		
Brevets Pledge MapReduce et follow-on innovation (groupe traite) x Co-brevets					0.02 [0.17]	0.01 [0.22]
Nombre d'inventeurs	0.07*** [0.02]	0.07*** [0.02]	0.07*** [0.02]	0.07*** [0.01]	0.07*** [0.02]	0.07*** [0.02]
Nombre de sous-groupes CPC	0.08*** [0.01]	0.05*** [0.01]	0.08*** [0.01]	0.05*** [0.01]	0.08*** [0.01]	0.05*** [0.01]
Taille de la famille	-0.16*** [0.02]	-0.00 [0.02]	-0.16*** [0.02]	-0.00 [0.02]	-0.17*** [0.02]	-0.00 [0.02]
Nombre de citations antérieurs	0.08*** [0.01]	0.03*** [0.01]	0.08*** [0.01]	0.03*** [0.01]	0.08*** [0.01]	0.03*** [0.01]
Nombre de citation hors brevets	0.09*** [0.01]	0.08*** [0.01]	0.09*** [0.01]	0.08*** [0.01]	0.09*** [0.01]	0.08*** [0.01]
Nombre de revendications	0.26*** [0.01]	0.09*** [0.01]	0.26*** [0.01]	0.09*** [0.01]	0.26*** [0.01]	0.09*** [0.01]
Constant	-0.24*** [0.06]	0.28*** [0.05]	-0.23*** [0.06]	0.28*** [0.05]	-0.24*** [0.06]	0.28*** [0.05]
EF Année	Oui	Oui	Oui	Oui	Oui	Oui
EF Technologie	Oui	Oui	Oui	Oui	Oui	Oui
EF Entreprise	Non	Oui	Non	Oui	Non	Oui
Observations	44106	42741	44106	42741	44106	42741
R-Squared	.099	.31	.1	.31	.099	.31

+ 0.10 ** 0.05 *** 0.01 Cluster brevets (Les écart-types des coefficients tiennent compte des corrélations intra-groupes définis au niveau des brevets pour tenir compte du fait que certaines lignes sont dupliquées en cas de co-brevet.)

Tableau 2.7: impact de la technologie Pledge MapReduce et Follow-on sur le nombre de citations reçues

	Nombre de citations sans autocitations					
	(1)	(2)	(3)	(4)	(5)	(6)
Brevets Pledge MapReduce et follow-on innovation (Groupe traité)	-0.13*** [0.04]	-0.11** [0.05]	-0.24*** [0.04]	-0.24*** [0.05]	-0.13*** [0.04]	-0.12*** [0.04]
Brevets Pledge MapReduce uniquement	0.74*** [0.23]	0.35+ [0.21]	0.59** [0.24]	0.25 [0.22]	0.74*** [0.23]	0.35+ [0.21]
Sponsors Apache	-0.08*** [0.01]		-0.09*** [0.01]		-0.08*** [0.01]	
Co-brevets	-0.06** [0.03]	-0.07** [0.03]	-0.06** [0.03]	-0.07** [0.03]	-0.06** [0.03]	-0.07*** [0.03]
Brevets Pledge MapReduce et follow-on innovation (Groupe traité) x Sponsors Apache			0.28*** [0.08]	0.24*** [0.09]		
Brevets Pledge MapReduce et follow-on innovation (Groupe traité) x Co-brevets					0.03 [0.17]	0.02 [0.22]
Nombre d'inventeurs	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]
Nombre de sous-groupes CPC	0.05*** [0.01]	0.03** [0.01]	0.05*** [0.01]	0.03** [0.01]	0.05*** [0.01]	0.03** [0.01]
Taille de la famille	-0.15*** [0.02]	-0.01 [0.02]	-0.15*** [0.02]	-0.01 [0.02]	-0.15*** [0.02]	-0.01 [0.02]
Nombre de citations antérieurs	0.06*** [0.01]	0.02*** [0.01]	0.06*** [0.01]	0.02*** [0.01]	0.06*** [0.01]	0.02*** [0.01]
Nombre de citation hors brevets	0.07*** [0.01]	0.07*** [0.01]	0.07*** [0.01]	0.07*** [0.01]	0.07*** [0.01]	0.07*** [0.01]
Nombre de revendications	0.22*** [0.01]	0.08*** [0.01]	0.22*** [0.01]	0.08*** [0.01]	0.22*** [0.01]	0.08*** [0.01]
Constant	-0.08 [0.05]	0.33*** [0.05]	-0.07 [0.05]	0.32*** [0.05]	-0.08 [0.05]	0.33*** [0.05]
EF Année	Oui	Oui	Oui	Oui	Oui	Oui
EF Technologie	Oui	Oui	Oui	Oui	Oui	Oui
EF Entreprise	Non	Oui	Non	Oui	Non	Oui
Observations	44106	42741	44106	42741	44106	42741
R-Squared	.089	.3	.09	.3	.089	.3

+ 0.10 ** 0.05 *** 0.01 Cluster brevets

Dans un premier temps nous allons étudier les résultats en considérant les autocitations des organisations et donc nous intéresser à la valeur privée des brevets de notre échantillon. Tout d'abord, on constate que les brevets pledge MapReduce et follow-on innovations (le groupe traité) reçoivent significativement moins de citations que les brevets ayant des technologies similaires sur l'ensemble des 6 estimations effectués. Il apparait donc que ces brevets, dans leur ensemble, n'ont pas participé à accroître significativement la valeur privée des entreprises à l'initiative de ces technologies. Si on s'intéresse uniquement aux 7 brevets pledge MapReduce on note qu'ils reçoivent significativement plus de citations que les brevets ayant des technologies similaires dans les estimations ayant des effets fixes sur les années de dépôt du brevet et sur les technologies, cependant ce résultat n'est plus significatif si on contrôle l'hétérogénéité entre les entreprises par un effet fixe. Ce résultat n'est pas surprenant en soi car en incluant un effet fixe sur les entreprises, le pouvoir explicatif provient uniquement de la variation au sein d'une entreprise, or ces 7 brevets n'appartiennent qu'à une seule entreprise ; Google. Ce dernier résultat ne nous permet donc pas de conclure en l'état sur l'impact des brevets pledge MapReduce.

Ensuite, il est intéressant d'examiner la variable « Apache Sponsors » permet d'évaluer le caractère innovant des brevets des sponsors Apache (la fondation communautaire regroupant de nombreuses technologies Big Data dont MapReduce et Hadoop) sur l'ensemble de l'échantillon. Il apparaît que sur l'ensemble des estimations ces brevets reçoivent moins de citations. Cependant dans les estimations 3 et 4, une interaction entre les sponsors Apache et les brevets MapReduce et follow-on innovation est introduite, le coefficient associé à cette interaction est positif et significatif sur les deux estimations réalisées (incluant notamment un effet fixe sur les entreprises). Il apparaît donc que les brevets MapReduce et follow-on innovation appartenant aux sponsors Apache reçoivent significativement plus de citations. Les estimations étant sur les citations avec autocitations, ces brevets auraient donc permis d'augmenter la valeur privée des sponsors Apache possédant ces technologies. Il est intéressant de noter que dans les estimations 5 et 6, une interaction entre les brevets MapReduce et follow-on innovation et les co-brevets est introduite afin d'avoir une référence supplémentaire sur les aspects collaboratifs (un co-brevets indiquant une collaboration entre plusieurs organisations) à l'intérieur de notre échantillon. Le coefficient associé à cette interaction est positif mais n'est pas statistiquement significatif dans les deux estimations effectuées.

L'étude du tableau 2.9 montre que les premiers résultats détaillés précédemment ne changent pas grandement si on considère les citations sans autocitations des organisations. En effet, les brevets pledge MapReduce et follow-on innovations reçoivent significativement toujours moins de citations, de même que les brevets des sponsors Apache. Cependant, les brevets MapReduce et follow-on innovation appartenant aux sponsors Apache reçoivent significativement plus de citations que les brevets ayant des technologies similaires. A l'instar de la valeur privée, les brevets MapReduce et follow-on innovations auraient donc également permis d'augmenter la valeur sociale des brevets des sponsors Apache. La seule différence notable entre les estimations avec et sans autocitations provient de l'impact des brevets Pledge MapReduce. En effet sur les estimations avec autocitations, il n'était pas possible de conclure sur l'impact de ces brevets, le coefficient associé à cette variable n'étant plus significatif avec l'introduction d'un effet fixe « entreprise ». Sur les estimations sans autocitations, on constate que ce coefficient devient faiblement significatif (10%) en présence d'effets fixes sur les entreprises (estimations 2 et 6, dans l'estimation 4 le pouvoir explicatif étant capté par l'interaction entre les brevets MapReduce/follow-on innovations et les sponsors Apache).

Tableau 2.8: Rôle de la Fondation Apache et type de follow-on innovation sur le nombre de citations

	Nombre de citations								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Brevets Pledge MapReduce uniquement	0.56**	0.17	0.40+	0.41+	0.19	0.41+	0.18	0.40+	0.17
	[0.23]	[0.21]	[0.23]	[0.23]	[0.21]	[0.23]	[0.21]	[0.23]	[0.21]
Sponsors Apache	-0.07***			-0.06					
	[0.01]			[0.17]					
Niveau de citation 1 (Groupe traité)	0.23**	0.17+	0.17	-0.25+	-0.25**	0.17+	0.18+	0.17	0.17+
	[0.10]	[0.10]	[0.10]	[0.13]	[0.11]	[0.10]	[0.10]	[0.10]	[0.10]
Niveau de citation 2 (Groupe traité)	-0.11+	-0.07	-0.14**	-0.13**	-0.07	-0.20***	-0.17+	-0.14**	-0.07
	[0.06]	[0.08]	[0.06]	[0.06]	[0.08]	[0.07]	[0.09]	[0.06]	[0.08]
Niveau de citation 3 (Groupe traité)	-0.43***	-0.40***	-0.47***	-0.47***	-0.40***	-0.47***	-0.40***	-0.44***	-0.36***
	[0.05]	[0.06]	[0.05]	[0.05]	[0.06]	[0.05]	[0.06]	[0.06]	[0.07]
Sponsors Apache Platine			0.09***	0.15		0.14		0.15	
			[0.02]	[0.17]		[0.17]		[0.17]	
Sponsors Apache Or			-0.22***	-0.16		-0.16		-0.16	
			[0.01]	[0.17]		[0.17]		[0.17]	
Sponsors Apache Argent			-0.39***	-0.33+		-0.31+		-0.33+	
			[0.05]	[0.17]		[0.17]		[0.17]	
Sponsors Apache Bronze			0.97**	1.03**		1.13**		1.11**	
			[0.47]	[0.50]		[0.50]		[0.51]	
Niveau de citation 1 (Groupe traité) x Sponsors Apache Platine				0.79***	0.74***				
				[0.20]	[0.19]				
Niveau de citation 1 (Groupe traité) x Sponsors Apache Or				0.34	0.33				
				[0.26]	[0.25]				
Niveau de citation 1 (Groupe traité) x Sponsors Apache Argent				0.74+	0.46***				
				[0.38]	[0.14]				
Niveau de citation 2 (Groupe traité) x Sponsors Apache Platine						0.28	0.30		
						[0.18]	[0.20]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Or						-0.03	0.01		
						[0.09]	[0.10]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Argent						-0.45***	-0.36**		
						[0.14]	[0.17]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Bronze						-1.59***	-1.56***		
						[0.48]	[0.47]		
Niveau de citation 3 (Groupe traité) x Sponsors Apache Platine								-0.15	-0.18
								[0.13]	[0.13]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Or								0.11	0.17+
								[0.09]	[0.10]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Argent								0.16**	0.33***
								[0.08]	[0.12]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Bronze								-1.33***	-1.32***
								[0.49]	[0.48]
Co-brevets	-0.12***	-0.11***	-0.15***	-0.15***	-0.11***	-0.15***	-0.11***	-0.15***	-0.11***
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
Nombre d'inventeurs	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***
	[0.02]	[0.01]	[0.02]	[0.02]	[0.01]	[0.02]	[0.01]	[0.02]	[0.01]
Nombre de sous-groupe CPC	0.08***	0.05***	0.08***	0.08***	0.05***	0.08***	0.05***	0.08***	0.05***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Taille de la famille	-0.17***	-0.00	-0.19***	-0.18***	-0.00	-0.19***	-0.00	-0.19***	-0.00
	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]
Nombre de citations antérieurs	0.08***	0.03***	0.08***	0.08***	0.03***	0.08***	0.03***	0.08***	0.03***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de citation hors brevets	0.09***	0.08***	0.09***	0.08***	0.07***	0.09***	0.08***	0.09***	0.08***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de revendications	0.26***	0.09***	0.21***	0.21***	0.09***	0.21***	0.09***	0.21***	0.09***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Constant	-0.24***	0.28***	-0.08	-0.08	0.28***	-0.08	0.28***	-0.08	0.28***
	[0.06]	[0.05]	[0.06]	[0.06]	[0.05]	[0.06]	[0.05]	[0.06]	[0.05]
EF Année	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
EF Technologie	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
EF Entreprise	Non	Oui	Non	Oui	Non	Non	Oui	Non	Oui
Observations	44106	42741	44106	44106	42741	44106	42741	44106	42741
R-Squared	.1	.31	.11	.11	.32	.11	.32	.11	.31

Tableau 2.9 : Rôle de la Fondation Apache et type de follow-on innovation sur le nombre de citations sans autocitations

	Nombre de citations sans autocitations								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Brevets Pledge MapReduce uniquement	0.62*** [0.23]	0.25 [0.21]	0.48** [0.23]	0.48** [0.23]	0.27 [0.21]	0.48** [0.23]	0.26 [0.21]	0.47** [0.23]	0.25 [0.21]
Sponsors Apache	-0.08*** [0.01]			-0.11 [0.16]		-0.10 [0.16]		-0.10 [0.16]	
Niveau de citation 1 (Groupe traité)	0.21** [0.10]	0.14 [0.10]	0.16+ [0.10]	-0.18 [0.13]	-0.19+ [0.10]	0.16+ [0.10]	0.14 [0.10]	0.16+ [0.10]	0.14 [0.10]
Niveau de citation 2 (Groupe traité)	-0.06 [0.06]	-0.04 [0.08]	-0.09 [0.06]	-0.08 [0.06]	-0.04 [0.08]	-0.16** [0.06]	-0.17+ [0.08]	-0.09 [0.06]	-0.04 [0.08]
Niveau de citation 3 (Groupe traité)	-0.37*** [0.05]	-0.34*** [0.05]	-0.40*** [0.05]	-0.40*** [0.05]	-0.34*** [0.05]	-0.40*** [0.05]	-0.34*** [0.05]	-0.38*** [0.05]	-0.32*** [0.06]
Sponsors Apache Platine			0.06*** [0.02]	0.16 [0.16]		0.15 [0.16]		0.17 [0.16]	
Sponsors Apache Or			-0.21*** [0.01]	-0.11 [0.16]		-0.11 [0.16]		-0.11 [0.16]	
Sponsors Apache Argent			-0.35*** [0.05]	-0.25 [0.17]		-0.23 [0.17]		-0.25 [0.17]	
Sponsors Apache Bronze			0.86** [0.42]	0.96** [0.45]		1.05** [0.45]		1.03** [0.46]	
Niveau de citation 1 (Groupe traité) x Sponsors Apache Platine				0.64*** [0.19]	0.57*** [0.18]				
Niveau de citation 1 (Groupe traité) x Sponsors Apache Or				0.31 [0.24]	0.31 [0.23]				
Niveau de citation 1 (Groupe traité) x Sponsors Apache Argent				0.64+ [0.36]	0.37*** [0.13]				
Niveau de citation 2 (Groupe traité) x Sponsors Apache Platine						0.29+ [0.17]	0.36+ [0.19]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Or						-0.04 [0.08]	0.02 [0.10]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Argent						-0.40*** [0.13]	-0.30+ [0.16]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Bronze						-1.41*** [0.43]	-1.40*** [0.41]		
Niveau de citation 3 (Groupe traité) x Sponsors Apache Platine								-0.11 [0.12]	-0.11 [0.12]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Or								0.11 [0.08]	0.15 [0.09]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Argent								0.15** [0.07]	0.28** [0.12]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Bronze								-1.17*** [0.44]	-1.22*** [0.42]
Co-brevets	-0.06** [0.03]	-0.07** [0.03]	-0.08*** [0.03]	-0.08*** [0.03]	-0.07** [0.03]	-0.08*** [0.03]	-0.07** [0.03]	-0.08*** [0.03]	-0.07** [0.03]
Nombre d'inventeurs	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]
Nombre de sous-groupes CPC	0.05*** [0.01]	0.03** [0.01]	0.05*** [0.01]	0.05*** [0.01]	0.03** [0.01]	0.05*** [0.01]	0.03** [0.01]	0.05*** [0.01]	0.03** [0.01]
Taille de la famille	-0.15*** [0.02]	-0.01 [0.02]	-0.17*** [0.02]	-0.17*** [0.02]	-0.01 [0.02]	-0.17*** [0.02]	-0.01 [0.02]	-0.17*** [0.02]	-0.01 [0.02]
Nombre de citations antérieurs	0.06*** [0.01]	0.02*** [0.01]	0.06*** [0.01]	0.06*** [0.01]	0.02*** [0.01]	0.06*** [0.01]	0.02*** [0.01]	0.06*** [0.01]	0.02*** [0.01]
Nombre de citation hors brevets	0.07*** [0.01]	0.06*** [0.01]	0.07*** [0.01]	0.07*** [0.01]	0.06*** [0.01]	0.07*** [0.01]	0.06*** [0.01]	0.07*** [0.01]	0.06*** [0.01]
Nombre de revendications	0.22*** [0.01]	0.08*** [0.01]	0.18*** [0.01]	0.18*** [0.01]	0.08*** [0.01]	0.18*** [0.01]	0.08*** [0.01]	0.18*** [0.01]	0.08*** [0.01]
Constant	-0.08 [0.05]	0.33*** [0.05]	0.06 [0.05]	0.07 [0.05]	0.33*** [0.05]	0.07 [0.05]	0.33*** [0.05]	0.06 [0.05]	0.33*** [0.05]
EF Année		Oui							
EF Technologie		Oui							
EF Entreprise		Non	Oui	Non	Non	Oui	Non	Oui	Non
Observations	44106	42741	44106	44106	42741	44106	42741	44106	42741
R-Squared	.091	.3	.1	.1	.3	.1	.3	.1	.3

A travers ce résultat, on pourrait donc penser que les brevets Pledge MapReduce auraient une valeur sociale plus importante que leur valeur privée, c'est-à-dire qu'ils auraient plus participé à l'essor du secteur des technologies Big Data, qu'à valoriser la société Google en elle-même. Ce résultat est à confirmer sur les prochaines estimations, notamment en termes de significativité statistiques.

Ces premiers résultats mettent plus particulièrement en avant les brevets Pledge MapReduce et follow-on innovations appartenant aux sponsors Apache. En effet, que ce soit avec ou sans autocitations des organisations, ces brevets reçoivent significativement plus de citations que les brevets ayant des technologies similaires. Il est donc logique de creuser ce résultat en détaillant deux points, d'une part les brevets follow-on innovations qui sont la résultante de trois niveaux de citations sur les brevets affiliés à la technologie MapReduce et d'autre part les sponsors Apache qui sont catégorisés en 4 types de sponsors : platine, or, argent et bronze (voir tableau 2.3). Les résultats des estimations économétriques avec la prise en compte avec le détail sur le type de follow-on innovation et sur les sponsors Apache sont présentés dans les tableaux suivants.

Le tableau 2.8 présente les résultats en tenant compte des autocitations des organisations. Les résultats sur les brevets pledge MapReduce sont comparables aux premières estimations, en effet les coefficients associés aux brevets déposés par la société Google ne sont pas significatifs dans l'ensemble des estimations qui incluent des effets fixes « entreprises ». Si on regarde maintenant le détail sur le type de follow-on innovations, on constate que dans la première estimation, qui ne tient pas compte du détail des sponsors Apache, ce sont les citations de niveau 1 qui reçoivent significativement le plus de citations, ce pouvoir explicatif diminue si on introduit des effets fixes « entreprises » (estimation 2). L'estimation 3 intègre les résultats sur le type de sponsors Apache, on constate que le niveau de citations 1 des follow-on innovation n'est plus significatif et les deux autres niveaux reçoivent significativement moins de citations. Concernant les sponsors Apache, ce sont les brevets des sponsors platines et bronzes qui reçoivent significativement le plus de citations alors qu'inversement ce sont les brevets des sponsors or et argent qui en reçoivent le moins. A l'image des premières estimations, des interactions entre le type de follow-on innovations et les sponsors Apache sont introduites à partir de l'estimation 4. On constate que ce sont les brevets follow-on innovations de niveau 1 (les brevets citant directement les brevets pledge MapReduce) appartenant aux sponsors Apache platines ou argent qui reçoivent significativement le plus de citations. On note

également que le coefficient associé aux brevets follow-on innovations de niveau 3 appartenant aux sponsors argent est également positif et significatif, le pouvoir explicatif est cependant légèrement inférieur par rapport au niveau 1 (5% pour le niveau 3 contre 1% pour le niveau 1). En résumé, ce sont donc les brevets les plus proches des technologies des brevets pledge MapReduce dont le déposant est un sponsor platine (Google, Apple, Amazon...) ou argent (Alibaba, Capital One, Line...) qui reçoivent le plus de citations.

En excluant les autocitations des organisations (Tableau 2.9), on constate que les résultats mentionnés précédemment ne changent pas, ce qui indiquerait que les brevets qui citent directement les brevets pledge MapReduce ont à la fois permis d'accroître la valeur privée des sponsors Apache platines et argents mais ils ont également contribué à l'essor du secteur des technologies Big Data. Enfin, il est important de noter que sur l'ensemble des estimations avec et sans autocitations, les coefficients associés aux brevets Pledge MapReduce sont positifs mais non significatifs, on ne peut donc pas statuer pour le moment sur l'apport de ces brevets et leur impact respectif sur la valeur privée (interne à l'entreprise Google) ou la valeur sociale (pour le secteur des technologies Big Data).

Afin de conforter les résultats de ces estimations, un troisième ensemble d'estimations a été réalisé afin d'intégrer des variables de contrôle supplémentaires sur la proximité entre les inventeurs des brevets cités et citant. En effet, les citations de brevets sont plus susceptibles de se produire parmi les inventeurs qui partagent une proximité sociale ou géographique (Breschi & Lissoni, 2009). Il semble donc utile de contrôler ces effets afin de s'assurer de la significativité de nos résultats.

Les résultats des estimations économétriques avec la prise en compte de la proximité entre les inventeurs des brevets cités et citant sont présentés dans les tableaux suivants.

Tableau 2.10 : Rôle de la Fondation Apache et type de follow-on innovation sur le nombre de citations

	Nombre de citations							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Groupe traité	-0.11*** [0.04]	-0.12*** [0.04]						
Brevets Pledge MapReduce uniquement	-0.35+ [0.21]	-0.58*** [0.21]	-0.40** [0.20]	-0.60*** [0.20]	-0.40** [0.20]	-0.60*** [0.20]	-0.41** [0.20]	-0.60*** [0.20]
Sponsors Apache	-0.13*** [0.01]		-0.25+ [0.14]				-0.25+ [0.14]	
Groupe traité x Sponsors Apache	0.16*** [0.06]	0.10 [0.07]						
Niveau de citation 1 (Groupe traité)			-0.14 [0.11]	-0.14 [0.10]	0.08 [0.07]	0.08 [0.07]	0.08 [0.07]	0.07 [0.07]
Niveau de citation 2 (Groupe traité)			-0.02 [0.05]	-0.02 [0.06]	-0.04 [0.06]	-0.05 [0.08]	-0.02 [0.05]	-0.02 [0.06]
Niveau de citation 3 (Groupe traité)			-0.21*** [0.03]	-0.22*** [0.04]	-0.21*** [0.03]	-0.22*** [0.04]	-0.21*** [0.04]	-0.20*** [0.05]
Sponsors Apache Platine			0.23 [0.14]		-0.01 [0.01]		0.24+ [0.14]	
Sponsors Apache Or			0.02 [0.14]		-0.22*** [0.01]		0.02 [0.14]	
Sponsors Apache Argent			0.03 [0.15]		-0.21*** [0.04]		0.03 [0.15]	
Sponsors Apache Bronze			0.97*** [0.34]		0.79** [0.31]		1.02*** [0.35]	
Niveau de citation 1 (Groupe traité) x Sponsors Apache Platine			0.43*** [0.15]	0.38*** [0.14]				
Niveau de citation 1 (Groupe traité) x Sponsors Apache Or			0.14 [0.18]	0.13 [0.18]				
Niveau de citation 1 (Groupe traité) x Sponsors Apache Argent			0.29 [0.18]	0.25** [0.12]				
Niveau de citation 2 (Groupe traité) x Sponsors Apache Platine					0.12 [0.14]	0.11 [0.16]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Or					-0.10 [0.07]	-0.07 [0.08]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Argent					-0.29** [0.13]	-0.24 [0.15]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Bronze					-0.90*** [0.31]	-1.04*** [0.32]		
Niveau de citation 3 (Groupe traité) x Sponsors Apache Platine							-0.05 [0.08]	-0.10 [0.08]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Or							0.15** [0.06]	0.17*** [0.07]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Argent							0.08 [0.08]	0.20 [0.14]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Bronze							-0.85*** [0.33]	-1.01*** [0.32]
Co-brevets	0.06** [0.02]	0.05+ [0.03]	0.04 [0.02]	0.05+ [0.03]	0.04 [0.02]	0.05+ [0.03]	0.04 [0.02]	0.05+ [0.03]
Autocitation de l'inventeur	0.00 [0.02]	0.01 [0.01]	0.00 [0.02]	0.01 [0.01]	0.00 [0.02]	0.01 [0.01]	0.00 [0.02]	0.01 [0.01]
Proximité géodésique inventeurs == 1 (0/1)	0.44*** [0.04]	0.37*** [0.03]	0.44*** [0.04]	0.37*** [0.03]	0.44*** [0.04]	0.37*** [0.03]	0.44*** [0.04]	0.37*** [0.03]
Proximité géodésique inventeurs == 2 (0/1)	0.10*** [0.03]	0.07*** [0.03]	0.11*** [0.03]	0.07*** [0.02]	0.11*** [0.03]	0.07*** [0.02]	0.11*** [0.03]	0.07*** [0.02]
Proximité géodésique inventeurs == 3 (0/1)	0.52*** [0.02]	0.50*** [0.02]	0.51*** [0.02]	0.50*** [0.02]	0.51*** [0.02]	0.50*** [0.02]	0.51*** [0.02]	0.50*** [0.02]
Nombre de co-inventeurs (max)	0.01*** [0.00]	0.01*** [0.00]	0.01*** [0.00]	0.01*** [0.00]	0.01*** [0.00]	0.01*** [0.00]	0.01*** [0.00]	0.01*** [0.00]
Même organisation	0.03*** [0.00]	0.03*** [0.00]	0.03*** [0.00]	0.03*** [0.00]	0.03*** [0.00]	0.03*** [0.00]	0.03*** [0.00]	0.03*** [0.00]
Même Etat	0.37*** [0.01]	0.31*** [0.01]	0.37*** [0.01]	0.31*** [0.01]	0.37*** [0.01]	0.31*** [0.01]	0.37*** [0.01]	0.31*** [0.01]
Nombre d'inventeurs	-0.16*** [0.01]	-0.14*** [0.01]	-0.15*** [0.01]	-0.14*** [0.01]	-0.15*** [0.01]	-0.14*** [0.01]	-0.15*** [0.01]	-0.14*** [0.01]
Nombre de sous-groupes CPC	-0.00 [0.01]	-0.01 [0.01]	0.00 [0.01]	-0.01 [0.01]	0.00 [0.01]	-0.01 [0.01]	0.00 [0.01]	-0.01 [0.01]
Taille de la famille	-0.11*** [0.01]	-0.02 [0.01]	-0.12*** [0.01]	-0.02 [0.01]	-0.12*** [0.01]	-0.02 [0.01]	-0.13*** [0.01]	-0.02 [0.01]
Nombre de citations antérieurs	0.04***	0.01**	0.04***	0.02**	0.04***	0.01**	0.04***	0.02**

	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de citation hors brevets	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de revendications	0.15***	0.05***	0.12***	0.05***	0.12***	0.05***	0.12***	0.05***	0.05***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Constant	0.23***	0.47***	0.34***	0.48***	0.34***	0.48***	0.34***	0.47***	0.47***
	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]
EF Année	Oui								
EF Technologie	Oui								
EF Entreprise	Non	Oui	Non	Oui	Non	Oui	Non	Oui	Oui
Observations	44106	42741	44106	42741	44106	42741	44106	42741	42741
R-Squared	.44	.56	.44	.56	.44	.56	.44	.56	.56

Groupe traité = Brevets Pledge MapReduce et follow-on innovation

Le tableau 2.10 détaille les résultats en considérant les autocitations des organisations. Tout d'abord, on constate que les brevets reçoivent plus de citations si les inventeurs du brevet cité et citant ont au moins une fois collaborer directement (le coefficient de la variable « proximité géodésique inventeurs == 1 » est systématiquement positif et significatif sur l'ensemble des estimations), l'effet est similaire si les inventeurs du brevet cité et citant ont été dans la même organisation ou le même état. La proximité sociale ou géographique est donc un facteur qui explique significativement le nombre de citations dans notre échantillon de brevets. Il est logique de se demander si ce résultat a eu un impact sur nos variables d'intérêt. Concernant les brevets pledge MapReduce/follow-on innovation et les brevets des sponsors Apache (sans distinction de type), on constate qu'ils reçoivent significativement moins de citations que les brevets ayant des technologies similaires, un résultat similaire aux précédents résultats. Cependant, si on s'intéresse uniquement aux brevets pledge MapReduce, on constate qu'ils reçoivent significativement moins de citations, en effet les coefficients associés à cette variable sont négatifs et significatifs sur l'ensemble des estimations (notamment sur les estimations incluant des effets fixes « entreprises ») alors qu'ils étaient positifs dans nos résultats précédents. Il semblerait donc que les brevets pledge MapReduce ont eu un impact limité sur les inventions du secteur des technologies Big Data. Si on s'intéresse au type de follow-on innovations, il apparaît que le niveau 1 et 2 n'est pas statistiquement significatif, quant aux brevets de niveau 3, ils reçoivent significativement moins de citations. Sur les sponsors Apache, on remarque que seuls les brevets des sponsors bronze, dans leur ensemble, reçoivent significativement plus de citations. Enfin concernant les interactions entre les brevets du groupe traité et les sponsors Apache, ce sont les brevets des sponsors platines qui citent directement les brevets pledge MapReduce qui reçoivent le plus de citations, ce résultat est également similaire à nos précédentes estimations. De plus, on constate également qu'avec l'introduction d'effets fixes sur les entreprises, le coefficient associé aux brevets des sponsors argent ayant un lien de citations directes avec les technologies MapReduce devient positif et significatif.

Tableau 2.11 : Rôle de la Fondation Apache et type de follow-on innovations sur le nombre de citations sans autocitations

	Nombre de citations sans autocitations							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Groupe traité	-0.10***	-0.12***						
	[0.04]	[0.04]						
Groupe traité x Sponsors Apache	0.19***	0.15**						
	[0.06]	[0.07]						
Sponsors Apache	-0.14***		-0.28+				-0.27+	
	[0.01]		[0.15]				[0.15]	
Brevets Pledge MapReduce uniquement	-0.19	-0.44**	-0.19	-0.40**	-0.19	-0.40**	-0.20	-0.41**
	[0.21]	[0.21]	[0.20]	[0.20]	[0.20]	[0.20]	[0.20]	[0.20]
Niveau de citation 1 (Groupe traité)			-0.09	-0.10	0.12	0.09	0.12	0.09
			[0.12]	[0.09]	[0.08]	[0.08]	[0.08]	[0.08]
Niveau de citation 2 (Groupe traité)			0.01	0.01	-0.03	-0.06	0.01	0.01
			[0.05]	[0.06]	[0.06]	[0.08]	[0.05]	[0.06]
Niveau de citation 3 (Groupe traité)			-0.19***	-0.19***	-0.19***	-0.19***	-0.20***	-0.19***
			[0.03]	[0.04]	[0.03]	[0.04]	[0.04]	[0.05]
Sponsors Apache Platine			0.23		-0.04***		0.24	
			[0.15]		[0.01]		[0.15]	
Sponsors Apache Or			0.05		-0.22***		0.05	
			[0.15]		[0.01]		[0.15]	
Sponsors Apache Argent			0.06		-0.20***		0.06	
			[0.16]		[0.04]		[0.16]	
Sponsors Apache Bronze			1.00***		0.80**		1.06***	
			[0.36]		[0.32]		[0.36]	
Niveau de citation 1 (Groupe traité) x Sponsors Apache Platine			0.39**	0.33**				
			[0.16]	[0.15]				
Niveau de citation 1 (Groupe traité) x Sponsors Apache Or			0.14	0.14				
			[0.17]	[0.16]				
Niveau de citation 1 (Groupe traité) x Sponsors Apache Argent			0.23	0.22**				
			[0.17]	[0.10]				
Niveau de citation 2 (Groupe traité) x Sponsors Apache Platine					0.20	0.24		
					[0.14]	[0.16]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Or					-0.09	-0.05		
					[0.07]	[0.08]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Argent					-0.27**	-0.19		
					[0.13]	[0.15]		
Niveau de citation 2 (Groupe traité) x Sponsors Apache Bronze					-0.93***	-1.07***		
					[0.33]	[0.33]		
Niveau de citation 3 (Groupe traité) x Sponsors Apache Platine							-0.01	-0.04
							[0.08]	[0.09]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Or							0.15**	0.15**
							[0.06]	[0.07]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Argent							0.09	0.20
							[0.07]	[0.13]
Niveau de citation 3 (Groupe traité) x Sponsors Apache Bronze							-0.87***	-1.05***
							[0.34]	[0.33]
Co-brevets	0.08***	0.06**	0.06**	0.06**	0.06**	0.06**	0.06**	0.06**
	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]
Autocitation de l'inventeur	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Proximité géodésique inventeurs == 1 (0/1)	0.24***	0.17***	0.23***	0.17***	0.23***	0.17***	0.23***	0.17***
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
Proximité géodésique inventeurs == 2 (0/1)	0.06**	0.04	0.07**	0.03	0.07**	0.04	0.07**	0.04
	[0.03]	[0.02]	[0.03]	[0.02]	[0.03]	[0.02]	[0.03]	[0.02]
Proximité géodésique inventeurs == 3 (0/1)	0.43***	0.41***	0.42***	0.41***	0.42***	0.41***	0.42***	0.41***
	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]
Nombre de co-inventeurs (max)	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Même organisation	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Même État	0.39***	0.33***	0.38***	0.33***	0.38***	0.33***	0.38***	0.32***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre d'inventeurs	-0.15***	-0.14***	-0.15***	-0.14***	-0.15***	-0.13***	-0.15***	-0.14***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de sous-groupes CPC	-0.01	-0.02	-0.01	-0.02	-0.01	-0.02	-0.01	-0.02
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]

Taille de la famille	-0.11***	-0.02	-0.12***	-0.02	-0.12***	-0.02	-0.12***	-0.02
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de citations antérieurs	0.03***	0.01+	0.03***	0.01**	0.03***	0.01+	0.03***	0.01**
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de citation hors brevets	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de revendications	0.13***	0.05***	0.11***	0.05***	0.11***	0.05***	0.11***	0.05***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Constant	0.28***	0.47***	0.37***	0.47***	0.37***	0.47***	0.37***	0.47***
	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]
EF Année	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
EF Technologie	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
EF Entreprise	Non	Oui	Non	Oui	Non	Oui	Non	Oui
Observations	44106	42741	44106	42741	44106	42741	44106	42741
R-Squared	.35	.49	.35	.49	.35	.49	.35	.49

Groupe traité = Brevets Pledge MapReduce et follow-on innovation

Au total, on constate donc que ce sont les brevets appartenant aux sponsors Apache platine qui citent directement les brevets Pledge MapReduce qui reçoivent significativement le plus de citations par rapport aux brevets ayant des technologies similaires. De plus, on constate également que les brevets des sponsors Apache bronze dans leur globalité sont les plus cités. Ces estimations ont également mis en avant que les brevets pledge MapReduce ne recevaient pas significativement plus de citations que les brevets ayant des technologies similaires. Un résultat qu'il est nécessaire d'approfondir compte tenu de la place centrale de ces technologies dans l'écosystème des technologies Big Data.

Dans la section suivante, de nouvelles estimations économétriques sont effectuées avec des données désagrégées qui tiennent compte de l'année de citation des brevets. La société Google ayant mis en place son programme « *Open Patent Non-Assertion Pledge* » en 2013, ces estimations vont permettre de distinguer l'année de déploiement de ce dispositif dans l'évaluation de l'impact des brevets pledge MapReduce.

2.6 Estimations désagrégées : citations reçues par brevet par année

Cette section continue à explorer l'impact des brevets pledge, des follow-on innovations et des contrôles sur les citations reçues mais cette fois en exploitant la dimension temporelle des données. On considère à présent le nombre de citations reçues chaque année (sans limite dans le temps) et les variables de proximité caractérisent la relation entre les inventeurs du brevet cité et les inventeurs des brevets citant pour une année donnée.

2.6.1 Variables et descriptions

La définition des variables reste semblable à celle de la section précédente. Toutefois une nouvelle variable est introduite, il s'agit de « Post-Pledge » qui indique si la citation intervient au cours des années suivant la déclaration en Pledge du brevet. Dans le cas des brevets Pledge MapReduce, il s'agit de l'année 2013. Ainsi la variable Post-Pledge prend la valeur 0 jusqu'en 2012 et la valeur 1 après 2012.

Le tableau 2.12 présente les principales statistiques descriptives des différentes variables pour les estimations désagrégées.

	Moyenne	Ecart-type	Minimum	Maximum
Tableau 2.12 : Statistiques descriptives				
Variable dépendantes				
Nombre de citations reçues (par année)	3.72	5.93	0.00	116.00
Nombre de citations reçues sans autocitations (par année)	3.34	5.35	0.00	116.00
Variables d'intérêt				
Brevets Pledge MapReduce et follow-on innovation (groupe traité)	0.03	0.17	0.00	1.00
Brevets Pledge MapReduce	0.00	0.04	0.00	1.00
Post-Pledge	0.79	0.41	0.00	1.00
Sponsors Apache	0.37	0.48	0.00	1.00
Sponsors Apache Platine	0.24	0.43	0.00	1.00
Sponsors Apache Or	0.13	0.33	0.00	1.00
Sponsors Apache Argent	0.01	0.07	0.00	1.00
Sponsors Apache Bronze	0.00	0.03	0.00	1.00
Niveau de citation 1	0.01	0.10	0.00	1.00
Niveau de citation 2	0.01	0.11	0.00	1.00
Niveau de citation 3	0.00	0.05	0.00	1.00
Variables de contrôle				
Co-brevets	0.08	0.27	0.00	1.00
Proximité entre inventeurs des brevets cités et citants				
Autocitation inventeur	0.03	0.17	0.00	1.00
Proximité géodésique inventeurs == 1 (0/1)	0.04	0.19	0.00	1.00
Proximité géodésique inventeurs == 2 (0/1)	0.04	0.18	0.00	1.00
Proximité géodésique inventeurs == 3 (0/1)	0.07	0.25	0.00	1.00
Nombre de co-inventeurs (Max)	14.39	15.35	0.00	144
Même organisation	0.14	0.34	0.00	1.00
Même ville	0.16	0.36	0.00	1.00
Même État	0.52	0.50	0.00	1.00
Caractéristiques des brevets cités				
Nombre d'inventeurs	3.76	3.14	1.00	44.00
Nombre de sous-groupes CPC	2.93	2.51	1.00	44.00
Taille de la famille	1.89	1.88	0.00	26.00
Nombre de citations antérieurs	32.05	66.61	0.00	4272.00
Nombre de citations hors brevets	11.46	24.48	0.00	1038.00
Nombre de revendications	20.58	10.14	0.00	229.00
Observations	56 994			

En moyenne, les brevets (traités ou contrôle) reçoivent 3.72 citations par an si on tient compte des autocitations des organisations. Ce chiffre baisse très légèrement à 3.34 en excluant les autocitations. A l'image de notre premier échantillon, ces brevets s'appuient sur un nombre conséquent de références, le nombre moyen de citations antérieurs et hors brevets est respectivement de 32.05 et 11.46.

Le tableau 2.13 détaille les différentes variables en distinguant les deux groupes « traité » et « contrôle ».

Tableau 2.13 : Test de comparaison des moyennes entre le groupe traité et les contrôles.

	Traités Contrôles		Différences	Ecart-types
	Moyennes			
Variable dépendantes				
Nombre de citations reçues	5.603	3.666	-1.938***	0.150
Nombre de citations reçues sans autocitations	4.600	3.305	-1.296***	0.135
Variables d'intérêt				
Post-Pledge	0.895	0.783	-0.110***	0.010
Brevets Pledge MapReduce et follow-on innovation (groupe traité)	1.000	0.000	-1.000	0.000
Brevets Pledge MapReduce	0.071	0.000	-0.071***	0.001
Sponsors Apache	0.571	0.362	-0.209***	0.012
Sponsors Apache Platine	0.522	0.230	-0.292***	0.011
Sponsors Apache Or	0.041	0.130	0.089***	0.008
Sponsors Apache Argent	0.009	0.005	-0.004*	0.002
Sponsors Apache Bronze	0.000	0.001	0.001	0.001
Niveau de citation 1	0.368	0.000	-0.368***	0.002
Niveau de citation 2	0.469	0.000	-0.469***	0.002
Niveau de citation 3	0.092	0.000	-0.092***	0.001
Variables de contrôle				
Co-brevets	0.137	0.080	-0.057***	0.007
Proximité entre inventeurs des brevets cités et citant				
Autocitation inventeur	0.034	0.030	-0.004	0.004
Proximité géodésique inventeurs == 1 (0/1)	0.034	0.037	0.003	0.005
Proximité géodésique inventeurs == 2 (0/1)	0.030	0.035	0.006	0.005
Proximité géodésique inventeurs == 3 (0/1)	0.076	0.065	-0.011*	0.006
Nombre de co-inventeurs (Max)	14.683	14.368	-0.298	0.389
Même organisation	0.149	0.137	-0.012	0.009
Même ville	0.202	0.157	-0.045***	0.009
Même État	0.564	0.517	-0.046***	0.013
Caractéristiques des brevets cités				
Nombre d'inventeurs	1.523	1.420	-0.103***	0.013
Nombre de sous-groupes CPC	1.504	1.213	-0.291***	0.013
Taille de la famille	0.953	0.938	-0.015	0.011
Nombre de citations antérieurs	3.540	2.920	-0.619***	0.026
Nombre de citation hors brevets	2.548	1.557	-0.991***	0.033
Nombre de revendications	3.064	2.976	-0.088***	0.012
Observations	1 604	55 390	56 994	

En moyenne, les brevets Pledge MapReduce et follow-on innovations (groupe traité) reçoivent deux citations de plus par an que les brevets ayant des technologies semblables (groupe

contrôle). En excluant les autocitations des organisations, cette différence entre les deux groupes n'est plus que de 1.3 citations. On constate également que les sponsors Apache sont en moyenne plus représentés dans le groupe traité. En termes de caractéristiques des brevets cités, les brevets pledge MapReduce et follow-on innovations s'appuient en moyenne sur un nombre de références antérieures (citations antérieures et citations hors brevets) plus importantes que les brevets ayant des technologies similaires.

La prochaine section expose les résultats de l'estimation désagrégées avec les citations reçues par brevet par année, nous nous intéresserons plus particulièrement à l'impact des brevets Pledge MapReduce et follow-on innovations.

2.6.2 Résultats des estimations désagrégées

A l'image de nos premières estimations réalisées précédemment, le tableau 9 présente des premiers résultats sans le détail sur le type de follow-on innovation. Les deux premières colonnes indiquent les résultats en tenant compte des autocitations des organisations, les deux dernières les excluent. Dans l'ensemble des estimations, des effets fixes « année de la citation » et « technologies » sont présents, dans les estimations 2 et 4, des effets fixes « entreprises citées » sont incluses.

Les résultats du tableau 2.14 nous indiquent que si on considère le nombre de citations reçues par année, les brevets MapReduce et follow-on innovation ne reçoivent pas plus de citations que les brevets ayant des technologies similaires avant l'introduction des Pledge, ce résultat est statistiquement significatif si on exclut les autocitations. Ce résultat est similaire à nos précédentes estimations sur les données agrégées par brevet. Cependant, il apparaît que les brevets Pledge MapReduce reçoivent significativement plus de citations avant l'introduction des Pledge et cela sur les deux périmètres d'estimations, à savoir avec et sans autocitations des organisations. En considérant le nombre de citations reçues par année, il apparaît donc que les 7 brevets spécifiques à la technologies MapReduce ont une valeur privée et sociale supérieures aux brevets ayant des technologies similaires avant le passage en Pledge. Comme mentionné dans l'introduction, la prise en compte de l'année de citation dans nos analyses permet d'introduire une variable « post-pledge » qui va tenir compte de la date effective de l'ouverture des brevets MapReduce, à savoir l'année 2013. Deux interactions sont effectuées avec cette

variable, d'une part avec les brevets pledge MapReduce et follow-on innovations et d'autre part, avec les brevets MapReduce uniquement.

Les résultats sur ces deux interactions sont opposés. En effet, si on considère les brevets pledge MapReduce et follow-on innovations sur la période ultérieure à 2013 (et l'ouverture des brevets MapReduce), ces brevets reçoivent sur cette période significativement plus de citations que les brevets ayant des technologies semblables. A l'inverse sur cette même période, les brevets MapReduce uniquement reçoivent significativement moins de citations.

Tableau 2.14 : Nombre de citations reçues par année

	Nombre de Citations reçues par année			
	Avec autocitations		Sans autocitations	
	(1)	(2)	(3)	(4)
Brevets Pledge MapReduce et follow-on innovation (groupe traité)	-0.08 [0.11]	-0.15 [0.11]	-0.19*** [0.07]	-0.29*** [0.08]
Brevets Pledge MapReduce	1.12*** [0.30]	1.03*** [0.29]	1.01*** [0.13]	0.94*** [0.13]
Post-Pledge x Brevets Pledge MapReduce et follow-on innovation (groupe traité)	0.22+ [0.12]	0.35*** [0.13]	0.30*** [0.10]	0.43*** [0.11]
Post-Pledge x Brevets Pledge MapReduce	-0.35** [0.15]	-0.48*** [0.16]	-0.47*** [0.14]	-0.60*** [0.16]
Sponsors Apache	0.01 [0.02]		-0.00 [0.02]	
Co-brevets	-0.13*** [0.05]	-0.07 [0.06]	-0.07 [0.05]	-0.02 [0.06]
Nombre d'inventeurs	0.03 [0.02]	0.03 [0.02]	0.02 [0.02]	0.02 [0.02]
Nombre de sous-groupes CPC	0.01 [0.02]	-0.03 [0.02]	-0.01 [0.02]	-0.04+ [0.02]
Taille de la famille	-0.03 [0.03]	0.02 [0.03]	-0.02 [0.03]	0.01 [0.03]
Nombre de citations antérieurs	0.03*** [0.01]	0.01 [0.01]	0.01 [0.01]	0.00 [0.01]
Nombre de citation hors brevets	0.03*** [0.01]	0.03*** [0.01]	0.02*** [0.01]	0.02*** [0.01]
Nombre de revendications	0.07*** [0.02]	0.01 [0.03]	0.06** [0.02]	0.00 [0.03]
Constant	0.84*** [0.08]	1.11*** [0.09]	0.94*** [0.07]	1.12*** [0.09]
EF Années de la citation	Oui	Oui	Oui	Oui
EF technologie	Oui	Oui	Oui	Oui
EF Entreprises citées	Non	Oui	Non	Oui
Observations	57063	56611	57063	56611
R-Squared	.038	.21	.026	.18
Cluster brevets cités				

A travers ces deux résultats, il apparaît que les brevets MapReduce ont eu un impact sur l'écosystème des technologies Big Data sur la période 2004-2012. A partir de 2013 et le déploiement du dispositif « *Open Patent Non-Assertion Pledge* » de l'entreprise Google, ce

sont les brevets follow-on innovations qui ont un lien de citations avec la technologie MapReduce qui ont capté la majorité des citations reçues. Il est important de rappeler que la technologie initiale MapReduce a fait l'objet d'un dépôt de brevet par l'entreprise Google en 2004 et a été accessible dès 2008 à travers la fondation Apache et le projet Hadoop. Le brevet MapReduce a été accordé par l'office USPTO en 2010, et donc pendant deux ans cette technologie a donc été accessible et utilisée dans le cadre du projet Hadoop, projet qui possède le plus grand nombre de contributeurs au sein de la fondation²². A l'obtention de son brevet, l'entreprise Google a immédiatement accordé une licence MapReduce aux utilisateurs de la technologie Hadoop qui étaient des membres de la fondation Apache (Metz Cade, 2010). Le dispositif « *Open Patent Non-Assertion Pledge* » mis en place par l'entreprise Google en 2013 avait donc principalement pour but de généraliser ce principe d'ouverture de la technologie MapReduce à l'ensemble des acteurs du secteur des technologies Big Data et non plus uniquement aux contributeurs de la fondation Apache.

Cette vision temporelle permet ainsi de comprendre pourquoi les brevets MapReduce n'ont pas reçu significativement plus de citations après la mise en place du dispositif « *Open Patent Non-Assertion Pledge* » en 2013, ces brevets étaient déjà largement utilisés dans le secteur des technologies Big Data notamment au travers du projet Hadoop de la fondation Apache. Concernant les brevets follow-on innovation, on peut penser que le dispositif « *Open Patent Non-Assertion Pledge* » a également permis de rassurer la communauté des utilisateurs de la technologie MapReduce sur les intentions de l'entreprise Google concernant son droit propriété intellectuelle. L'abaissement de cette barrière sur la propriété intellectuelle a eu un « effet ricochet » sur les technologies ayant un lien de filiation avec la technologie MapReduce, les utilisateurs ayant l'assurance que ces technologies ne seraient pas non plus bloquées par les brevets détenus par l'entreprise Google. Le risque de litige étant écarté, l'ensemble des acteurs du secteur des technologies des Big Data ont pu donc utiliser ces technologies pour leur propre développement, d'où un nombre de citations plus important sur la période qui a suivi l'ouverture directe des brevets MapReduce.

Les estimations suivantes complètent cette première analyse avec le niveau de citations des follow-on innovations de MapReduce afin de voir si un niveau spécifique capte une majorité de citations. Les premiers résultats sont présentés dans le tableau 2.15. Les résultats confirment

²² [Apache Projects List](https://projects.apache.org/projects.html?number) : <https://projects.apache.org/projects.html?number>

les premiers résultats obtenus sur ces estimations avec des données désagrégées. Les brevets pledge MapReduce reçoivent significativement plus de citations que les brevets ayant des technologies similaires dans les deux périmètres d'estimations avec et sans autocitations avant l'ouverture en Pledge. L'interaction de ces brevets avec l'indicatrice « post-pledge » ne donne pas de résultats statistiquement significatifs sur l'ensemble des estimations. Concernant le type de follow-on innovation, il apparaît que l'ensemble des niveaux de citations reçoivent significativement moins de citations avant l'ouverture, un résultat semblable à notre estimations précédente.

Tableau 2.15: Nombre de citations reçues par année avec niveaux de citations des follow-on innovations

	Nombre de Citations reçues par année			
	Avec autocitations		Sans autocitations	
	(1)	(2)	(3)	(4)
Niveau de citation 1	0.07 [0.13]	-0.02 [0.13]	-0.12 [0.08]	-0.23** [0.09]
Niveau de citation 2	-0.40*** [0.04]	-0.44*** [0.12]	-0.32*** [0.04]	-0.43*** [0.12]
Niveau de citation 3	-0.61*** [0.03]	-0.59*** [0.08]	-0.50*** [0.03]	-0.40*** [0.12]
Brevets Pledge MapReduce	1.05*** [0.28]	0.89*** [0.27]	0.82*** [0.12]	0.66*** [0.12]
Post-pledge x Niveau de citation 1	0.25** [0.12]	0.27** [0.11]	0.43*** [0.10]	0.44*** [0.10]
Post-pledge x Niveau de citation 2	0.50*** [0.16]	0.66*** [0.17]	0.39** [0.16]	0.57*** [0.18]
Post-pledge x Niveau de citation 3	0.31*** [0.09]	0.48*** [0.11]	0.16*** [0.05]	0.23** [0.10]
Post-pledge x Brevets Pledge MapReduce	-0.13 [0.09]	-0.13 [0.10]	-0.16 [0.11]	-0.17 [0.11]
Sponsors Apache	0.01 [0.02]		-0.01 [0.02]	
Co-brevets	-0.13*** [0.05]	-0.07 [0.05]	-0.06 [0.05]	-0.02 [0.05]
Nombre d'inventeurs	0.03 [0.02]	0.03 [0.02]	0.02 [0.02]	0.02 [0.02]
Nombre de sous-groupes CPC	0.02 [0.02]	-0.03 [0.02]	-0.01 [0.02]	-0.03 [0.02]
Taille de la famille	-0.02 [0.03]	0.02 [0.03]	-0.02 [0.03]	0.01 [0.03]
Nombre de citations antérieurs	0.03*** [0.01]	0.01 [0.01]	0.01 [0.01]	0.00 [0.01]
Nombre de citation hors brevets	0.03*** [0.01]	0.03*** [0.01]	0.02*** [0.01]	0.02*** [0.01]
Nombre de revendications	0.07*** [0.02]	0.01 [0.03]	0.06** [0.02]	0.00 [0.03]
Constant	0.83*** [0.08]	1.11*** [0.09]	0.94*** [0.07]	1.12*** [0.09]
EF Années de la citation	Oui	Oui	Oui	Oui
EF technologie	Oui	Oui	Oui	Oui
EF Entreprises citées	Non	Oui	Non	Oui
Observations	57063	56611	57063	56611
R-Squared	.04	.21	.028	.18

Cependant si on considère la période ultérieure à 2013, les coefficients associés aux trois niveaux de citations deviennent tous significatifs et positifs sur l'ensemble des estimations avec ou sans autocitations. Ce résultat vient confirmer notre analyse précédente. Si on regarde plus attentivement les coefficients des différentes interactions, il apparaît que les follow-on innovations de niveau 2 reçoivent en moyenne un peu plus de citations sur la période ultérieure à 2013 que les deux autres niveaux si on tient compte des autocitations (coefficient à 0.66 contre 0.27 pour le niveau 1 et 0.48 pour le niveau 3 sur l'estimation 2 intégrant des effets fixes « entreprises »). Cette tendance est conservée si on exclut les autocitations.

A l'instar des estimations agrégées au niveau des brevets cités, les estimations présentées dans le tableau 2.16 intègrent des variables de contrôle supplémentaires sur la proximité entre les inventeurs des brevets cités et citant afin conforter nos analyses précédentes.

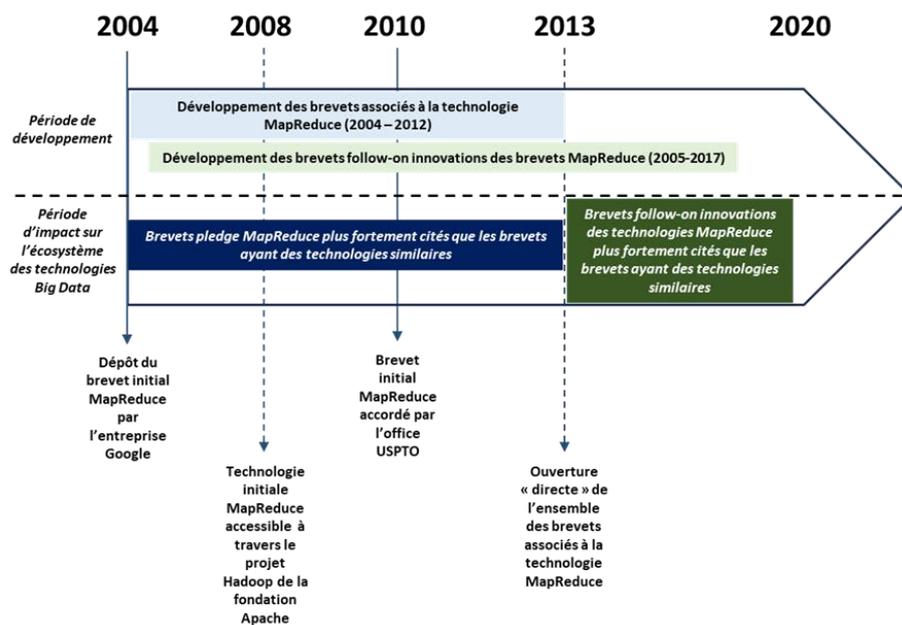


Figure 2.11 : synthèse de l'impact des brevet pledge MapReduce et follow-on innovations

(Charles XAVIER, 2023)

L'ajout des variables de contrôle sur la proximité entre les inventeurs des brevets cités et citant ne modifie pas la tendance des résultats des estimations précédentes. Sur la période antérieure à leur ouverture directe par le dispositif « *Open Patent Non-Assertion Pledge* » en 2013, les brevets pledge MapReduce reçoivent plus de citations que les brevets ayant des technologies similaires, tandis que les brevets follow-on innovations de ces technologies

reçoivent un nombre de citations plus important sur la période ultérieure au déploiement de ce même dispositif. Ces résultats peuvent être synthétisés par la Figure 2.11.

2.7 Conclusion

Ce chapitre avait pour objectif d'étudier les mécanismes d'innovation ouverte associés à l'une des principales technologies de traitement Big Data : MapReduce. En effet, la technologie MapReduce a bénéficié de deux mécanismes distincts d'innovation ouverte :

- L'accessibilité à travers la fondation communautaire Apache et son projet open source Hadoop
- L'accessibilité directe par le dispositif « *Open Patent Non-Assertion Pledge* » mis en place par l'entreprise Google en 2013

Des estimations économétriques sur les citations de 32 320 brevets de l'office USPTO ont été réalisées. Cet échantillon de brevets se décompose en deux groupes, d'une part, le groupe « traité » qui est l'ensemble des brevets associés à la technologie MapReduce et à ses follow-on innovations et d'autre part, le groupe « contrôle » constitué de technologies de traitement de données similaires mais n'ayant aucun lien avec la technologie MapReduce. Deux types d'estimations ont été réalisées, d'une part, des estimations agrégées sur le nombre de citations reçues par brevet et d'autre part, des estimations désagrégées sur le nombre de citation reçues par brevet et par année.

Les estimations agrégées sur le nombre de citations reçues par brevet ont mis en évidence que les brevets appartenant aux sponsors Apache platine qui citent directement les brevets pledge MapReduce reçoivent significativement plus de citations que les brevets ayant des technologies similaires. Ce constat est le même si on considère les autocitations des organisations. Par ailleurs, si l'on considère les niveaux de citations, il apparaît donc que les brevets follow-on innovations de niveau 1 des technologies MapReduce ont permis d'augmenter la valeur privée des sponsors platines, tout en favorisant l'essor des technologies de traitement de données Big Data. A travers les estimations désagrégées sur le nombre de citations reçues par brevet et par année, il a été possible d'affiner notre approche empirique en tenant compte de la dimension temporelle. Il apparaît que sur la période 2004-2012, période antérieure au dispositif d'ouverture directe des technologies MapReduce, les brevets affiliés à cette technologie reçoivent un nombre de citations significativement plus élevé sur les deux périmètres

d'estimations avec et sans autocitations. Ce résultat s'explique par le fait que le dispositif d'ouverture des brevets MapReduce déployé par la société Google en 2013 a été une généralisation d'un principe acquis depuis 2008, à savoir l'accessibilité de la technologie MapReduce au travers du projet open source Hadoop de la fondation Apache. Ce résultat sur les brevets pledge MapReduce est semblable avec et sans autocitations. Ces brevets, sur la période 2004-2012, ont donc permis d'accroître la valeur privée de la société Google mais ils ont également participé à l'essor du secteur des technologies Big Data. Sur la période ultérieure à 2013 et le déploiement du dispositif d'ouverture directe des technologies MapReduce, ce sont les brevets follow-on innovations, de tous types, qui reçoivent significativement plus de citations que ce soit avec ou sans autocitations. Ces brevets ont donc eu un impact sur la valorisation des entreprises que les ont développées, mais également sur le secteur des technologies Big Data.

Les analyses empiriques de ce chapitre ont mis en avant que l'accessibilité forte de la technologie MapReduce a favorisé l'innovation au sein des secteurs des technologies Big Data. Ce constat nous amène à une question centrale de notre analyse : est-ce l'accessibilité aisée à la technologie MapReduce ou son caractère innovant par rapport aux autres technologies de traitement de données qui a permis à cette technologie de devenir l'une des technologies majeures du secteur des technologies Big Data ? Pour répondre à cette question, il est nécessaire d'élargir notre champ d'analyse et de s'interroger sur l'impact global de l'ouverture des technologies de traitement de données « Big Data » sur la valorisation de l'espace technologique des organisations.

Tableau 2.16 : Nombre de citations reçues par année avec niveaux de citations des follow-on innovations

	Nombre de Citations reçues par année							
	Avec autocitation		Sans autocitation		Avec autocitations		Sans autocitation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Brevets Pledge MapReduce et follow-on innovation (groupe traite)	-0.07 [0.10]	-0.14 [0.11]	-0.19*** [0.06]	-0.28*** [0.08]				
Brevets Pledge MapReduce	1.13*** [0.29]	1.03*** [0.28]	1.03*** [0.13]	0.95*** [0.13]	1.06*** [0.27]	0.90*** [0.26]	0.84*** [0.11]	0.67*** [0.11]
Post-pledge x Brevets Pledge MapReduce et follow-on innovation (groupe traite)	0.21+ [0.12]	0.34*** [0.13]	0.30*** [0.10]	0.42*** [0.11]				
Post-pledge x Brevets Pledge MapReduce	-0.35** [0.15]	-0.48*** [0.16]	-0.47*** [0.15]	-0.60*** [0.16]	-0.14 [0.10]	-0.14 [0.10]	-0.17 [0.11]	-0.18 [0.11]
Sponsors Apache	0.02 [0.02]		0.00 [0.02]		0.01 [0.02]		-0.00 [0.02]	
Co-brevet	-0.14*** [0.05]	-0.08 [0.05]	-0.07 [0.05]	-0.02 [0.06]	-0.14*** [0.05]	-0.08 [0.05]	-0.07 [0.05]	-0.02 [0.05]
Autocitation inventeur	-0.20*** [0.04]	-0.12*** [0.03]	-0.16*** [0.04]	-0.12*** [0.03]	-0.19*** [0.04]	-0.12*** [0.03]	-0.15*** [0.04]	-0.11*** [0.03]
Proximité géodésique inventeurs == 1 (0/1)	-0.00 [0.04]	-0.02 [0.04]	-0.05 [0.04]	-0.07** [0.03]	-0.01 [0.04]	-0.02 [0.04]	-0.05 [0.04]	-0.07** [0.03]
Proximité géodésique inventeurs == 2 (0/1)	0.11*** [0.03]	0.05+ [0.03]	0.05+ [0.03]	0.01 [0.03]	0.12*** [0.03]	0.05+ [0.03]	0.06+ [0.03]	0.01 [0.03]
Proximité géodésique inventeurs == 3 (0/1)	0.14*** [0.02]	0.12*** [0.02]	0.12*** [0.02]	0.11*** [0.02]	0.14*** [0.02]	0.12*** [0.02]	0.12*** [0.02]	0.11*** [0.02]
Nombre de co-inventeurs (Max)	-0.00+ [0.00]	-0.00 [0.00]	-0.00** [0.00]	-0.00 [0.00]	-0.00 [0.00]	-0.00 [0.00]	-0.00+ [0.00]	-0.00 [0.00]
Même organisation	0.16*** [0.02]	0.13*** [0.02]	0.13*** [0.02]	0.11*** [0.02]	0.16*** [0.02]	0.13*** [0.02]	0.13*** [0.02]	0.11*** [0.02]
Même État	0.07*** [0.01]	0.05*** [0.01]	0.07*** [0.01]	0.05*** [0.01]	0.07*** [0.01]	0.05*** [0.01]	0.07*** [0.01]	0.05*** [0.01]
Nombre d'inventeurs	0.02 [0.02]	0.01 [0.02]						
Nombre de sous-groupes CPC	0.01 [0.02]	-0.03 [0.02]	-0.01 [0.02]	-0.04+ [0.02]	0.01 [0.02]	-0.03 [0.02]	-0.01 [0.02]	-0.04 [0.02]
Taille de la famille	-0.03 [0.03]	0.02 [0.03]	-0.03 [0.03]	0.01 [0.03]	-0.03 [0.03]	0.02 [0.03]	-0.02 [0.03]	0.01 [0.03]
Nombre de citations antérieurs	0.03*** [0.01]	0.01 [0.01]	0.02 [0.01]	0.01 [0.01]	0.03*** [0.01]	0.01 [0.01]	0.02+ [0.01]	0.01 [0.01]

Nombre de citation hors brevets	0.03***	0.03***	0.02***	0.02**	0.03***	0.03***	0.02***	0.02**
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de revendications	0.07***	0.01	0.06**	0.00	0.07***	0.01	0.06**	0.00
	[0.02]	[0.03]	[0.02]	[0.03]	[0.02]	[0.03]	[0.02]	[0.03]
Niveau de citation 1					0.07	-0.01	-0.13	-0.22**
					[0.13]	[0.13]	[0.08]	[0.09]
Niveau de citation 2					-0.38***	-0.43***	-0.31***	-0.42***
					[0.04]	[0.12]	[0.04]	[0.12]
Niveau de citation 3					-0.57***	-0.56***	-0.47***	-0.39***
					[0.03]	[0.09]	[0.04]	[0.12]
Post-pledge x Niveau de citation 1					0.24**	0.26**	0.43***	0.43***
					[0.12]	[0.11]	[0.10]	[0.10]
Post-pledge x Niveau de citation 2					0.48***	0.64***	0.37**	0.55***
					[0.16]	[0.17]	[0.16]	[0.18]
Post-pledge x Niveau de citation 3					0.27***	0.45***	0.14***	0.21**
					[0.08]	[0.11]	[0.05]	[0.10]
Constant	0.81***	1.09***	0.91***	1.10***	0.81***	1.09***	0.91***	1.10***
	[0.07]	[0.09]	[0.07]	[0.09]	[0.07]	[0.09]	[0.07]	[0.09]
EF Années de la citation	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
EF technologie	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
EF Entreprises citées	Non	Oui	Non	Oui	Non	Oui	Non	Oui
Observations	57063	56611	57063	56611	57063	56611	57063	56611
R-Squared	.056	.22	.039	.19	.058	.22	.041	.19

2.8 Références bibliographiques

- Ayerbe, C., & Azzam, J. E. (2015). Pratiques coopératives dans l'Open Innovation : Les enseignements des patent pools. *Management International*, 19(2), 95–114. <https://doi.org/10.7202/1030389ar>
- Baudry, M., & Dumont, B. (2020). Que nous apprennent les brevets de Tesla Motors sur sa stratégie PI ? https://www.openscience.fr/IMG/pdf/iste_techinn20v5n4_6.pdf
- Bengtsson, M., & Kock, S. (1999). Cooperation and competition in relationships between competitors in business networks. *Journal of Business & Industrial Marketing*, 14(3), 178–194. <https://doi.org/10.1108/08858629910272184>
- Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, 9(4), 439–468. <https://doi.org/10.1093/jeg/lbp008>
- Bronzo, N. (2022). Les Patent Pledges. *Revue Générale de Droit Médical, Panorama de Droit Pharmaceutique*, Pp.103. <https://doi.org/10.17863/CAM.48822>
- Chang, V., Mills, H., & Newhouse, S. (2007). From Open Source to long-term sustainability: Review of Business Models and Case studies. https://www.researchgate.net/publication/39995080_From_Open_Source_to_long-term_sustainability_Review_of_Business_Models_and_Case_studies
- Charleux, A. (2020). L'open source entre concurrents - Approche de la création et de l'appropriation de valeurs par les business models et la coopération. <https://hal.archives-ouvertes.fr/tel-02523313>
- Chen, Y. S., & Chen, B. Y. (2011). Utilizing patent analysis to explore the cooperative competition relationship of the two LED companies: Nichia and Osram. *Technological Forecasting and Social Change*, 78(2), 294–302. <https://doi.org/10.1016/j.techfore.2010.06.017>
- Clark, J. (2000). Patent pools: a solution to the problem of access in biotechnology patents? United States Patent And Trademark Office.
- Cohendet, P., Pénin, J., Farcot, M., Henri Tudor, C., & Matthieufarcot, L. (2006). Entre incitation et coordination : Repenser le rôle du brevet d'invention dans une économie fondée sur la connaissance. <https://www.researchgate.net/publication/255599313>
- Contreras, J. L. (2013). Technical standards and “ex ante” disclosure: results and analysis of an Empirical study. *Jurimetrics*, Vol. 53, No. 1

- Contreras, J. L. (2015a). A Market Reliance Theory for FRAND Commitments and Other Patent Pledges. *In Utah Law Review (Vol. 2015, Issue 2)*. <http://dc.law.utah.edu/ulr> Available at: <http://dc.law.utah.edu/ulr/vol2015/iss2/5>
- Contreras, J.L.(2015b). Patent Pledges. <http://ssrn.com/abstract=2525947> <https://ssrn.com/abstract=2525947> Electronic copy available at: <http://ssrn.com/abstract=2525947>
- Dagnino, G. B., Leroy, F., & Yami, S. (2007). La dynamique des stratégies de coopération. *Revue Française de Gestion, 33(176)*, 87–98. <https://doi.org/10.3166/rfg.176.87-98>
- de Rassenfosse, G., & Palangkaraya, A. (2023). Do patent pledges accelerate innovation? *Research Policy, 52(5)*. <https://doi.org/10.1016/j.respol.2023.104745>
- Dhyani, B., & Barthwal, A. (2014). Big Data Analytics using Hadoop. *In International Journal of Computer Applications (Vol. 108, Issue 12)*.
- Ehrnsperger, J. F., & Tietze, F. (2019). Patent pledges, open IP, or patent pools? Developing taxonomies in the thicket of terminologies. *PLoS ONE, 14(8)*. <https://doi.org/10.1371/journal.pone.0221411>
- Gueguen, G. (2009). Coopetition and business ecosystems in the information technology sector: the example of Intelligent Mobile Terminals. *In Int. J. Entrepreneurship and Small Business (Vol. 8, Issue 1)*.
- Hall, B. H., & Harhoff, D. (2012). Recent research on the economics of patents. *Annual Review of Economics, 4*, 541–565. <https://doi.org/10.1146/annurev-economics-080511-111008>
- Hall, B. H., & Helmers, C. (2013). Innovation and diffusion of clean/green technology: Can patent commons help? *Journal of Environmental Economics and Management, 66(1)*, 33–51. <https://doi.org/10.1016/j.jeem.2012.12.008>
- Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market Value and Patent Citations. *In The RAND Journal of Economics (Vol. 36, Issue 1)*.
- Hall, B. H., & Ziedonis, R. H. (2001). The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995. *In The RAND Journal of Economics (Vol. 32, Issue 1)*.
- Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation Frequency and the Value of Patented Inventions. *In Source: The Review of Economics and Statistics (Vol. 81, Issue 3)*.
- Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *In Research Policy (Vol. 32)*.

- Khezr, S. N., & Navimipour, N. J. (2017). MapReduce and Its Applications, Challenges, and Architecture: a Comprehensive Review and Directions for Future Research. *In Journal of Grid Computing (Vol. 15, Issue 3, pp. 295–321)*. Springer Netherlands. <https://doi.org/10.1007/s10723-017-9408-0>
- Lado, A. A., Boyd, N. G., & Hanlon, S. C. (1997). Competition, cooperation, and the search for economic rents: a syncretic model. *In Academy of Management Review (Vol. 22, Issue 1)*.
- Lampe, R., & Moser, P. (2010). Do Patent Pools Encourage Innovation? Evidence from the 19th-Century Sewing Machine Industry. *The Journal of Economic History, vol. 70, no. 4, pp. 898–920*. JSTOR, <http://www.jstor.org/stable/40984782>.
- Lanjouw, J. O., Pakes, A., Putnam, J., & Putnam', J. (1998). How to count patents and value intellectual property: uses of patent renewal and application data. *The Journal of Industrial Economics, vol. 46, no. 4, pp. 405–32*. JSTOR, <http://www.jstor.org/stable/117497>
- Lanjouw, J. O., & Schankerman, M. (2001). Characteristics of Patent Litigation: A Window on Competition. *The RAND Journal of Economics, vol. 32, no. 1, pp. 129–51*. JSTOR, <https://doi.org/10.2307/2696401>.
- Lerner, J., Strojwas, M., & Tirole, J. (2007). The design of patent pools: the determinants of licensing rules. *In Journal of Economics (Vol. 38, Issue 3)*.
- Lévêque, F. (2007). La normalisation et le droit de la concurrence face au hold-up. *Revue Lamy de La Concurrence, N° 12*.
- Lim, K., Chesbrough, H., & Yi, R. (2010). Open innovation and patterns of R&D competition. *In Int. J. Technology Management (Vol. 52)*.
- Metz Cade. (2010, April 27). *Google blesses Hadoop with MapReduce patent license*. https://www.TheRegister.Com/2010/04/27/Google_licenses_mapreduce_patent_to_hadoop/.
- Nalebuff, B. J., & Brandenburger, A. M. (1997). Co-opetition: Competitive and cooperative business strategies for the digital economy. *In Strategy & Leadership (Vol. 25, Issue 6, pp. 28–33)*. <https://doi.org/10.1108/eb054655>
- Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between U.S. technology and public science. *In Elsevier Research Policy (Vol. 26)*.
- Porter, M. E. (1991). Towards a dynamic theory of strategy. *Strategic Management Journal, 12(2 S), 95–117*. <https://doi.org/10.1002/smj.4250121008>

- Rysman, M., & Simcoe, T. S. (2005). Patents and the Performance of Voluntary Standard Setting Organizations. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.851245>
- Sapsalis, E., van Pottelsberghe de la Potterie, B., & Navon, R. (2006). Academic versus industry patenting: An in-depth analysis of what determines patent value. *Research Policy*, *35*(10), 1631–1645. <https://doi.org/10.1016/j.respol.2006.09.014>
- Shapiro, Carl., & Varian, H. R. (1999). Information rules: a strategic guide to the network economy. *Harvard Business School Press*.
- Singh, J., & Agrawal, A. (2011). Recruiting for ideas: How firms exploit the prior inventions of new hires. *Management Science*, *57*(1), 129–150. <https://doi.org/10.1287/mnsc.1100.1253>
- Vakili, K. (2012). Competitive Effects of Modern Patent Pools: The Effect of the MPEG-2 Pool on Incumbents? Innovative Performance. *Paper to be presented at the DRUID 2012 on June 19 to June 21 at CBS, Copenhagen, Denmark*.
- Vakili, K. (2016). Collaborative promotion of technology standards and the impact on innovation, industry structure, and organizational capabilities: Evidence from modern patent pools. *Organization Science*, *27*(6), 1504–1524. <https://doi.org/10.1287/orsc.2016.1098>
- Yang, N., Ferreira, I., Serebrenik, A., & Adams, B. (2022). Why do projects join the apache software foundation? *ICSE-SEIS '22: Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society*, 161–171. <https://doi.org/10.1145/3510458.3513006>

CHAPITRE 3. IMPACT DES PATENTS PLEDGE SUR LA VALEUR DES PORTEFEUILLES TECHNOLOGIQUES DES ENTREPRISES

Le chapitre 2 a mis en lumière que la technologie MapReduce, technologie initialement développée par deux ingénieurs de l'entreprise Google en 2004, avait eu un impact sur le secteur des technologies de traitement des Big Data. En effet, les estimations empiriques sur le nombre de citations reçues par brevet par année ont mis en évidence que les brevets MapReduce avaient reçu significativement plus de citations que les brevets ayant des technologies similaires sur la période 2004-2012. Cette technologie a donc eu un impact sur l'innovation au sein du secteur des technologies Big Data, d'une part pour l'entreprise Google dans la construction de sa propriété intellectuelle (à travers les autocitations) et d'autre part, sur les autres entreprises du secteur (cf. les estimations empiriques sans autocitation).

La technologie MapReduce a fait l'objet d'une double ouverture à travers le projet open-source Hadoop initié par la fondation Apache en 2008 mais également grâce au dispositif « *Open Patent Non-Assertion Pledge* » déployé par la firme Google en 2013. A ce stade, une question se pose : est-ce l'accessibilité facilitée qui en a fait une technologie majeure du traitement Big Data ou est-ce le caractère innovant intrinsèque de la technologie qui a permis d'impacter le secteur d'activité ?

L'ouverture de la technologie MapReduce n'est pas le seul exemple d'ouverture d'une technologie, en effet, le dispositif « *Open Patent Non-Assertion Pledge* » de l'entreprise Google comporte plus de 200 brevets. L'initiative de la société Google n'est pas une initiative isolée, des entreprises concurrentes telles que Microsoft ou IBM ont également ouvert une partie de leur propriété intellectuelle sur les technologies de traitement de données. De plus, ces initiatives ne sont pas uniquement concentrées sur le secteur des TIC, les entreprises Tesla ou Toyota pour l'automobile mais également Monsanto pour le secteur agricole ou Moderna pour l'industrie pharmaceutique ont également mis en place des dispositifs similaires permettant d'accéder à une partie de leur propriété intellectuelle. L'ouverture des technologies est donc une stratégie de propriété intellectuelle qui tend à se populariser au sein de nombreux secteurs d'activités.

Afin de comprendre si la technologie MapReduce a été plus portée par son accessibilité ou son caractère innovant, il est nécessaire d'élargir notre champ d'analyse à d'autres technologies

ouvertes et ainsi avoir des références supplémentaires sur l'impact de l'ouverture des technologies sur la valeur des portefeuilles technologiques des entreprises. Les stratégies d'ouverture des technologies trouvent leur origine dans le principe de « *l'innovation ouverte* » développé par Henry Chesbrough dans les années 2000. Depuis les premiers écrits de Chesbrough sur l'innovation ouverte, la littérature économique s'est grandement intéressée à l'impact de l'ouverture des technologies sur l'innovation des entreprises. Dans un premier temps, il est nécessaire de s'intéresser aux différentes stratégies d'innovation ouverte afin de pouvoir positionner la technologie MapReduce dans ce cadre d'analyse.

3.1 Les stratégies d'innovation ouverte

Avant de s'intéresser aux stratégies d'innovation ouverte, il est important de comprendre les fondements du principe d'innovation ouverte.

3.1.1 Le principe de l'innovation ouverte

Dans ses différentes publications, Chesbrough définit une stratégie d'innovation ouverte comme « *l'antithèse* » des stratégies traditionnelles d'intégration verticale où l'innovation interne conduit à la production de biens et services vendus par l'entreprise (Chesbrough, 2006). Chesbrough caractérise ces stratégies traditionnelles comme des approches « *d'innovation fermée* » qui conduisent à la création d'un cercle vertueux décrit dans la Figure 3.1.

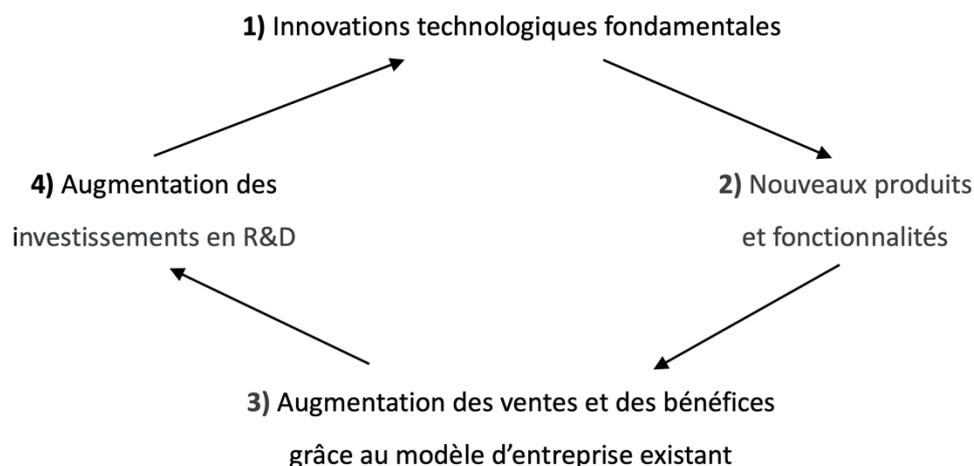


Figure 3.1 Cercle vertueux de l'innovation fermée (source : Chesbrough, 2003)

Cette stratégie d'innovation fermée repose sur le fait que les entreprises développent des innovations technologiques fondamentales, celles-ci leur permettent de créer de nouveaux produits et fonctionnalités. L'augmentation des profits qui découlent de la vente de ces produits, conduit à accroître les investissements en recherche et développement, ce qui aboutit à de nouvelles innovations. Au cours du XXème siècle, cette stratégie d'innovation fermée a grandement été utilisée par des entreprises dans tous les secteurs d'activité, comme l'industrie chimique allemande qui s'est appuyée sur de nombreux laboratoires de recherche pour se développer, ou des sociétés américaines telles que Général Electric, société fondée par Thomas Edison afin de promouvoir ses propres inventions (Chesbrough, 2003).

Ce modèle d'innovation fermée a été remis en cause à la fin du XXème de siècle au travers de plusieurs facteurs (Chesbrough, 2003b) :

- L'accroissement de la mobilité des employés hautement qualifiés (chercheurs, ingénieurs) qui, en quittant leurs entreprises, emportent avec eux les connaissances acquises à travers leur expérience ;
- L'augmentation de l'offre de formation de l'enseignement, qui s'accompagne d'une hausse des diplômés hautement qualifiés. Les connaissances scientifiques et techniques ne sont plus uniquement concentrées dans les laboratoires de recherche des grandes entreprises, elles se sont diffusées dans des entreprises de toute taille ;
- L'émergence des « start-ups », des petites structures agiles et à la pointe des dernières technologies qui sont venues concurrencer les grandes entreprises et leur modèle d'innovation ;
- La diminution du cycle de vie des produits qui oblige les entreprises à innover toujours plus rapidement afin de conserver leur leadership sur les marchés.

Ces facteurs font que l'innovation fermée, en s'appuyant exclusivement sur des ressources internes, peuvent limiter la capacité d'innover des entreprises, et il est donc nécessaire que les entreprises s'ouvrent également aux connaissances externes afin de continuer à innover. Cette combinaison entre ressources internes et connaissances externes est le point de départ du principe d'innovation ouverte, principe que Chesbrough définit comme « *l'utilisation volontaire de flux de connaissances internes et externes pour accélérer l'innovation interne afin d'élargir les marchés par une utilisation externe de l'innovation* » (Chesbrough, 2006, traduction Xavier, 2023). Afin de créer ce « flux de connaissances internes et externes », Chesbrough met en avant l'importance du partage des connaissances à travers différentes

formes de collaborations entre les organisations : alliances, joint-ventures ou licences de propriété intellectuelle, l'objectif étant de créer un tunnel d'innovation (Chesbrough, 2003a).

Prolongeant les travaux de Chesbrough, Pénin (2008) propose des conditions plus strictes sur le partage des connaissances pour définir une stratégie d'innovation ouverte. Selon lui, trois facteurs sont nécessaires pour caractériser une telle stratégie :

1. La divulgation volontaire des connaissances ;
2. L'ouverture des connaissances ;
3. Des interactions continues entre les parties prenantes.

Pour Pénin, les stratégies d'innovation ouverte ne sont donc pas uniformes, certaines approches sont dites « faibles » et d'autres « fortes » selon le degré de partage des connaissances et donc d'ouverture des technologies. C'est en analysant ce degré d'ouverture que Euchner (2010) va caractériser deux approches d'innovation ouverte : « *l'innovation à frontières ouvertes* » et « *l'innovation open-source* » que nous allons présenter dans les paragraphes suivants.

3.1.2 La stratégie d'innovation à « frontières ouvertes »

La stratégie d'innovation à « frontières ouvertes » développée par Euchner se rapproche du concept initial d'innovation ouverte introduit par Chesbrough, c'est-à-dire utiliser les savoirs internes et externes à l'entreprise pour créer de nouvelles innovations. Euchner va compléter cette vision en précisant que l'objectif de cette approche est de pouvoir générer des profits grâce à la vente des innovations créées. Deux principales caractéristiques définissent la stratégie d'innovation à « frontière ouverte » :

- **Propriété de l'innovation créée** : l'organisation à l'initiative de la stratégie va conserver la propriété des innovations développées. Les droits de propriété intellectuelle sont essentiels dans cette approche car ils permettent d'avoir un avantage concurrentiel.
- **Contrôle du processus d'innovation** : l'organisation à l'initiative du projet va contrôler toutes les étapes du processus d'innovation (définition des priorités, moyens déployés, choix des partenaires...) afin de s'assurer de l'efficacité de la démarche.

En pratique, cette stratégie repose sur la collaboration entre une organisation porteuse d'un projet, que l'on nomme « le chercheur » avec une entité capable d'accélérer le développement de ce projet, qui est désignée comme « le solveur ». Le « solveur » n'est pas forcément un expert

dans le domaine du « chercheur » mais il doit posséder une expertise permettant d'accélérer le processus d'innovation. Le « solveur » sera rémunéré pour sa contribution au projet tandis que le « chercheur » conservera le contrôle de l'innovation créée et des droits des propriétés intellectuelles qui en résultent.

Un exemple qui illustre concrètement cette stratégie d'innovation à « frontières ouvertes » est donné par l'entreprise Bell Atlantic, une entreprise américaine de télécommunications. Cette entreprise a souhaité équiper ces techniciens de terrain avec un appareil de test léger et robuste ayant une interface utilisateur simple et attrayante. Pour le côté « léger et robuste » Bell Atlantic a collaboré avec une entreprise spécialisée dans l'alpinisme, quant à l'interface, elle a été confiée à des experts en jeux vidéo (Euchner, 2013). A travers cet exemple, on retrouve également un élément initial de l'innovation ouverte portée, à savoir la volonté de collaborer avec tout type d'organisations sans distinction du secteur d'activité (Chesbrough, 2003a).

La stratégie d'innovation à « frontières ouvertes » peut également être illustrée par le concept des « hackathons ». Le terme hackathon est la combinaison entre les mots « hack » et « marathon », le mot « hack » n'est pas employé ici comme une activité de piratage illégale mais plutôt dans le sens de l'exploration et la recherche (Briscoe & Mulligan, 2018). Lors d'un hackathon, l'organisateur soumet une problématique précise à un groupe d'individus présentant des compétences diverses. Ces individus doivent résoudre cette problématique sur une période courte, généralement 2 à 3 jours et se voient récompenser par l'organisateur en cas de succès. A la fin du hackathon, l'organisateur conserve les éléments produits durant l'événement. On retrouve dans ce concept de hackathon, les principales caractéristiques de la stratégie d'innovation à « frontières ouvertes. Tout d'abord, l'appropriation des innovations créées par l'organisateur, puis le contrôle du processus d'innovation (sélection des participants, définition du temps imparti) et enfin la rémunération sous forme de récompenses pour les participants. Le concept de hackathon s'inscrit dans un modèle de production participative (plus connu sous le terme anglais de « crowdsourcing ») ayant pour objectif de stimuler l'innovation parmi des individus d'horizons divers grâce au partage de leurs connaissances afin de résoudre un objectif commun (Temiz, 2021).

De nombreuses organisations ont mis en place des hackathons, on peut citer l'exemple de l'entreprise Méta (Facebook) qui en organise depuis 2007²³. Il est intéressant de noter que Méta utilise les hackathons, d'une part, pour des usages internes avec des événements réservés aux employés pour stimuler leur créativité, et d'autre part, pour des usages externes afin d'identifier de nouvelles tendances technologiques. Les hackathons ne sont pas exclusifs au secteur des TIC, la marque de luxe Louis Vuitton a organisé un hackathon en 2015 afin de s'ouvrir aux nouvelles technologies sur le traitement des données Big Data²⁴.

Au total, il apparaît que le degré d'ouverture des technologies est limité dans une stratégie d'innovation à « frontières ouvertes », les organisations conservant la propriété de l'innovation créée et les droits de propriété intellectuelle qui en découlent. Dans la deuxième stratégie développée par Euchner, la stratégie d'innovation « open-source », une plus forte ouverture des technologies est au centre de l'approche.

3.1.3 La stratégie d'innovation « open-source »

La stratégie d'innovation à « frontières ouvertes » s'inscrit dans un paradigme classique de l'innovation ouverte en s'appuyant sur les premiers écrits de Chesbrough en la matière. Cette approche est portée par un modèle d'entreprise « classique » dans lequel l'entreprise est au cœur de la stratégie d'innovation. En revanche, la stratégie d'innovation « open-source » remet en cause ce modèle d'entreprise en déplaçant la source de l'innovation de l'entreprise vers une communauté des utilisateurs.

Pour Euchner (2013), c'est la communauté des utilisateurs anonymes qui va non seulement être porteuse des idées novatrices, mais également produire les innovations qui en résultent. Les membres de la communauté étant à la source de l'innovation, ils vont également maîtriser le processus d'innovation et bénéficier directement des résultats. Les caractéristiques d'une stratégie d'innovation « open-source » diffèrent donc fortement de celles à « frontières ouvertes ». En effet, dans une stratégie « open-source » la propriété intellectuelle est ouverte à l'ensemble de la communauté, elle n'est plus la propriété unique des organisations. De plus, le processus d'innovation va faire l'objet d'une gouvernance ouverte entre les différents membres de la communauté, le contrôle n'est plus exclusif aux organisations. En déplaçant la source de

²³ <https://mockitt.wondershare.com/hackathon/facebook-hackathons.html>

²⁴ <https://www.lvmh.fr/actualites-documents/actualites/hackathon-decoder-louis-vuitton/>

l'innovation des entreprises à la communauté des utilisateurs, la stratégie d'innovation « open-source » va avoir pour objectif l'amélioration des usages et non plus uniquement la recherche du profit.

L'industrie du logiciel ouvert, « open-source software », est une parfaite illustration de cette approche. Pour comprendre le principe d'un logiciel ouvert, il est important de définir les deux principales composantes d'un logiciel qui sont le code source et le code objet. L'agence pour la protection des programmes définit le code source « *comme un ensemble d'instructions à la source d'un programme informatique. Ces instructions sont exprimées dans un langage que l'homme est capable de comprendre et de modifier aisément* ». Le code objet est défini « *comme le code binaire du programme, compréhensible par une machine et obtenu en compilant le code source* »²⁵. Il apparaît que le code source est le cœur de l'innovation dans la création d'un logiciel et c'est sur ce point que va reposer le principe d'un logiciel ouvert. En effet, un logiciel est dit « ouvert » à partir du moment où son code source (et non uniquement son code objet) est librement accessible à tous et que les modifications apportées au code source par des utilisateurs sont librement accessibles par tous (Lerner & Tirole, 2001).

De nombreuses initiatives se sont développées à travers ce principe du logiciel ouvert comme le système d'exploitation Linux créé en 1991 par un étudiant finlandais Linus Torvalds ou encore le serveur HTTP de la fondation Apache (cf. chapitre précédent). Les entreprises ont également utilisé ce type d'approche pour distribuer certains de leur logiciel comme la suite bureautique OpenOffice mise à disposition par l'entreprise américaine Sun Microsystems. L'exemple d'OpenOffice est un cas d'étude intéressant dans l'univers du logiciel libre. En effet, cette suite bureautique historiquement provient du logiciel StarOffice développé par une entreprise allemande Star Division. En 2000, la société américaine Sun Microsystems rachète la société allemande et rend public le code source de StarOffice qui devient un logiciel ouvert nommé OpenOffice. En 2009, l'entreprise Sun Microsystems est à son tour rachetée par l'entreprise Oracle, qui décide deux ans plus tard de se désengager du projet OpenOffice.org et le cède à la fondation Apache (Wikipédia²⁶). L'exemple d'OpenOffice illustre concrètement le

²⁵ Agence pour la Protection des Programmes :<https://www.app.asso.fr/centre-information/base-de-connaissances/code-logiciels/la-protection-des-elements-composant-un-logiciel/la-protection-par-le-droit-dauteur-specifique-des-logiciels#:~:text=Le%20code%20source%20et%20le%20code%20objet&text=Ces%20instructions%20sont%20exprim%C3%A9es%20dans,en%20compilant%20le%20code%20source.>

²⁶ <https://fr.wikipedia.org/wiki/OpenOffice.org>

passage d'une stratégie d'innovation fermée (StarOffice) à une stratégie d'innovation « open-source » (OpenOffice) avec l'ambition d'offrir une alternative à la suite bureautique Office de la société Microsoft²⁷.

L'industrie du logiciel ouvert n'est pas le seul secteur d'activité à s'appuyer sur une stratégie d'innovation « open-source », le monde médical commence également à s'ouvrir à ce type d'approche. Dans leurs recherches sur les maladies tropicales (paludisme, dengue...), Maurer et al., (2004) constatent que seuls 1% des nouveaux médicaments produits sont destinés à soigner ce type de maladie. Ils attribuent ce constat au modèle d'innovation fondé sur les brevets qui caractérise cette industrie. Ce modèle fonctionne si les laboratoires peuvent vendre leurs médicaments brevetés afin de couvrir leurs frais en recherche et développement. Cependant, dans les pays en voie de développement qui sont les plus touchés par les maladies tropicales, les patients n'ont pas les moyens de payer des médicaments brevetés, ce qui n'incite pas les laboratoires à investir dans la recherche pour lutter contre ces maladies. Afin de palier l'échec du modèle d'entreprise de l'industrie pharmaceutique, Maurer et al proposent une approche open-source qu'ils nomment « initiative sur les maladies tropicales ». Cette initiative repose sur une communauté digitale décentralisée composée de laboratoires, d'universités, d'instituts et d'entreprises qui collaborent pour faire progresser la recherche sur ces maladies. La vision open-source de Maurer et al s'inspire largement du modèle de l'industrie du logiciel ouvert, ils précisent d'ailleurs que la recherche médicale a des points communs avec le développement informatique : « *dans les deux cas, la recherche consiste à trouver et résoudre de petits problèmes cachés dans un océan de code* ».

A travers sa vision communautaire, la stratégie d'innovation « open-source » est une stratégie dont le degré d'ouverture des technologies est fort. De plus, les différents exemples cités précédemment montrent que cette stratégie entraîne systématiquement une remise en cause du modèle d'entreprise classique basé sur la possession exclusive des innovations.

Les deux stratégies d'innovation ouverte que nous avons détaillées précédemment s'opposent par leur degré d'ouverture des technologies. En effet, pour la stratégie d'innovation à « frontière ouverte » la propriété de l'innovation est un enjeu important et il en résulte un degré d'ouverture limité des technologies. A l'inverse, dans la stratégie d'innovation « open-source », la source

²⁷ https://www.openoffice.org/press/sun_release.html

de l'innovation provient de la communauté d'utilisateurs et de ce fait le degré d'ouverture des technologies est fort. Le choix entre l'une des deux stratégies va donc être un choix motivé par les objectifs des entreprises relatifs aux débouchés des innovations créées. Pour la stratégie d'innovation à « frontière ouverte » c'est la recherche du profit qui va motiver ce type d'approche tandis que pour la stratégie « open-source » c'est l'amélioration des usages par la communauté qui sera la motivation première. Il est important de noter que le choix d'une stratégie d'innovation « open-source » n'est pas anodin pour une organisation car elle comporte des risques sur les revenus du fait de la renonciation à ses droits de propriété intellectuelle et du partage de ses connaissances avec une communauté indépendante. Cependant la stratégie d'innovation « open-source » comporte également des bénéfices certains comme la réduction des coûts et des temps de production des innovations, des bénéfices liés à la proximité avec une communauté d'utilisateurs-experts qui va accélérer le processus d'innovation (Euchner, 2013).

Après avoir défini ces deux types de stratégie, il est nécessaire de s'intéresser plus spécifiquement à la technologie MapReduce et d'essayer de positionner cette technologie dans le champ de l'innovation ouverte.

3.2 Positionnement de la technologie MapReduce dans le champ de l'innovation ouverte

3.2.1 Définition d'un cadre d'analyse

Les deux stratégies d'innovations ouvertes présentées précédemment diffèrent fortement par le degré d'ouverture des technologies qu'elles nécessitent. L'innovation à « frontière ouverte » se caractérise par une forte maîtrise du processus d'innovation et de la propriété intellectuelle, le degré d'ouverture des technologies est faible. A l'opposé la stratégie d'innovation « open-source » cède le processus d'innovation et la propriété intellectuelle à la communauté d'utilisateurs, le degré d'ouverture des technologies est donc fort.

Ces deux stratégies « à frontières ouvertes » et « open-source » s'appuient sur le principe fondateur de l'innovation ouverte défini par Chesbrough (2003) à savoir s'appuyer sur les connaissances internes et externes à l'entreprise pour développer des innovations, cependant elles sont antagoniques dans le degré d'ouverture des technologies qu'elles nécessitent. De ce fait, Euchner (2013) va indiquer qu'il y a un « *continuum d'innovation ouverte* » entre les

stratégies d'innovation à « frontières ouvertes » et « open-source » et qu'entre ces deux stratégies il y a une multitude de variantes. Ce « *continuum d'innovation ouverte* » est caractérisé par plusieurs facteurs comme le détaille la Figure 3.2.

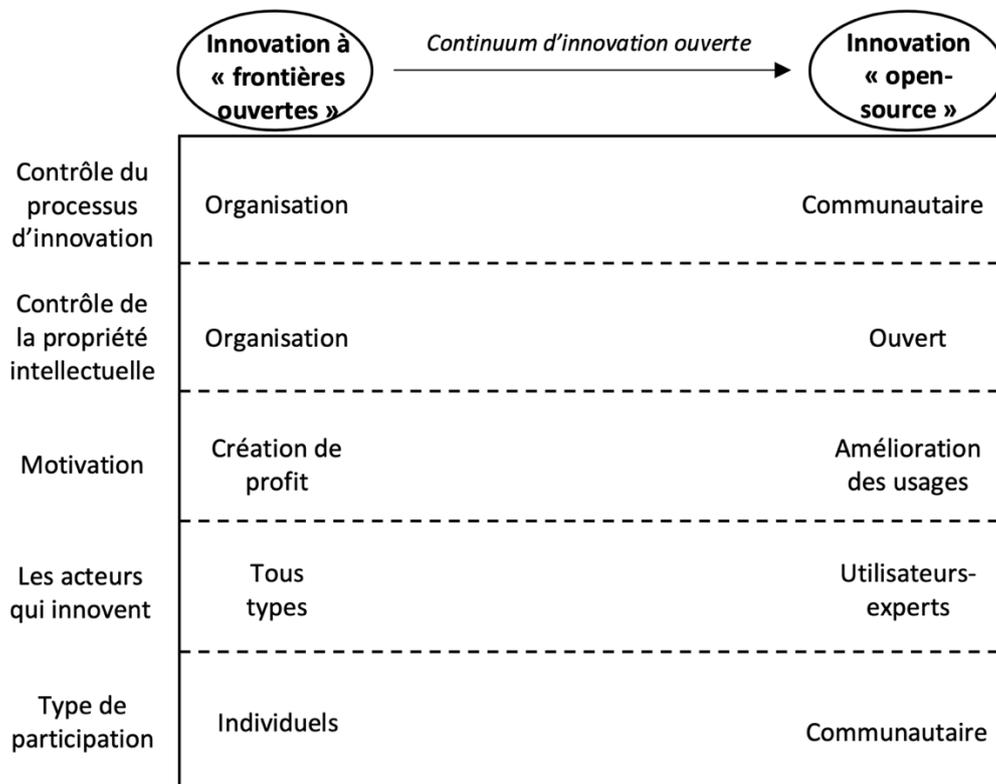


Figure 3.2 : le continuum d'innovation ouverte (Euchner, 2013)

Le « *continuum d'innovation ouverte* » repose sur 5 facteurs :

- Le contrôle du processus de l'innovation ;
- Le contrôle de la propriété intellectuelle ;
- La motivation ;
- Les acteurs qui innovent ;
- Le type de participation.

En suivant ce cadre d'analyse, il est possible de positionner une technologie et ainsi voir si elle se rapproche plus d'une stratégie d'innovation « à frontières ouvertes » ou d'une stratégie « open-source ». La section suivante analyse la technologie MapReduce selon ce cadre d'analyse.

3.2.2 Analyse des mécanismes d'innovation ouverte de la technologies MapReduce

Comme nous l'avons vu dans le chapitre 2, la technologie MapReduce a profité de deux mécanismes d'innovation ouverte, d'une part, un mécanisme communautaire grâce à l'accessibilité de la technologie par le projet Hadoop de la fondation Apache et d'autre part, un mécanisme d'ouverture de la propriété intellectuelle à travers le dispositif de l'entreprise Google « *Open Patent Non-Assertion Pledge* ». Si on se réfère au cadre d'analyse présenté précédemment, ces deux mécanismes ne se positionnent pas au même niveau dans le « *continuum d'innovation ouverte* ». En effet, le mécanisme lié à l'accessibilité de la technologie MapReduce par la fondation Apache nous indique que la stratégie d'innovation ouverte de cette technologie se rapproche d'une stratégie « open-source ». Cependant, la technologie MapReduce a fait l'objet de brevets et même si ceux-ci ont été ouverts par un mécanisme de patent pledge, ce mécanisme ne remet pas en cause le droit de propriété intellectuelle conféré par les brevets obtenus par l'entreprise Google. A travers ce mécanisme des patent pledge, la stratégie d'innovation ouverte de la technologie MapReduce tendrait plus vers une stratégie « à frontières ouvertes ». Au total, la juxtaposition de deux mécanismes d'innovation ouverte rend difficile le positionnement de la stratégie d'innovation de la technologie MapReduce. Il est donc nécessaire d'analyser individuellement les 5 facteurs du cadre d'analyse que nous avons énoncés précédemment afin de statuer sur la stratégie d'innovation de la technologie MapReduce.

Les deux premiers facteurs sont le contrôle du processus d'innovation et de la propriété intellectuelle. Comme la technologie MapReduce a fait l'objet d'un brevet, cela conduit mécaniquement à limiter ces deux facteurs. En effet, l'entreprise Google, à travers ses brevets, peut encadrer le processus d'innovation en choisissant d'accorder ou non des licences. Comme mentionné dans le chapitre précédent, la société Google a d'ailleurs accordé une licence sur le brevet initial MapReduce à la fondation Apache afin qu'elle puisse continuer à exploiter cette technologie au moment où le brevet a été accordé en 2010²⁸. De plus, le mécanisme des patent pledge associé à cette technologie amoindrit les droits de propriété intellectuelle mais ne les supprime pas. Concernant les motivations, l'entreprise Google a clairement mentionné à travers son dispositif « *Open Patent Non-Assertion Pledge* » qu'elle souhaitait « *s'engager à promouvoir l'innovation afin de favoriser la croissance et le progrès global des technologies*

²⁸ https://www.theregister.com/2010/04/27/google_licenses_mapreduce_patent_to_hadoop/

de l'information »²⁹. On peut donc penser que c'est l'amélioration des usages sur la technologie MapReduce qui motive la société Google et non la recherche du profit. Sur ce point, on peut aussi considérer que la recherche d'un standard technologique (cf. le chapitre 2) a été un point déterminant dans les motivations à ouvrir la technologie MapReduce. Enfin, concernant les deux derniers facteurs, à savoir les « acteurs qui innovent » et le type de participation, l'approche sur la technologie MapReduce est une approche hybride. D'une part, les acteurs qui innovent autour de cette technologie sont de tous types (via le mécanisme des patent pledge) mais également des experts de la communauté Apache. D'autre part, le type de participation est également individuel et communautaire.

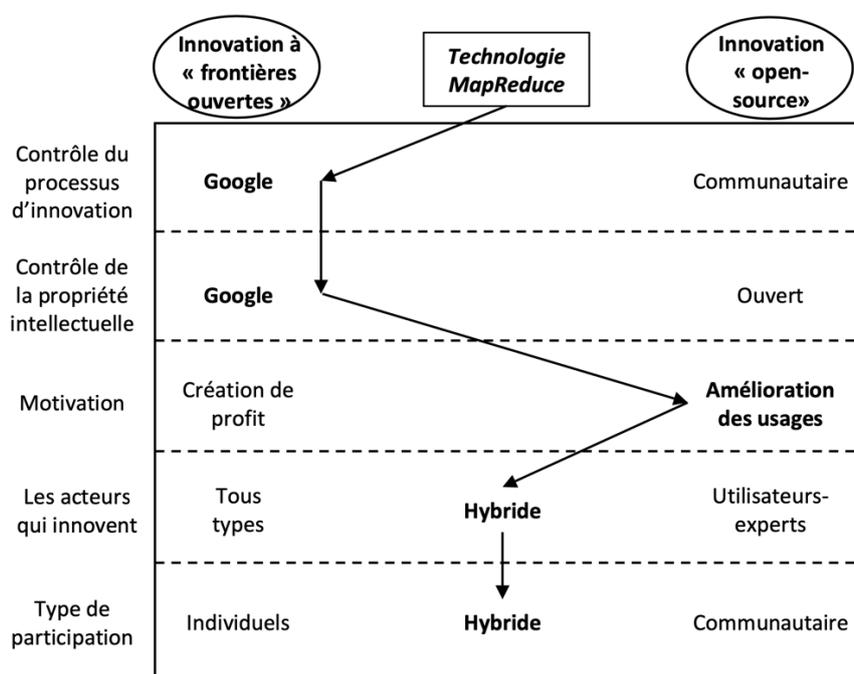


Figure 3.3 : la technologie MapReduce dans le continuum d'innovation ouverte (source : Euchner, 2013, modifié par Xavier, 2023)

Au total, le positionnement de la technologie MapReduce ne semble pas « spécifique » à l'une des deux stratégies d'innovation ouverte. La Figure 3.1 positionne la technologie MapReduce dans le « continuum d'innovation ouverte » et montre son *caractère hybride*.

En effet, sur les deux premiers facteurs liés aux contrôles du processus d'innovation et de la propriété intellectuelle, la technologie MapReduce se rapproche d'une stratégie d'innovation à

²⁹ <https://www.google.com/patents/opnpledge/pledge/>

« frontières ouvertes ». Sur le troisième facteur, la motivation, la technologie s'oriente vers une stratégie d'innovation « open source ». Enfin sur les deux derniers facteurs, la technologie adopte une approche hybride entre les deux stratégies. Si on compte le nombre de facteurs exclusifs par type de stratégie, il apparaît que la technologie MapReduce répond davantage au cadre de l'innovation à « frontières ouvertes » avec 2 facteurs sur 5 en faveur de cette approche.

Par ailleurs, il apparaît clairement que le « poids » des brevets MapReduce dans l'analyse de la stratégie d'innovation de cette technologie n'est pas négligeable. En effet, ces brevets permettent à la société Google de conserver le contrôle sur le processus d'innovation et sur la propriété intellectuelle de la technologie. Il en résulte que l'ouverture de la technologie reste limitée notamment pour la communauté des utilisateurs. Notre analyse du rôle des brevets MapReduce dans la définition du degré d'ouverture de cette technologie est corroborée par les analyses de Pénin (2008). En effet, il constate qu'une innovation peut être diffusée à de nombreuses parties prenantes tout en faisant l'objet d'une protection de propriété intellectuelle. Dans ces conditions, il considère que la stratégie d'innovation ouverte est « faible ». Il apparaît donc que la stratégie d'innovation de la technologie MapReduce est une stratégie d'innovation ouverte faible, du fait de l'existence de brevets qui limite son ouverture.

L'analyse précédente a montré que la technologie MapReduce se caractérise par un degré d'ouverture faible malgré la présence de deux mécanismes d'innovation ouverte étudiés au chapitre 2. L'objet du présent chapitre est d'aller plus loin dans la réflexion et de comprendre si c'est l'accessibilité aisée à la technologie MapReduce ou son caractère innovant intrinsèque qui a permis d'impacter le secteur d'activité des technologies Big Data. Pour ce faire, ce chapitre s'intéresse aux technologies ayant un degré d'ouverture semblable par le biais des patent pledge afin d'en étudier l'impact. Les facteurs qui limitent l'ouverture de la technologie MapReduce sont les brevets associés à cette technologie qui sont des brevets « pledge ». Il est donc pertinent de se concentrer sur l'étude des technologies ayant fait l'objet d'une ouverture grâce à un mécanisme de patent pledge. Comme nous l'avons mentionné au début de ce chapitre, de nombreuses entreprises ont utilisé ce mécanisme sans pour autant renoncer à leur droit de propriété intellectuelle. Avant d'étudier empiriquement l'impact des pledge, la section suivante discute le rôle des pledge à travers la littérature.

3.3 Impact des pledge à travers la littérature

Dans le chapitre 2, nous avons détaillé le principe des patent pledge qui est un engagement unilatéral et volontaire d'une organisation de mettre en accès libre ses brevets (Bronzo, 2022). Un patent pledge n'est pas un acte de renonciation par les organisations de la propriété intellectuelle de leurs brevets, il s'agit d'une volonté des organisations d'amoindrir leur droit de propriété intellectuelle sur leurs technologies. Avec l'émergence de la philosophie sur l'innovation ouverte que nous avons présentée dans la section précédente, de nombreuses organisations ont fait le choix d'ouvrir leurs brevets par un mécanisme de patent pledge afin d'initier « le flux de connaissances internes et externes » porté par les stratégies d'innovation ouverte. Dans cette section, nous allons nous intéresser aux analyses qui ont été réalisées sur les patent pledge à travers deux études de cas sectorielles. La première étude de cas porte sur les technologies vertes et la deuxième sur les véhicules électriques.

3.3.1 Les patent pledge pour promouvoir les technologies vertes

Afin de s'engager pour la protection de l'environnement, la société IBM annonce en 2006 son intention de promouvoir la diffusion des technologies vertes à travers l'initiative « *eco-patent commons* » qui consiste en l'ouverture de brevets permettant d'améliorer ou de protéger directement ou indirectement l'environnement (Reuchlin, 2008³⁰). En janvier 2008, le dispositif est effectivement lancé avec la participation des entreprises IBM, Nokia, Pitney Bowes et Sony. Au démarrage, 36 brevets provenant des quatre entreprises initiatrices du projet ont été intégrés au dispositif. En 2011, ils sont au nombre de 94 brevets issus de 13 entreprises. La date d'intégration au dispositif et le nombre de brevets par entreprise sont présentés dans le tableau 3.1. A sa création, le dispositif « éco-brevets communs » avait pour ambition de promouvoir la diffusion des technologies vertes en accélérant leur adoption et en encourageant les innovations ultérieures (Contreras et al., 2018).

Ce dispositif a cependant pris fin en 2016, faute de résultats probants. Afin de comprendre pourquoi ce dispositif, qui devait encourager la diffusion des technologies vertes, n'a pas été efficace, Contreras, Hall et Helmers (2018) ont mis en place deux types d'analyses distinctes :

³⁰https://www.un.org/esa/sustdev/sdissues/energy/op/beijing_hlccc_nov08/EcoPatentCommons-China_P.Reuchlin.pdf (page 6)

- Une analyse qualitative composée d’entretiens avec des employés des entreprises ayant rejoint l’initiative « éco-brevets communs ».
- Une analyse économétrique basée sur l’étude initiale de Hall & Helmers (2013)

Tableau 3.1 Eco-brevets – date d’intégration et nombre de brevets par entreprise (source : Contreras et al., 2018)

Entreprises	Date d’intégration à « Eco-brevets »	Nombre de brevets pledge ³¹
IBM	14 janvier 2008	29
Nokia	14 janvier 2008	1
Pitney Bowes	14 janvier 2008	2
Sony	14 janvier 2008	4
Bosch	8 septembre 2008	24
DuPont ³²	8 septembre 2008	11
Xerox	8 septembre 2008	13
Taisei	23 mars 2008	2
Ricoh	23 mars 2008	1
Dow	20 octobre 2009	1
Fuji Xerox	20 octobre 2009	2
Hewlett-Packard	1 ^{er} juillet 2010	3
Hitachi ³²	25 juillet 2011	1

Pour réaliser l’analyse qualitative, les auteurs ont identifié des employés des entreprises participantes au dispositif « éco-brevets communs ». Ces employés devaient être impliqués dans le choix de leur entreprise de rejoindre le dispositif. Les auteurs ont ainsi pu interroger sept personnes appartenant à neuf entreprises, ils ont complété leur échantillon avec des employés de la « World Business Council for Sustainable Development » (WBCSD) et de la « Environmental Law Institute » (ELI) également impliqués dans le dispositif. Les questions des différents entretiens concernent :

- Les raisons ayant poussé l’entreprise à rejoindre le dispositif « éco-brevets communs » ;
- La sélection des brevets à partager dans le dispositif ;
- Les objectifs de l’entreprise à travers ce dispositif ayant été atteints.

³¹ Brevets prioritaire (i.e. famille de brevets)

³² DuPont et Hitachi se sont retirés du dispositif « eco-brevets communs » en 2013 à la suite du transfert de direction de la « World Business Council for Sustainable Development » (WBCSD) à la « Environmental Law Institute » (ELI)

- L'engagement de l'entreprise au sein du dispositif ;
- Les forces et les faiblesses du dispositif ;
- La gestion du dispositif « éco-brevets communs » pour les employés de la WBCSD et ELI

L'analyse qualitative a mis évidence des lacunes dans le dispositif « éco-brevets communs ». Six points d'insatisfaction ont été relevés par Contreras et al. :

- *Adhésion et recrutement* : malgré la présence de grands groupes avec des portefeuilles de brevets conséquents, le nombre de brevets au sein du dispositif ne représente qu'une infime part du potentiel des brevets qui pouvaient être partagés (l'entreprise IBM qui est à l'initiative du dispositif n'a partagé que 29 brevets de son portefeuille qui en comporte plus d'une centaine de milliers). Grâce aux échanges avec les employés, il a été relevé que trois points expliquent ce faible partage. Tout d'abord, la difficulté et les coûts nécessaires pour identifier les brevets pertinents susceptibles de rejoindre le dispositif. Ensuite, la conviction de certains membres qu'ils ne disposaient pas de brevets adéquats qui pourraient faire l'objet d'un partage à travers ce dispositif. Enfin certains départements des entreprises, tels que le département juridique ou de propriété intellectuelle, ne concevaient pas de partager des brevets sans compensation financière.
- *Absence de suivi* : l'ensemble des entretiens a mis en lumière le manque de suivi dans l'utilisation des brevets pour favoriser la diffusion des technologies vertes. Ce manque de retour conduit à une incapacité à statuer sur l'efficacité du dispositif. De ce fait, il est difficile pour les entreprises de justifier la poursuite de leur implication dans ce dispositif.
- *Site Web peu convivial* : les employés des entreprises participantes à l'initiative « éco-brevets communs » ont rapporté que le site web du dispositif n'était pas assez intuitif pour rechercher les informations pertinentes sur les brevets « pledge ». Par exemple, les brevets étaient classés par entreprise et non par technologie, ce qui nécessitait des travaux de recherches supplémentaires pour identifier les brevets pertinents.
- *Pas de transfert de technologie* : les répondants aux entretiens ont également pointé du doigt que le dispositif cherchait à favoriser la diffusion des technologies vertes par le seul biais de l'ouverture des brevets. Or, dans ce type de technologie, il est également nécessaire

d'avoir un accompagnement par du transfert de technologies afin de pouvoir comprendre et exploiter les technologies. Un constat déjà mis en avant par Barton, (2004). La seule divulgation des brevets est donc insuffisante pour promouvoir efficacement les technologies vertes.

- *Écart avec la réalité économique des pays en voie de développement* : une des motivations principales du dispositif « éco-brevets communs » était l'accessibilité de ces technologies pour les pays en voie de développement. Cependant, de nombreuses technologies n'étaient pas pertinentes avec la réalité économique de ces pays. Contreras et al. donnent l'exemple d'un brevet IBM sur une technique de nettoyage des tranches de semi-conducteurs qui utilise de l'ozone gazeux pour éliminer des contaminants produits par les processus de nettoyage chimique. Ce brevet a un intérêt principalement dans les pays industrialisés qui possèdent majoritairement les usines de fabrication des semi-conducteurs. De plus, le manque de transfert de technologie mentionné plus haut, impacte également l'exploitation des « éco-brevets » par les pays en voie de développement.

- *Changement dans les priorités des entreprises* : de nombreux répondants ont indiqué que les initiatives en faveur de la responsabilité sociétale des entreprises (RSE) avaient diminué pendant la période d'activité des « éco-brevets ». Cette baisse était due à un manque de budget et de ressources permettant le soutien des actions RSE.

En parallèle de l'analyse qualitative, Contreras et al. ont également mis à jour l'analyse empirique de Hall & Helmers (2013) en actualisant les données utilisées. Cette analyse empirique a pour objectif de quantifier l'impact des « éco-brevets » sur l'incitation à innover dans les technologies vertes. La méthodologie retenue est une estimation économétrique en différence sur le nombre de citations reçues par les « éco-brevets ». Deux groupes ont été mis en place ; d'une part, un groupe traité avec les « éco-brevets » et d'autre part, un groupe contrôle avec des brevets ayant des technologies similaires aux « éco-brevets » (même année de priorité de dépôt, même classe technologique, même autorité de publication). Les auteurs évaluent également si les « éco-brevets » ont reçu significativement plus de citations après la mise sous forme de patent pledge. L'estimation empirique a été réalisée à partir de données issues de la base PATSTAT d'avril 2017.

L'analyse empirique de Contreras et al. met en évidence que les « éco-brevets » sont significativement moins cités que les brevets ayant des technologies similaires que ce soit avant

ou après leur entrée dans le dispositif. Afin de compléter cette analyse économétrique, les auteurs ont également interrogé 71 inventeurs des brevets ayant cité un « éco-brevet » après son ouverture par le mécanisme de patent pledge. Dix inventeurs ont répondu à leurs questions, ce sont des inventeurs qui ont travaillé pour les entreprises IBM, DuPont, Bosch et Xerox, des entreprises participant au dispositif. Il apparaît que 50% des inventeurs ne savaient pas que le brevet qu'ils avaient cité, faisait partie du dispositif « éco-brevet communs ». Les auteurs précisent que bien que l'échantillon des inventeurs ayant répondu au questionnaire était très faible, ce résultat va dans le sens de l'analyse empirique, à savoir l'inefficacité du dispositif « éco-brevets communs » dans la diffusion des technologies vertes.

Au total, l'étude de Contreras et al. (2018) montre que le dispositif « éco-brevets communs » n'a pas permis de favoriser le développement des technologies vertes. Il s'avère que dans le cas des technologies vertes, le caractère peu innovant des technologies a été un élément déterminant dans l'échec. En effet, les technologies qui ont rejoint le dispositif n'étaient pas plus citées avant leur mise sous forme de patent pledge que des brevets ayant des technologies similaires. Ce résultat indique que ces technologies n'étaient donc pas particulièrement innovantes. En ouvrant ces technologies par le biais du dispositif « éco-brevets communs » cette tendance ne s'est pas inversée.

Afin de compléter l'analyse de l'impact des patent pledge, une seconde étude de cas sur les véhicules électriques va être étudiée.

3.3.2 Les patent pledge pour promouvoir les véhicules électriques

En 2014, Elon Musk, président du constructeur automobile Tesla, a déclaré : « *Tous nos brevets vous appartiennent* »³³. A travers cette phrase, Musk a lancé le programme d'ouverture de la propriété intellectuelle de Tesla sur les technologies des véhicules électriques. Dans sa déclaration, Musk indique que la production de véhicules électriques est quasi-inexistante, ce qui ne permet pas de lutter efficacement contre les émissions de CO2 dans l'atmosphère. A travers ce programme d'ouverture de ses brevets, Tesla souhaite donc inciter les autres constructeurs automobiles à accélérer leur programme sur les véhicules électriques afin de ralentir de manière durable les émissions carbonées. Le communiqué indique que dorénavant « *Tesla n'entamera donc aucune poursuite judiciaire pour défendre ses brevets contre ceux qui*

³³ https://www.tesla.com/fr_fr/blog/all-our-patent-are-belong-you

veulent, de toute bonne foi, utiliser notre technologie ». L'initiative de Tesla n'est pas restée dans le secteur automobile. En effet, en janvier 2015, le constructeur japonais Toyota annonce que plus de 5600 de ses brevets sur les technologies de véhicules à hydrogène seront également accessibles librement³⁴. A l'instar de Tesla, le constructeur japonais indique dans son communiqué que cette initiative « *a pour objectif d'inciter d'autres constructeurs automobiles à construire des véhicules à hydrogène mais également aider les fournisseurs de carburant à ouvrir de nouvelles stations de ravitaillement en hydrogène* »³⁴.

Dans leur étude sur les patent pledge de Tesla, Baudry & Dumont (2020) étudient la stratégie de propriété intellectuelle de la firme californienne. Afin d'évaluer la décision de Tesla d'adopter une stratégie de propriété intellectuelle ouverte à travers des patent pledge, Baudry & Dumont réalisent une analyse complète des portefeuilles de brevets de Tesla et de ses concurrents en s'appuyant sur la base PATSTAT sur la période 2006 (date du premier dépôt de brevet Tesla) à 2018. A travers l'étude des brevets Tesla, Baudry & Dumont ont pu mettre en évidence plusieurs points spécifiques.

Tout d'abord, les auteurs constatent que la taille du portefeuille de brevets Tesla est petite au regard de la taille des portefeuilles des autres constructeurs automobiles, et cela même si on limite l'analyse aux classes technologiques ciblées par les brevets de Tesla. Ils constatent même que certains constructeurs chinois, aussi âgés que Tesla, possèdent un portefeuille de brevets plus important indiquant que le constructeur californien ne fonde pas sa stratégie d'entreprise sur une grande variété d'innovations.

Ensuite, Baudry & Dumont étudient l'impact de l'annonce de l'ouverture de la propriété intellectuelle de Tesla en 2014 sur sa stratégie de dépôt de brevets. Pour ce faire, ils étudient le flux de nouvelles inventions dans le portefeuille des familles de Tesla et de ses concurrents. Pour Tesla, ce flux augmente jusqu'en 2012 puis baisse, et pour les concurrents de Tesla, on constate la même tendance à la baisse à partir de 2013. Le flux de nouvelles inventions pour Tesla et ses concurrents étant proche, les auteurs concluent que l'annonce de l'ouverture de la propriété intellectuelle de Tesla en 2014 n'a pas modifié significativement la stratégie d'entreprise de Tesla par rapport à ses concurrents. Par ailleurs, Baudry & Dumont ont également constaté en étudiant la taille des familles de brevets que Tesla protège fortement ses brevets à l'international, une protection comparable à celle des constructeurs tels que Toyota,

³⁴ <https://pressroom.toyota.com/toyota-fuel-cell-patents-ces-2015/>

Hyundai ou Audi. Ce constat indique une ambition internationale du constructeur californien, ambition similaire à des constructeurs historiques du secteur de l'automobile.

Enfin, concernant les coopérations de Tesla en matière de R&D, Baudry & Dumont montrent que les co-brevets peuvent être considérés comme un proxy des coopérations sur des projets de R&D en phase de lancement. Pour évaluer ce type de coopération chez les constructeurs automobiles, deux indicateurs sont croisés : la proportion de familles de brevets contenant au moins un co-brevet et la proportion moyenne de co-déposants dans les co-dépôts. Le croisement de ces deux indicateurs montre que Tesla est un des constructeurs automobiles les moins enclins à coopérer sur des projets de R&D en phase de lancement. Les auteurs expliquent qu'un positionnement atypique de Tesla en matière d'innovation pourrait expliquer ce résultat. Pour vérifier cette hypothèse, ils utilisent un indice de similitude technologique pour chaque constructeur. Plus cet indice est proche de 1 et plus deux portefeuilles de brevets sont proches. Ainsi, les auteurs mettent en évidence que les brevets de Tesla se rapprochent technologiquement de ceux des constructeurs Renault, PSA, Honda, Mitsubishi, Hyundai et Greatwall (un constructeur chinois). Parmi ces constructeurs, certains coopèrent fortement dans des projets R&D au regard de leurs co-brevets (notamment Mitsubishi, Honda, Hyundai) alors que leur indice de similitude technologique est proche de celui de Tesla. Les auteurs concluent donc que ce n'est pas un positionnement atypique en matière d'innovation qui explique le niveau faible de coopération en R&D de Tesla.

Enfin, Baudry & Dumont examinent si la stratégie d'ouverture par les patent pledge n'a pas pour objectif d'affirmer le leadership technologique de Tesla. Les auteurs rappellent qu'une des motivations premières dans le choix d'ouvrir ses technologies à travers un pledge est la volonté d'imposer un standard technologique à travers un potentiel effet de réseau et d'échelle (point également abordé dans la section 2.2.2 du chapitre 2). Ils indiquent également que la production de véhicules électriques s'appuie justement sur ces deux effets. D'une part, les bornes de recharges nécessaires au chargement des véhicules électriques vont induire des effets de réseau par l'augmentation de leur nombre sur un territoire. D'autre part, le coût de production des batteries va baisser avec l'augmentation du volume de production des véhicules électriques d'où des économies d'échelle. Les auteurs précisent également que pour générer effectivement ces effets de réseau et d'échelle, il faut que les technologies ouvertes par les patent pledge aient un réel intérêt technologique pour les autres constructeurs automobile. Afin de vérifier ce point, Baudry & Dumont vont s'intéresser au nombre de citations (sans autocitation) reçues par les brevets de Tesla afin de vérifier si ce nombre est significativement plus élevé pour le

constructeur californien par rapport aux autres constructeurs du secteur. Il s'avère que les brevets de Tesla sont nettement plus cités que ceux de ses concurrents. Tesla possède donc moins de brevets que ses concurrents mais ceux-ci sont plus innovants compte tenu du nombre de citations reçues.

Au total, Baudry & Dumont concluent leur étude en indiquant que la stratégie des patent pledge de Tesla a pour but d'affirmer le leadership technologique du constructeur californien et de le transformer en leadership de marché, un objectif en accord avec la stratégie basée sur un modèle d'intégration verticale des systèmes de recharge et des batteries. La firme californienne, en ouvrant ses technologies par le mécanisme des patent pledge, a donc souhaité imposer un standard technologique à ses concurrents et bénéficier des effets de réseau et d'échelle qui en résultent.

Les études de cas sur les technologies vertes et les véhicules électriques ont mis en évidence qu'un mécanisme d'ouverture de la propriété intellectuelle grâce à des patent pledge pouvait avoir un impact différent selon les technologies considérées. D'une part, le dispositif « éco-brevets communs » s'est avéré inefficace pour promouvoir les technologies vertes, le dispositif s'étant d'ailleurs arrêté en 2016. D'autre part, pour l'entreprise Tesla, le mécanisme des patent pledge a permis d'affirmer le leadership technologique de la firme californienne. Dans ces deux études de cas, c'est l'innovation portée par les technologies qui a joué un rôle central dans le succès ou non du dispositif d'ouverture par les patent pledge. En effet, Contreras et al. (2018) ont mis en évidence que les technologies ayant rejoint le dispositif « éco-brevets communs » n'étaient pas plus citées avant leur mise sous forme de patent pledge que des brevets ayant des technologies similaires. Ces technologies n'étaient donc pas particulièrement innovantes. Inversement, Baudry & Dumont (2020) ont montré que même si le portefeuille de brevets de Tesla est moins important que ceux de ses concurrents, les brevets de la firme californienne sont plus cités que les brevets des autres constructeurs. Les technologies Tesla sont donc particulièrement innovantes.

Afin de déterminer si la technologie MapReduce a impacté le secteur des technologies Big Data du fait de son caractère innovant intrinsèque et non uniquement à cause de son ouverture, nous allons analyser la propriété intellectuelle de trois acteurs du secteur des TIC, Google, IBM et Microsoft afin d'évaluer l'impact sectoriel de leur patent pledge, en incluant évidemment la technologie MapReduce.

3.4 Stratégie d'analyse empirique des patent pledge : le cas de trois acteurs des TIC

3.4.1 Modèles d'estimation et hypothèses

L'analyse empirique de ce chapitre va évaluer l'impact des patents pledge sur un périmètre plus large que celui du secteur des technologies Big Data. L'objectif est de tester **l'impact des patents pledge (incluant la technologie MapReduce) sur la valeur privée et sociale des brevets** inclus dans le portefeuille de trois acteurs majeurs des TIC que sont **Google, Microsoft et IBM**.

En termes de méthodologie, il s'agit d'estimer la valeur des pledge par rapport aux non pledge au sein de deux cadres complémentaires :

1. Comparer la valeur des brevets pledge à celle des brevets non pledge et d'un groupe de contrôle de brevets ayant des caractéristiques technologiques semblables (même date de priorité et même technologie).
2. Comparer la valeur des brevets pledge à celle des brevets non pledge au sein des portefeuilles des entreprises.

Pour définir les hypothèses à tester dans ce chapitre, il importe de repartir des résultats du chapitre 2 et de la littérature empirique sur les pledge. Les résultats du chapitre 2 ont mis en évidence que l'ouverture des brevets MapReduce avait eu un impact positif sur les follow-on innovations, en revanche la valeur privée et sociale du pledge a subi un impact négatif une fois que le pledge a été octroyé. L'impact négatif des pledge sur la valeur sociale est également confirmé par la littérature sur les « eco-patent ». La question qui se pose à présent est de savoir si ce résultat peut être généralisé à l'ensemble des pledge dans le domaine des TIC des entreprises Google, Microsoft et IBM.

La première hypothèse s'appuie sur le chapitre 2 et se présente comme suit :

Le chapitre 2 a montré qu'en se comparant à des brevets de contrôle, la valeur privée et sociale des brevets pledge était plus faible, il est donc possible de proposer l'hypothèse suivante :

Hypothèse 1 : les brevets pledge ont une valeur privée et sociale plus faible que celle de technologies similaires non pledge

Le mécanisme a été exposé précédemment : les brevets ont pour objectif de créer un standard technologique mais la qualité des technologies sous-jacente n'est pas forcément à la hauteur (cas des eco-patent) et par ailleurs une fois le brevet ouvert, les développements technologiques ont tendance à se ralentir.

Toutefois une seconde question se pose. En effet, les pledge peuvent avoir une valeur pour l'entreprise qui les détient. Comme l'entreprise cherche à créer un standard, elle a tendance à s'appuyer sur ces technologies pour ses développements et, ainsi, on peut s'attendre à ce que la valeur privée soit plus élevée par rapport aux autres brevets du portefeuille.

Hypothèse 2 : les brevets pledge ont une valeur privée plus élevée que les autres brevets du portefeuille.

Le modèle utiliser pour tester les hypothèses est semblable à celui du chapitre précédent.

$$CITATIONS_{i,t} = \beta_0 + \beta_P PLEDGE_i + \beta_X X_i + \beta_t + \delta_{t-appAnnée} + \epsilon_{i,t}$$

$CITATIONS_{i,t}$ représente le nombre de citations reçues par le brevet i (pledge ou non pledge) à l'année t .

A l'image des estimations du chapitre 2, nous distinguons les citations avec autocitations et sans autocitation afin de différencier l'impact des brevets en termes de valeur privée et sociale. $PLEDGE_i$ est une indicatrice égale à 1 si le brevet est un brevet pledge.

Enfin, nous intégrons un vecteur X_i représentant les variables qui permettent de contrôler les variations du nombre de citations reçues (nombre d'inventeurs, qualité technologique du brevet...)

3.4.1 Constitution des deux groupes de travail

L'échantillon considéré pour ce chapitre extrait de l'office USPTO l'ensemble des brevets des entreprises Google, IBM et Microsoft sur la période 1994-2017. L'échantillon identifie tous les brevets « non pledge » et les brevets pledge. La liste des brevets pledge pour chacune des entreprises a été collectée à travers les sources suivantes :

- Google : site institutionnel de l'entreprise³⁵
- IBM : site d'informations technologiques eweek.com (Galli, 2005)

³⁵ <https://www.google.com/patents/opnpledge/patents/>

- Microsoft : site institutionnel de l'entreprise³⁶

L'analyse des données collectées a mis en évidence que certains brevets pledge n'ont pas été déposés initialement par le propriétaire actuel du brevet. Ce point n'est pas surprenant car ces trois entreprises font régulièrement des acquisitions de propriété intellectuelle (rachat de brevets, de sociétés...). Par exemple, Google a racheté plus de 1000 brevets à IBM en 2011 (Lagane, 2011). Il semble donc pertinent de considérer l'origine des pledge et d'identifier ceux qui ont été développés en interne et déposés par l'entreprise des autres brevets.

La base de données finale s'appuie sur les 101 879 brevets des portefeuilles de Google, IBM et Microsoft. Le tableau suivant présente la répartition des brevets à travers les entreprises.

Tableau 3.2 : Périmètre des brevets Google, IBM et Microsoft ayant au moins un brevet de contrôle

	Nombre de brevets	Nombre de brevets ayant au moins un contrôle	Pourcentage de complétude
Brevets Google	10 997	7 360	66.9%
Brevets Pledge	101	86	85.1%
Brevets Pledge déposés initialement par Google	18	15	83.3%
Brevets Pledge déposés initialement par Google (MapReduce)	7	7	100.0%
Brevets Pledge déposés initialement par Google (hors MapReduce)	11	8	72.7%
Brevets Pledge non déposés initialement par Google	83	71	85.5%
Brevets IBM	67 202	54 286	80.8%
Brevets Pledge	402	342	85.1%
Brevets Pledge déposés initialement par IBM	371	317	85.4%
Brevets Pledge non déposés initialement par IBM	31	25	80.6%
Brevets Microsoft	23 680	18 550	78.3%
Brevets Pledge	211	155	73.5%
Brevets Pledge déposés initialement par Microsoft	210	154	73.3%
Brevets Pledge non déposés initialement par Microsoft	1	1	100.0%
Total	101 879	80 196	78.7%

Notre échantillon de données est constitué majoritairement de brevets IBM (66% du total). Ce point n'est pas surprenant, IBM étant une firme très active dans le domaine de la propriété intellectuelle (Beky, 2012). Les brevets pledge représentent une très faible part de notre échantillon, moins de 1% des brevets, ce pourcentage est globalement similaire pour les trois entreprises de l'échantillon : pour Google, les brevets pledge représentent 0.92% de ses brevets, ce pourcentage est de 0.60% et 0.89% respectivement pour IBM et Microsoft. Le nombre de

³⁶ <https://msopenspecs.azureedge.net/files/public-patents-export/patmap-tool.html>

brevets pledge non déposés initialement par l'entreprise est également répertorié. Il est de 8% pour IBM et quasi-nul pour Microsoft et de 83% pour Google. Cette différence entre Google d'un côté et IBM et Microsoft de l'autre n'est pas étonnante compte tenu de l'ancienneté de chacune des entreprises. En effet, IBM a été fondé en 1911, Microsoft en 1975 et enfin Google en 1998. Par leur ancienneté dans le secteur des TIC, IBM et Microsoft ont construit graduellement leur propriété intellectuelle au fil du temps. Google, pour se mettre à niveau technologiquement, a fait plus d'acquisitions. Il est intéressant de noter que parmi les technologies développées par Google et qui font l'objet d'un patent pledge, la technologie MapReduce représente à elle seule 39% de ces brevets, ce qui constitue une part assez importante.

Les contrôles ont été sélectionnés en utilisant les mêmes critères qu'au chapitre 2 à savoir même année de priorité et même combinaison de classes technologiques (CPC à 5 caractères) que le brevet traité. Concernant les contrôles trouvés pour les brevets des trois entreprises, environ 79% de leurs brevets ont au moins un brevet de contrôle. Ce chiffre est légèrement plus faible pour l'entreprise Google (environ 67%) mais les brevets pledge de la firme ont un taux de complétude plus élevé à 85%. Pour l'entreprise IBM, environ 81% des brevets ont au moins un brevet de contrôle, ce taux s'élève à 85% pour les brevets pledge. Enfin pour Microsoft, 78% des brevets du périmètre initial ont au moins un brevet de contrôle, ce taux baisse légèrement à 73% pour les brevets pledge.

Au total, **deux échantillons sont utilisés pour les régressions** :

- un échantillon de 1 355 856 brevets qui comprend les brevets des trois entreprises pledge et non pledge ayant un contrôle (80 196) et leurs contrôles (1 275 660) et
- un échantillon constitué des portefeuilles complets des trois entreprises (sans contrôle) et comprenant 101878 brevets.

La suite de cette section présente les brevets de chacune des trois entreprises étudiées à travers trois indicateurs : le nombre de brevets déposés par année, leurs principales technologies et enfin le nombre de citations reçues sur les 5 premières années suivant la publication du brevet.

3.4.2 Description des portefeuilles des brevets de Google, IBM et Microsoft

- **Nombre de brevets déposés par année**

Ces trois Figures 3.4 à 3.6 confirment la faible part des brevets pledge dans la propriété intellectuelle des entreprises. Les figures mettent en évidence la position atypique de Google par rapport à IBM et Microsoft. En effet, la majorité des brevets Google dont l'année de dépôt est antérieure à 2004 sont des brevets pledge qui n'ont pas été développés initialement par Google. Il apparaît qu'entre sa création (1998) et 2004, Google a acquis un certain nombre de brevets pour ses développements internes et qu'il les a par la suite ouverts grâce au mécanisme des patent pledge. A partir de 2004, on constate que le nombre de dépôts de brevets croît graduellement pour atteindre un pic en 2012, majoritairement constitué de brevets non pledge. On peut donc raisonnablement penser que les brevets pledge acquis par Google lors de ses premières années d'existence ont permis à la firme de construire sa propriété intellectuelle sur les années suivantes.

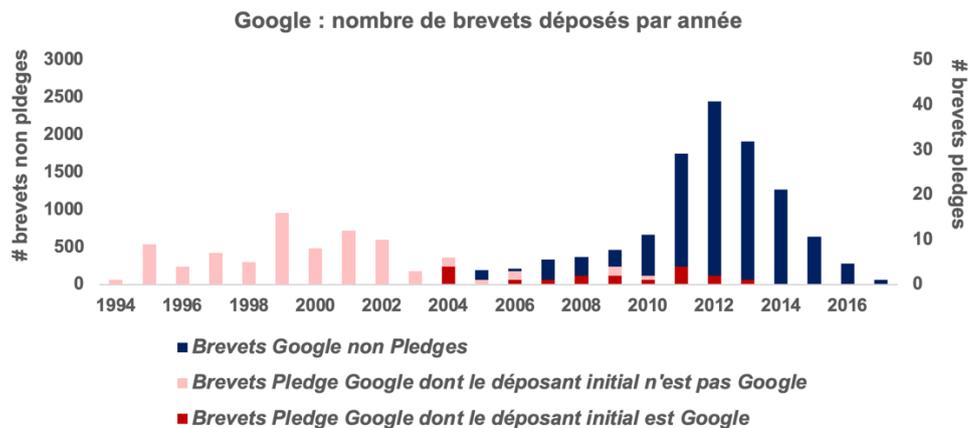


Figure 3.4 nombre de brevets déposés par Google entre 1994 et 2017 (source : USPTO)

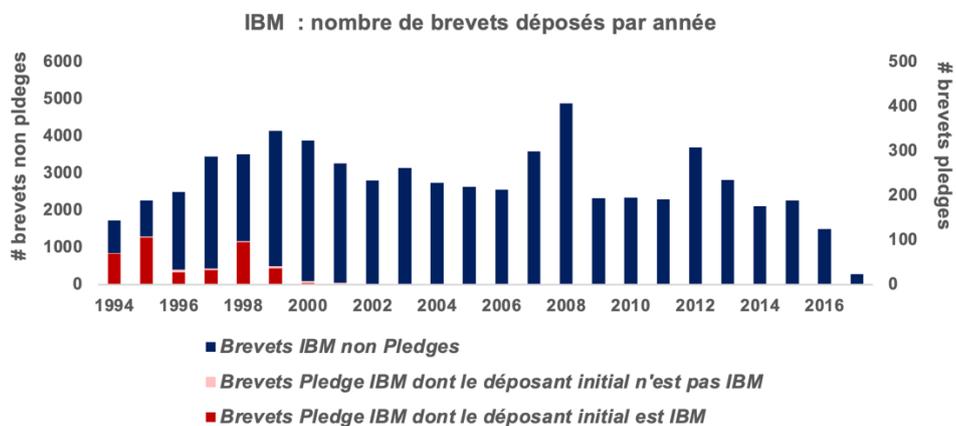


Figure 3.5 nombre de brevets déposés par IBM entre 1994 et 2017 (source : USPTO)

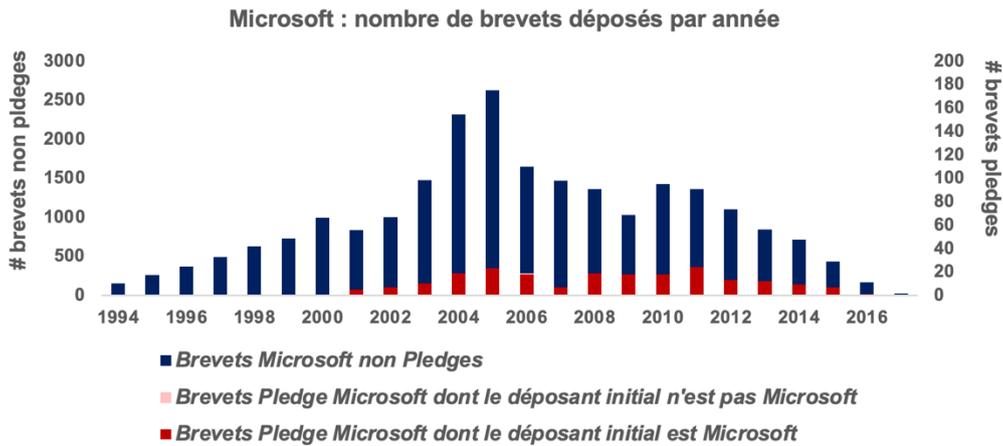


Figure 3.6 nombre de brevets déposés par Microsoft entre 1994 et 2017 (source : USPTO)

Pour l'entreprise IBM, on constate que la majorité des brevets pledge ont été déposés avant les années 2000. Le nombre de brevets déposés par an est assez constant sur la période étudiée, même si on constate un ralentissement des dépôts après le pic de 2008.

Enfin pour Microsoft, on peut voir qu'à la différence des deux autres sociétés, l'entreprise a régulièrement déposé des brevets pledge entre 2003 et 2013, des brevets développés en quasi-totalité par la firme elle-même. Le nombre de brevets déposés par Microsoft progresse graduellement de 1994 à 2005 puis ce nombre ralentit sur les années suivantes.

▪ Principales technologies des brevets (classification WIPO)

Les principales technologies des brevets Google, IBM et Microsoft sont détaillées dans les graphiques suivants (Figure 3.7 à 3.9). Le top 3 des technologies est systématiquement indiqué pour chacune des entreprises.

Les technologies des trois entreprises sont globalement similaires à une exception près pour IBM qui a la technologie « semiconducteur » dans son top 3 là où Google et Microsoft ont la technologie « Méthodes de traitement des données à des fins de gestion ». Cette différence s'explique par le fait qu'historiquement IBM était un fabricant de matériel informatique qui s'est peu à peu orienté vers des activités de logiciels et de service informatique. Pour les trois entreprises, le domaine majeur reste celui de l'informatique, qui représente environ 50% des domaines technologiques de leurs brevets. Si on s'intéresse plus particulièrement aux brevets pledge, pour Google et Microsoft ce sont les domaines « informatique » et « communication

numérique » qui sont majoritairement représentés alors que pour IBM les brevets pledge sont concentrés très majoritairement dans le domaine de l’informatique.

Au total, on constate que la propriété intellectuelle de ces trois entreprises est assez semblable, majoritairement concentrée dans le domaine de l’informatique.

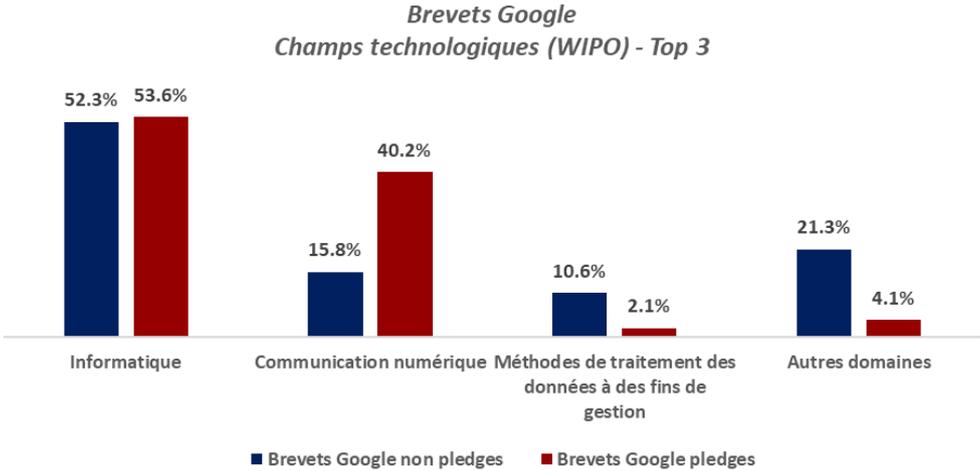


Figure 3.7 Top 3 des technologies (WIPO) des brevets Google entre 1994 et 2017 (source : USPTO)

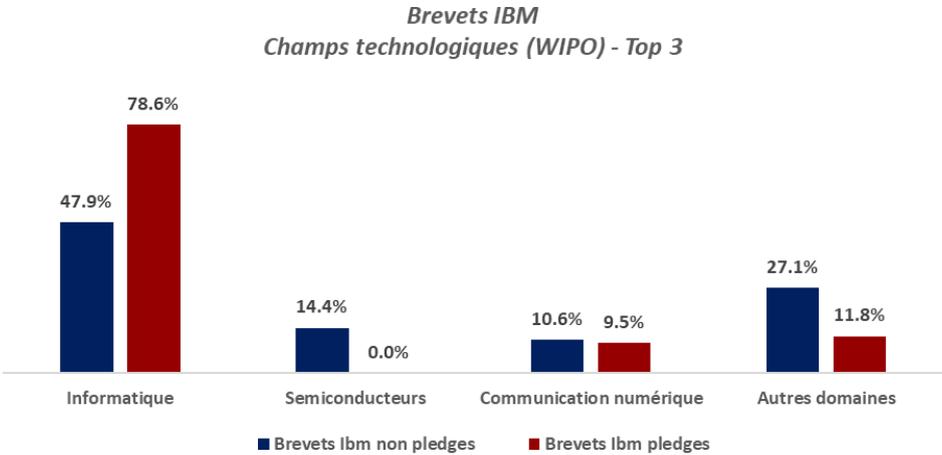


Figure 3.8 Top 3 des technologies (WIPO) des brevets IBM entre 1994 et 2017 (source : USPTO)

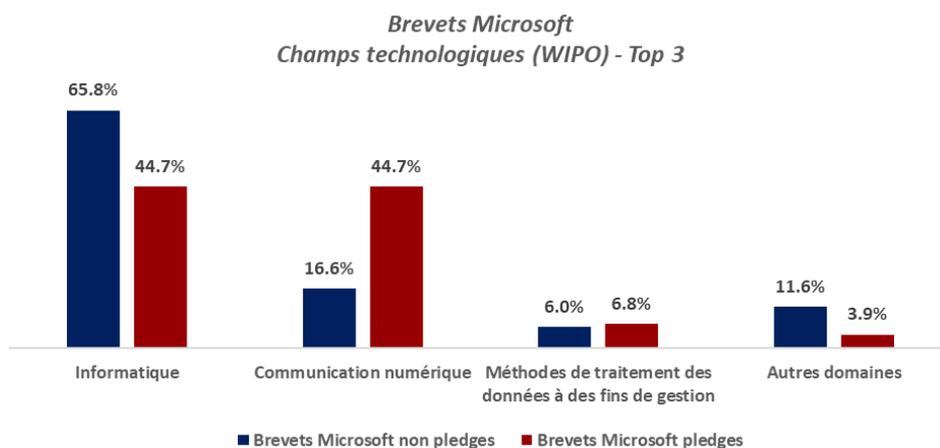


Figure 3.9 Top 3 des technologies (WIPO) des brevets Microsoft entre 1994 et 2017 (source : USPTO)

▪ **Nombre de citations reçues sur les 5 premières années après le dépôt du brevet**

Les Figures de 3.10 à 3.15 détaillent le nombre de citations reçues sur les 5 premières années après la publication du brevet pour les brevets de Google, IBM et Microsoft.

Pour chaque entreprise, il y a une Figure sur les brevets non pledge et une sur les brevets pledge. Chaque figure distingue le nombre de citations avec et sans autocitation.

On constate des périodes de fortes citations pour les brevets Google suivant la nature du brevet. Pour les brevets non pledge, la période de forte citation se situe entre 2010 et 2015, à savoir la période de fort dépôt de ses brevets par Google. Pour les brevets pledge, on distingue deux périodes en fonction des autocitations. Si on tient compte des autocitations, les brevets pledge sont fortement cités sur la période 1994-2004, période de dépôt des brevets pledge non développés initialement par Google. Google a donc fortement cité ces brevets pledge pour son développement interne. Si on exclut les autocitations, on constate un pic de citations des brevets pledge sur la période 2010-2013 là où l'entreprise Google a déposé un nombre plus important de brevets pledge développés en interne. Sur cette période 2010-2013, ses brevets pledge développés par Google ont donc été plus cités par les acteurs du secteur des TIC.

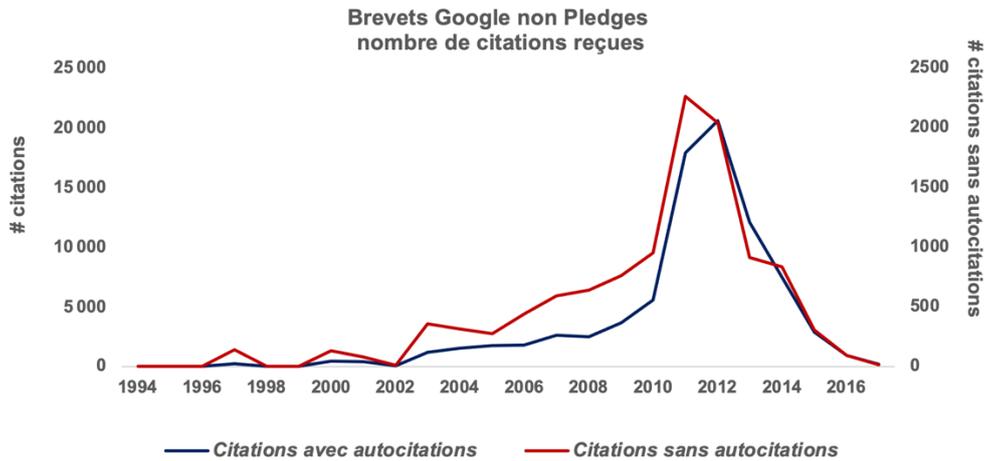


Figure 3.10 : nombre de citations reçues par les brevets Google non pledge entre 1994 et 2017 – agrégés sur les 5 années après la publication du brevet (source : USPTO)

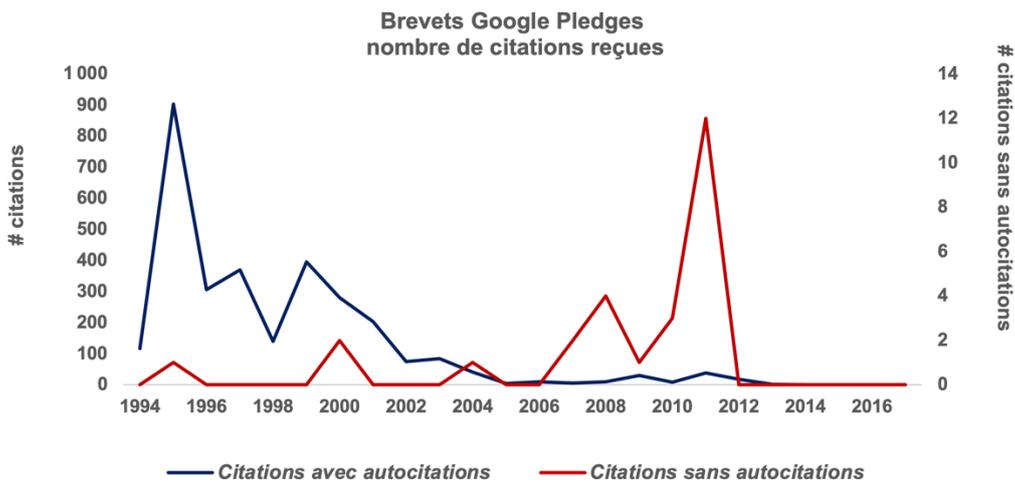


Figure 3.11 nombre de citations reçues par les brevets Google pledge entre 1994 et 2017 - agrégés sur les 5 années après la publication du brevet (source : USPTO)

Pour les brevets IBM, on constate que le nombre de citations décroît graduellement sur la période analysée quelque que soit la nature du brevet. Pour les brevets non pledge, la période de fortes citations se situe entre 1994 et 2000. A partir de cette date, le nombre de citations des brevets non pledge IBM ne cesse de décroître avec néanmoins des pics de citations sur les années 2008 et 2012 correspondants à des pics de dépôts. Pour les brevets pledge IBM, ils sont

cités exclusivement sur la période 1994-2002, c'est-à-dire sur la période de leur dépôt. Les brevets pledge IBM ont donc un impact très spécifique dans le temps.

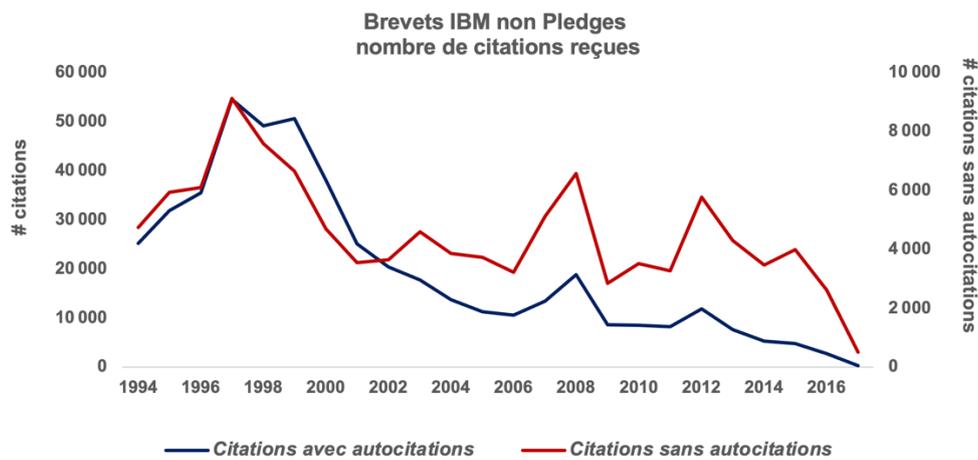


Figure 3.12 nombre de citations reçues par les brevets IBM non pledge entre 1994 et 2017 - agrégés sur les 5 années après la publication du brevet (source : USPTO)

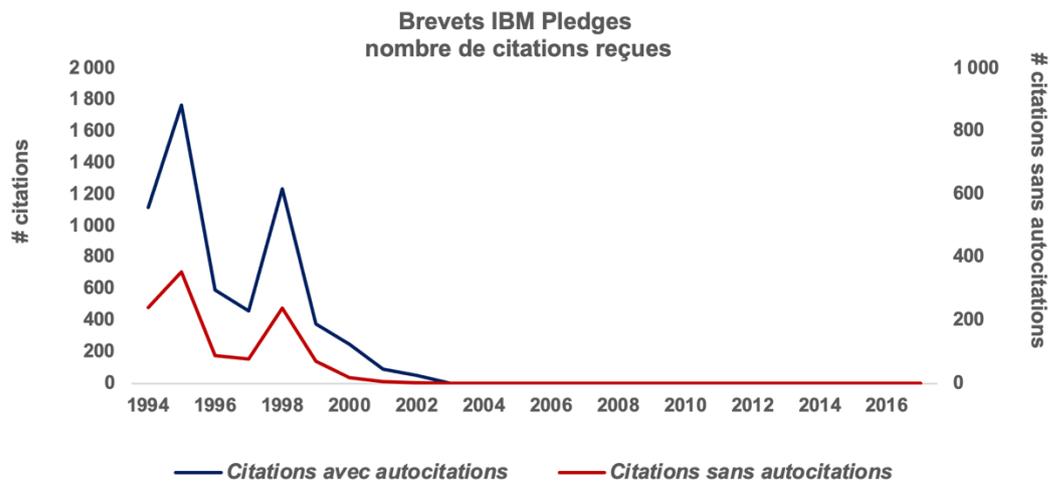


Figure 3.13 nombre de citations reçues par les brevets IBM pledge entre 1994 et 2017 - agrégés sur les 5 années après la publication du brevet (source : USPTO)

Pour l'entreprise Microsoft, le nombre de citations reçues par leurs brevets non pledge progresse fortement jusqu'à l'année 2004, ensuite ce nombre décroît graduellement. Les brevets pledge de la firme sont majoritairement cités sur la période 2003-2013 juste après leur publication.

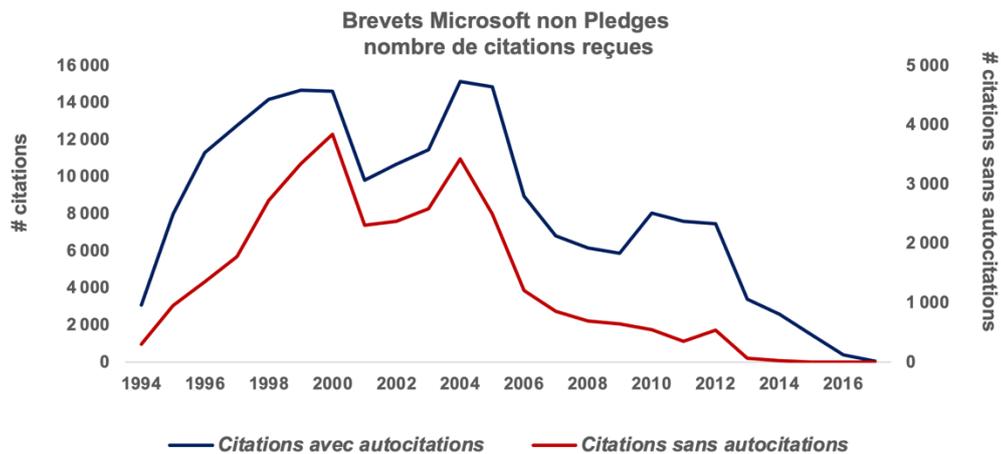


Figure 3.14 nombre de citations reçues par les brevets Microsoft non pledge entre 1994 et 2017 - agrégés sur les 5 années après la publication du brevet (source : USPTO)

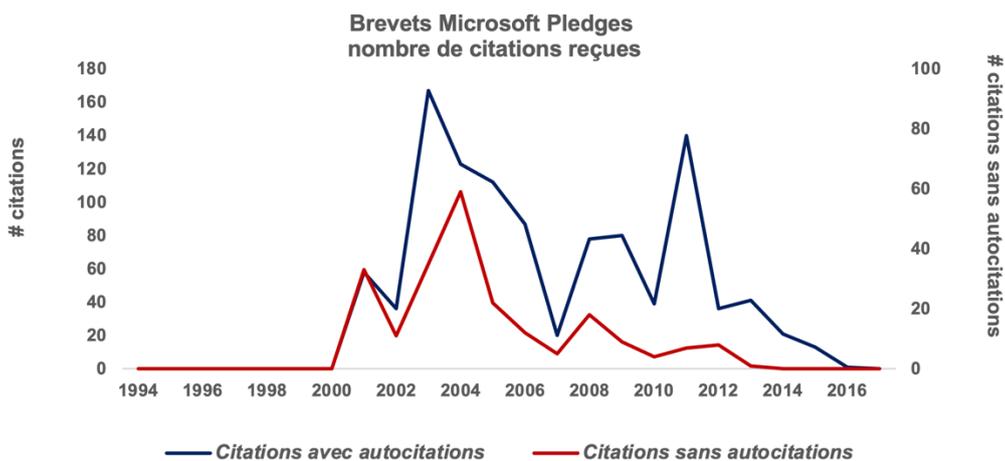


Figure 3.15 nombre de citations reçues par les brevets Microsoft pledge entre 1994 et 2017 - agrégés sur les 5 années après la publication du brevet (source : USPTO)

3.5 Estimations agrégées avec groupe de contrôle

3.5.1 Variables et statistiques descriptives

L'analyse se situe, comme au chapitre 2, au niveau du brevet cité et agrège les citations reçues au cours des 5 années suivant la publication.

Variables dépendantes

- **Nombre de citations de brevets reçues (5 ans)** est le nombre de citations reçues par le brevet dans une fenêtre de 5 années après la publication.
- **Nombre de citations de brevets reçues sans autocitation (5 ans)** est le nombre de citations reçues par le brevet dans une fenêtre de 5 années après la publication sans les autocitations définies au niveau des déposants.

Variables d'intérêts

- **Brevets (groupe traité)** : indicatrice qui prend la valeur 1 si le brevet appartient au portefeuille des trois entreprises (Google, IBM et Microsoft) et 0 sinon ; les contrôles sont les brevets ayant les mêmes caractéristiques technologiques et même année de dépôt que les brevets traités.
- **Brevets Pledge** : indicatrice qui prend la valeur 1 si le brevet est un pledge dans le portefeuille d'une des trois entreprises étudiées (Google, IBM et Microsoft) et 0 sinon
- **Brevets Pledge déposés initialement** : indicatrice qui prend la valeur 1 si le pledge a été déposé par l'entreprise qui en a la propriété et 0 sinon
- **Brevets Pledge non déposés initialement** : indicatrice qui prend la valeur 1 si le pledge a été déposé initialement par une autre organisation que l'une des trois entreprises considérées et 0 sinon
- **Brevets Pledge Google** : indicatrice qui prend la valeur 1 si le brevet est un pledge appartenant à Google et 0 sinon
- **Brevets Pledge MapReduce Google** : indicatrice qui prend la valeur 1 si le brevet est un pledge MapReduce et 0 sinon
- **Brevets Pledge déposés initialement par Google hors MapReduce** : indicatrice qui prend la valeur 1 si le pledge a été déposé initialement par Google et 0 sinon

- **Brevets Pledge non déposés initialement par Google** : indicatrice qui prend la valeur 1 si le pledge a été déposé initialement par un autre déposant que Google
- **Brevets Pledge IBM** : indicatrice qui prend la valeur 1 si le brevet est un pledge appartenant à la société IBM et 0 sinon
- **Brevets Pledge déposés initialement par IBM** : indicatrice qui prend la valeur 1 si le brevet est un pledge déposé initialement par IBM et 0 sinon
- **Brevets non déposés initialement par IBM** : indicatrice qui prend la valeur 1 si le brevet est un pledge déposé initialement par un autre déposant que IBM et 0 sinon
- **Brevets Pledge Microsoft** : indicatrice qui prend la valeur 1 si le brevet est un pledge appartenant à la société Microsoft et 0 sinon.

Variables de contrôle

Les variables de contrôle sont les mêmes qu’au chapitre précédent. Nous rappelons les définitions.

- **Co-brevets (Déposants)** représentent les brevets qui ont plusieurs déposants. Ils apparaissent plusieurs fois dans les observations.
- **Nombre d’inventeurs** listés sur le document du brevet est une mesure de la valeur des brevets, il indique aussi le montant des ressources investies dans l’invention (Sapsalis et al., 2006).
- **Nombre de sous-groupes CPC** caractérisant le brevet est un proxy de l’étendue technologique du brevet est un indicateur de complexité technologique.
- **Taille de la famille** est un proxy de la valeur de l’invention (Lanjouw et al., 1998; Harhoff et al., 2003).
- **Le nombre de citations antérieures** (*Backward citations*) (Lanjouw & Schankerman, 2001)
- **Le nombre de citations hors brevets**, ce sont les références ajoutées à un brevet, elles sont considérées comme un proxy de la proximité à la science (Narin et al., 1997)
- **Le nombre de revendications** aussi appelées *claims* dans la littérature (Lanjouw & Schankerman, 2001)

3.5.2 Statistiques descriptives

Le tableau 3.3 présente les statistiques descriptives pour l'ensemble des brevets traités et des contrôles. Les brevets traités qui représentent les brevets des trois entreprises représentent 5,9% de l'ensemble des brevets, en revanche les pledge représentent une part très faible dans l'échantillon complet.

Tableau 3.3 Statistiques descriptives – Brevets Google, IBM, Microsoft et contrôles				
	Moyenne	Ecart-type	Min	Max
Nombre de citations de brevets reçues (5 ans)	6.037	16.220	0.000	2846.000
Nombre de citations de brevets reçues sans autocitation (5 ans)	4.863	14.496	0.000	2522.000
Pledge	0.000	0.021	0.000	1.000
Brevets (groupe traité)	0.059	0.236	0.000	1.000
Brevets Pledge IBM	0.000	0.016	0.000	1.000
Brevets Pledge déposés initialement par IBM	0.000	0.015	0.000	1.000
Brevets non déposés initialement par IBM	0.000	0.004	0.000	1.000
Brevets Pledge Google	0.000	0.008	0.000	1.000
Brevets Pledge MapReduce Google	0.000	0.002	0.000	1.000
Brevets Pledge déposés initialement par Google	0.000	0.003	0.000	1.000
Brevets Pledge déposés initialement par Google hors MapReduce	0.000	0.002	0.000	1.000
Brevets Pledge non déposés initialement par Google	0.000	0.007	0.000	1.000
Brevets Pledge Microsoft	0.000	0.011	0.000	1.000
Brevets non déposés initialement par Microsoft	0.000	0.001	0.000	1.000
Co-brevets	0.024	0.154	0.000	1.000
Nombre de sous-groupes CPC	4.134	4.954	1.000	190.000
Nombre d'inventeurs	2.592	1.768	1.000	76.000
Taille de la famille	2.569	2.289	0.000	41.000
Nombre de citations antérieures	21.382	47.959	0.000	4891.000
Nombre de citation hors brevets	5.105	18.232	0.000	1413.000
Nombre de revendications (claims)	18.314	11.373	0.000	760.000
Observations	1 355 856			

Le tableau 3.4. présente les tests de comparaison des moyennes des observations traitées et des contrôles. Il apparaît à première vue que les traités reçoivent plus de citations dans l'absolu mais largement moins de citations que les contrôles ayant les mêmes caractéristiques lorsque les autocitations ne sont pas prises en compte. Ceci indique que les contrôles ont une valeur technologique et sans doute économique plus élevée que les brevets des portefeuilles de Google, IBM et Microsoft. Les traités sont aussi plus complexes (plus de sous-groupes CPC), une famille plus grande plus de citations antérieures. Mais ils ont moins d'inventeurs et moins de revendications.

Tableau 3.4 Tests de comparaison des moyennes

	Traités		Contrôles	
	Moyennes		Différences	Ecart-types
Nombre de citations de brevets reçues (5 ans)	7.216	5.962	-1.254***	0.059
Nombre de citations de brevets reçues sans autocitation (5 ans)	1.524	5.073	3.549***	0.053
Pledge	0.007	0.000	-0.007***	0.000
Brevets Pledge IBM	0.004	0.000	-0.004***	0.000
Brevets Pledge déposés initialement par IBM	0.004	0.000	-0.004***	0.000
Brevets Pledge non déposés initialement par IBM	0.000	0.000	-0.000***	0.000
Brevets Pledge Google	0.001	0.000	-0.001***	0.000
Brevets Pledge MapReduce Google	0.000	0.000	-0.000***	0.000
Brevets Pledge déposés initialement par Google	0.000	0.000	-0.000***	0.000
Brevets Pledge non déposés initialement par Google	0.001	0.000	-0.001***	0.000
Brevets Pledge déposés initialement par Google hors MapReduce	0.000	0.000	-0.000***	0.000
Brevets Pledge Microsoft	0.002	0.000	-0.002***	0.000
Brevets non déposés initialement par Microsoft	0.000	0.000	-0.000***	0.000
Co-brevets	0.019	0.025	0.006***	0.001
Nombre de sous-groupes CPC	4.034	4.141	0.107***	0.018
Nombre d'inventeurs	3.319	2.547	-0.772***	0.006
Taille de la famille	2.010	2.604	0.595***	0.008
Nombre de citations antérieures	20.969	21.408	0.439**	0.175
Nombre de citation hors brevets	7.040	4.984	-2.056***	0.066
Nombre de revendications	18.497	18.303	-0.194***	0.041
Observations	80196	1 275 660	1 355 856	

3.5.3 Résultats des estimations avec contrôles

Les trois tableaux suivants 3.5 à 3.7 présentent les résultats des estimations de la valeur des pledge des entreprises Google, IBM et Microsoft par rapport aux brevets non pledge de Google, IBM et Microsoft mais également par rapport aux brevets ayant des technologies similaires à la propriété intellectuelle de ces trois entreprises.

Les tableau 3.5 présente les résultats des estimations de **l'impact de l'ensemble des brevets pledge et des pledge de chaque entreprise**. Si l'on considère les citations avec autocitations, les pledge ont un impact faiblement positif et significatif et lorsque l'on considère le détail, les pledge d'IBM et de Google ont un impact positif alors que ceux de Microsoft ont un impact négatif. Ainsi les brevets pledge ont donc permis d'accroître la valeur privée des deux premières. Lorsque l'on considère les citations hors autocitations, les résultats sont inversés. Les pledge ont globalement un impact négatif et significatif comme ceux des entreprises IBM et Google et seul les pledge de Microsoft ont un impact significatif et positif. Les entreprises du secteur des TIC se sont donc appuyées sur les brevets pledge Microsoft pour développer leurs innovations, les brevets pledge ont donc permis d'augmenter la valeur sociale de Microsoft.

Tableau 3.5 Ensemble des brevets (Google, IBM et Microsoft) et leurs contrôles

	# Citations		# Citations sans autocitations	
	(1)	(2)	(3)	(4)
Brevets Pledge	0.12+		-0.21**	
	[0.07]		[0.10]	
Brevets Pledge IBM		0.17***		-0.19***
		[0.06]		[0.05]
Brevets Pledge Microsoft		-0.17***		0.07**
		[0.06]		[0.03]
Brevets Pledge Google		0.42***		-0.82***
		[0.10]		[0.07]
Co-brevets	-0.03	-0.03	-0.03+	-0.03+
	[0.02]	[0.02]	[0.02]	[0.02]
Nombre de sous-groupes CPC	0.07**	0.07**	0.06**	0.06**
	[0.03]	[0.03]	[0.02]	[0.02]
Nombre d'inventeurs	0.05**	0.05**	0.04**	0.04**
	[0.02]	[0.02]	[0.02]	[0.02]
Taille de la famille	0.04***	0.04***	0.03***	0.03***
	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de citations antérieures	0.01	0.01	0.01	0.01
	[0.01]	[0.01]	[0.00]	[0.00]
Nombre de citation hors brevets	0.03**	0.03**	0.03**	0.03**
	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de revendications	0.09***	0.09***	0.07***	0.07***
	[0.03]	[0.03]	[0.02]	[0.02]
Constant	0.64***	0.64***	0.61***	0.61***
	[0.18]	[0.18]	[0.14]	[0.14]
Observations	1341610	1341610	1341610	1341610
R-Squared	.68	.68	.65	.65

Effets fixes année de priorité, technologie Wipo et entreprise. Entre parenthèses les écarts-types de coefficients avec un cluster sur les technologies et les entreprises. + 0.10 ** 0.05 *** 0.01

Les tableaux 3.6 et 3.7 détaillent les estimations pour Google et IBM en précisant l'origine du brevet pledge. Pour rappel, Microsoft a développé la quasi-totalité de ses brevets pledge, il n'y a donc pas de variable sur l'origine du dépôt de brevet.

Concernant Google (Tableau 3.6), lorsque l'on considère l'impact des pledge sur les citations incluant les autocitations, l'impact est positif et significatif pour l'ensemble des pledge Google et pour les Pledge non déposés initialement par cette entreprise. Ce sont donc les brevets pledge acquis par Google qui ont permis à la firme d'accroître sa valeur privée. En revanche, les Pledge associés à MapReduce n'ont pas d'impact. Concernant l'impact des pledge sur les citations sans autocitation, l'impact est négatif, que les brevets aient été inventés ou non par Google.

Concernant IBM (Tableau 3.7), les résultats sont semblables à ceux de Google. Les pledge ont un impact positif et significatif sur les citations avec autocitations, donc sur la valeur privée des brevets et ce quelle que soit leur origine. Et l'impact est négatif hors autocitation, donc sur la valeur sociale.

Tableau 3.6 Impact des pledge Google par rapport à l'ensemble des brevets et contrôles

	# Citations					
	Toutes		Sans autocitations			
	(1)	(2)	(3)	(4)	(5)	(6)
Brevets Pledge Google	0.41*** [0.09]			-0.82*** [0.06]		
Brevets Pledge déposés initialement par Google		-0.08 [0.07]			-0.16** [0.07]	
Brevets Pledge non déposés initialement par Google		0.52*** [0.10]	0.52*** [0.10]		-0.96*** [0.08]	-0.96*** [0.09]
Brevets Pledge MapReduce Google			0.00 [0.03]			-0.30*** [0.04]
Brevets Pledge déposés initialement par Google hors MapReduce			-0.16 [0.11]			-0.03 [0.02]
Co-brevets	-0.03 [0.02]	-0.03 [0.02]	-0.03 [0.02]	-0.03+ [0.02]	-0.03+ [0.02]	-0.03+ [0.02]
Nombre de sous-groupes CPC	0.07** [0.03]	0.07** [0.03]	0.07** [0.03]	0.06** [0.02]	0.06** [0.02]	0.06** [0.02]
Nombre d'inventeurs	0.05** [0.02]	0.05** [0.02]	0.05** [0.02]	0.04** [0.02]	0.04** [0.02]	0.04** [0.02]
Taille de la famille	0.04*** [0.01]	0.04*** [0.01]	0.04*** [0.01]	0.03*** [0.01]	0.03*** [0.01]	0.03*** [0.01]
Nombre de citations antérieures	0.01 [0.01]	0.01 [0.01]	0.01 [0.01]	0.01 [0.00]	0.01 [0.00]	0.01 [0.00]
Nombre de citation hors brevets	0.03** [0.01]	0.03** [0.01]	0.03** [0.01]	0.03** [0.01]	0.03** [0.01]	0.03** [0.01]
Nombre de revendications	0.09*** [0.03]	0.09*** [0.03]	0.09*** [0.03]	0.07*** [0.02]	0.07*** [0.02]	0.07*** [0.02]
Constant	0.64*** [0.18]	0.64*** [0.18]	0.64*** [0.18]	0.61*** [0.14]	0.61*** [0.14]	0.61*** [0.14]
Observations	1341610	1341610	1341610	1341610	1341610	1341610
R-Squared	.68	.68	.68	.65	.65	.65

Effets fixes année de priorité, technologie Wipo et entreprise. Entre parenthèses les écarts-types de coefficients avec un cluster sur les technologies et les entreprises. + 0.10 ** 0.05 *** 0.01

Au total, une tendance claire se dégage concernant l'utilisation des brevets pledge par ces trois entreprises. L'entreprise Google s'est appuyée sur les pledge dont elle a fait l'acquisition pour développer sa propriété intellectuelle et donc sa valeur privée. De même pour IBM qui a également utilisé l'ensemble de ses brevets pledge pour ses développements internes. Les pledge de Microsoft sont les seuls à accroître la valeur sociale des brevets.

Quand on élargit le périmètre d'analyse à trois acteurs majeurs des TIC mais également à un périmètre incluant des technologies similaires à la propriété intellectuelle de ces trois firmes, les brevets associés à la technologie MapReduce ne reçoivent pas significativement plus de citations.

Tableau 3.7 Impact des pledge IBM par rapport à l'ensemble des brevets et contrôles

	# Citations			
	Toutes (1)	(2)	Sans autocitations (3)	(4)
Brevets Pledge IBM	0.17*** [0.05]		-0.19*** [0.05]	
Brevets Pledge déposés initialement par IBM		0.12*** [0.04]		-0.20*** [0.05]
Brevets non déposés initialement par IBM		0.82*** [0.15]		-0.10 [0.18]
Co-brevets	-0.03 [0.02]	-0.03 [0.02]	-0.03 [0.02]	-0.03 [0.02]
Nombre de sous-groupes CPC	0.07** [0.03]	0.07** [0.03]	0.06** [0.02]	0.06** [0.02]
Nombre d'inventeurs	0.05** [0.02]	0.05** [0.02]	0.04** [0.02]	0.04** [0.02]
Taille de la famille	0.04*** [0.01]	0.04*** [0.01]	0.02*** [0.01]	0.02*** [0.01]
Nombre de citations antérieures	0.01 [0.01]	0.01 [0.01]	0.01 [0.00]	0.01 [0.00]
Nombre de citation hors brevets	0.03** [0.01]	0.03** [0.01]	0.03** [0.01]	0.03** [0.01]
Nombre de revendications	0.09*** [0.03]	0.09*** [0.03]	0.07*** [0.02]	0.07*** [0.02]
Constant	0.64*** [0.18]	0.64*** [0.18]	0.61*** [0.14]	0.61*** [0.14]
Observations	1341610	1341610	1341610	1341610
R-Squared	.68	.68	.65	.65

Effets fixes année de priorité, technologie Wipo et entreprise. Entre parenthèses les écarts-types de coefficients avec un cluster sur les technologies et les entreprises. + 0.10 ** 0.05 *** 0.01

3.6 Estimations agrégées centrées sur les des portefeuilles de brevets

Cette dernière section s'intéresse à la valeur des pledge pour les entreprises Google, IBM et Microsoft en comparaison à leurs propres brevets. **La question est de savoir si les pledge ont une valeur privée et sociale plus élevée que le reste des brevets de leur portefeuille.** L'échantillon se réduit à présent à 101 878 brevets.

Les estimations introduisent systématiquement des effets fixes sur l'année de priorité, la technologie et les entreprises.

3.6.1 Statistiques descriptives

Le tableau 3.8 présente les statistiques descriptives. Les chiffres indiquent que les citations sans autocitation sont cinq fois moins nombreuses témoignant ainsi du fait que l'accumulation des connaissances s'appuie largement sur les portefeuilles et stocks de connaissance interne. La part des pledge dans les portefeuilles représente 0,7 % des brevets, une partie ayant été acquise

par l'entreprise (0,1% des brevets sont des pledge ayant été initialement déposés par d'autres organisations).

Tableau 3.8 Statistiques descriptives				
	Moyenne	Ecart-type	Minimum	Maximum
Variabes dépendantes				
Nombre de citations de brevets reçues (5 ans)	7.475	13.435	0.000	459.000
Nombre de citations de brevets reçues sans autocitation (5 ans)	1.504	3.310	0.000	101.000
Variabes d'intérêt				
Brevets Pledge des 3 entreprises (Google, IBM, Microsoft)	0.007	0.083	0.000	1.000
Brevets Pledge déposés initialement par les 3 entreprises	0.006	0.076	0.000	1.000
Brevets Pledge non déposés initialement par les 3 entreprises	0.001	0.034	0.000	1.000
Brevets Pledge Google	0.001	0.031	0.000	1.000
Brevets Pledge IBM	0.004	0.063	0.000	1.000
Brevets Pledge Microsoft	0.002	0.045	0.000	1.000
Brevets Pledge MapReduce Google	0.000	0.008	0.000	1.000
Brevets Pledge déposés initialement par Google hors MapReduce	0.000	0.010	0.000	1.000
Brevets Pledge non déposés initialement par Google	0.001	0.029	0.000	1.000
Brevets Pledge déposés initialement par IBM	0.004	0.060	0.000	1.000
Brevets non déposés initialement par IBM	0.000	0.017	0.000	1.000
Brevets Pledge Microsoft	0.002	0.045	0.000	1.000
Variabes de contrôle				
Co-brevets	0.020	0.139	0.000	1.000
Nombre de sous-groupes CPC	5.128	6.065	0.000	300.000
Nombre d'inventeurs	3.348	1.996	1.000	76.000
Taille de la famille	2.127	2.363	0.000	33.000
Nombre de citations antérieures	22.317	34.734	0.000	1922.000
Nombre de citations hors brevets	7.532	19.268	0.000	953.000
Nombre de revendications	18.497	10.382	0.000	202.000
TIC	0.709	0.454	0.000	1.000
Observations	101 878			

3.6.2 Résultat des estimations

3.6.2.1 Impact des pledge sur la valeur privée et sociale des brevets

Le tableau 3.9 présente une estimation de l'impact de pledge des trois entreprises confondues sans explorer leur origine ni les entreprises auxquelles elles appartiennent.

Concernant les citations, l'impact est non significatif, il est négatif pour les pledge déposés initialement par les entreprises et positif pour les pledge ayant été acquis par les entreprises. Lorsqu'il s'agit des citations excluant les autocitations, les impacts sont significatifs et négatifs pour les pledge acquis par les entreprises. Autrement dit, Google, IBM et Microsoft ont acquis des brevets qui ont une valeur pour leur propre développement technologique augmentant de fait leurs valeurs privées mais leur valeur sociale (citations hors autocitation) est inférieure au

reste des brevets de leur portefeuille. Il apparaît donc que les autres entreprises du secteur ne se sont pas appuyées sur ces pledge pour développer leurs innovations.

En résumé, les brevets pledge ne sont pas plus cités que les brevets non pledge voire ils le sont moins si on exclut les autocitations.

3.6.2.2 Impact des pledge selon les domaines technologiques des TIC

Les tableaux 3.10 et 3.11 ont pour objet de tester **la robustesse de ces régressions** en évaluant la valeur des pledge sur deux sous-échantillons :

- Les domaines technologiques caractérisant les **TIC** (Tableau 3.10) correspondants aux domaines **WIPO 4 à 7**, à savoir « Communication numérique », « techniques de communication de base », « informatique » et « méthodes de traitement des données à des fins de gestion ».
- Le domaine technologique de l'« **informatique** » (**WIPO 6**) (Tableau 3.11) qui représente à lui seul plus de 50% des technologies de notre échantillon.

Les résultats sont globalement similaires à ceux obtenus pour l'ensemble des portefeuilles, cependant certains résultats voient leur significativité statistique s'améliorer suivant le périmètre technologique considéré. Ainsi, si on exclut les autocitations, la variable associée aux pledge des trois entreprises devient plus fortement significative ; sur l'ensemble des domaines technologiques, la variable est négative avec une significativité de 10%, en considérant uniquement 4 domaines, cette significativité est de 5% et enfin si on se concentre uniquement sur le domaine « informatique », elle est de 1%. Au total, il apparaît donc que les brevets pledge de Google, IBM et Microsoft dans leur ensemble ont reçu significativement moins de citations que leurs brevets non pledge si on exclut les autocitations. Ce résultat est d'autant plus significatif si on se concentre sur les TIC. Ensuite, pour les brevets initialement développés par Google, IBM et Microsoft, le coefficient associé à ces brevets devient faiblement significatif pour les estimations excluant les autocitations. Enfin, pour les brevets qui n'ont pas été déposés initialement par ces trois entreprises, le coefficient devient faiblement significatif si on considère les TIC pour les estimations incluant les autocitations.

Tableau 3.9 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft

	# Citations				# Citations sans autocitations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Brevets Pledge	0.007 [0.031]				-0.060+ [0.033]			
Brevets Pledge déposés initialement par les 3 entreprises		-0.078*** [0.027]		-0.077*** [0.027]		0.001 [0.030]		0.000 [0.030]
Brevets Pledge non déposés initialement par les 3 entreprises			0.445*** [0.123]	0.444*** [0.122]			-0.374*** [0.100]	-0.374*** [0.100]
Co-brevets	0.031 [0.030]	0.031 [0.030]	0.032 [0.030]	0.032 [0.030]	-0.061*** [0.016]	-0.061*** [0.016]	-0.061*** [0.016]	-0.061*** [0.016]
Nombre de sous-groupes CPC	0.108*** [0.015]	0.108*** [0.015]	0.107*** [0.015]	0.107*** [0.015]	0.062*** [0.018]	0.062*** [0.018]	0.062*** [0.018]	0.062*** [0.018]
Nombre d'inventeurs	0.084*** [0.013]	0.084*** [0.013]	0.084*** [0.013]	0.084*** [0.013]	0.094*** [0.016]	0.094*** [0.016]	0.094*** [0.016]	0.094*** [0.016]
Taille de la famille	0.039** [0.015]	0.039** [0.015]	0.038** [0.016]	0.038** [0.016]	0.018** [0.008]	0.018** [0.008]	0.019** [0.008]	0.019** [0.008]
Nombre de citations antérieures	0.012 [0.010]	0.012 [0.010]	0.013 [0.010]	0.013 [0.010]	0.004 [0.007]	0.004 [0.007]	0.004 [0.007]	0.004 [0.007]
Nombre de citation hors brevets	0.049*** [0.006]	0.049*** [0.006]	0.049*** [0.006]	0.049*** [0.006]	0.057*** [0.008]	0.057*** [0.008]	0.057*** [0.008]	0.057*** [0.008]
Nombre de revendications	0.125*** [0.008]	0.125*** [0.008]	0.125*** [0.008]	0.125*** [0.008]	0.058*** [0.009]	0.058*** [0.009]	0.058*** [0.009]	0.058*** [0.009]
Constant	0.772*** [0.039]	0.772*** [0.039]	0.771*** [0.039]	0.771*** [0.039]	0.084+ [0.046]	0.084+ [0.046]	0.085+ [0.046]	0.085+ [0.046]
Observations	101878	101878	101878	101878	101878	101878	101878	101878
R-Squared	.26	.26	.26	.26	.085	.085	.085	.085

Effets fixes année de priorité, technologie Wipo et entreprise. Entre parenthèses les écarts-types de coefficients avec un cluster sur les technologies et les entreprises.
+ 0.10 ** 0.05 *** 0.01

**Tableau 3.10 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft
– Sous échantillon des brevets TIC**

	# Citations				# Citations sans autocitations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Brevets Pledge	-0.038+ [0.014]				-0.105** [0.022]			
Brevets Pledge déposés initialement par les 3 entreprises		-0.123*** [0.016]		-0.121*** [0.016]		-0.033 [0.031]		-0.035 [0.031]
Brevets Pledge non déposés initialement par les 3 entreprises			0.406+ [0.158]	0.404+ [0.158]			-0.479*** [0.073]	-0.480*** [0.073]
Co-brevets	-0.042 [0.026]	-0.043 [0.026]	-0.041 [0.026]	-0.042 [0.026]	-0.095+ [0.031]	-0.095+ [0.031]	-0.095+ [0.031]	-0.096+ [0.031]
Nombre de sous-groupes CPC	0.122*** [0.006]	0.122*** [0.006]	0.122*** [0.006]	0.122*** [0.006]	0.084*** [0.010]	0.084*** [0.010]	0.084*** [0.010]	0.084*** [0.010]
Nombre d'inventeurs	0.100*** [0.006]	0.100*** [0.006]	0.099*** [0.006]	0.100*** [0.006]	0.108** [0.019]	0.107** [0.019]	0.108** [0.019]	0.108** [0.019]
Taille de la famille	0.022** [0.005]	0.022** [0.005]	0.021** [0.005]	0.021** [0.005]	0.014 [0.007]	0.014 [0.007]	0.015 [0.007]	0.015 [0.007]
Nombre de citations antérieures	0.010 [0.012]	0.010 [0.012]	0.010 [0.012]	0.010 [0.012]	0.009+ [0.003]	0.009+ [0.003]	0.008+ [0.003]	0.008+ [0.003]
Nombre de citation hors brevets	0.052*** [0.007]	0.052*** [0.007]	0.052*** [0.007]	0.052*** [0.007]	0.061** [0.011]	0.061** [0.011]	0.061** [0.011]	0.061** [0.011]
Nombre de revendications	0.113*** [0.006]	0.113*** [0.006]	0.113*** [0.006]	0.113*** [0.006]	0.042*** [0.004]	0.042*** [0.004]	0.042*** [0.004]	0.042*** [0.004]
Constant	0.781*** [0.057]	0.781*** [0.057]	0.781*** [0.057]	0.781*** [0.056]	0.060 [0.047]	0.060 [0.047]	0.060 [0.047]	0.060 [0.047]
Observations	72253	72253	72253	72253	72253	72253	72253	72253
R-Squared	.28	.28	.28	.28	.11	.11	.11	.11

Effets fixes année de priorité, technologie Wipo et entreprise. Entre parenthèses les écarts-types de coefficients avec un cluster sur les technologies et les entreprises.

+ 0.10 ** 0.05 *** 0.01

**Tableau 3.11 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft
– Sous échantillon des brevets « Informatique »**

	# Citations				# Citations sans autocitations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Brevets Pledge	-0.046 [0.044]				-0.127*** [0.036]			
Brevets Pledge déposés initialement par les 3 entreprises		-0.126*** [0.044]		-0.125*** [0.044]		-0.068+ [0.038]		-0.069+ [0.038]
Brevets Pledge non déposés initialement par les 3 entreprises			0.507*** [0.154]	0.504*** [0.154]			-0.528*** [0.096]	-0.529*** [0.096]
Co-brevets	-0.070+ [0.037]	-0.070+ [0.037]	-0.069+ [0.037]	-0.070+ [0.037]	-0.120*** [0.030]	-0.119*** [0.030]	-0.119*** [0.030]	-0.120*** [0.030]
Nombre de sous-groupes CPC	0.124*** [0.008]	0.124*** [0.008]	0.123*** [0.008]	0.123*** [0.008]	0.091*** [0.006]	0.091*** [0.006]	0.091*** [0.006]	0.091*** [0.006]
Nombre d'inventeurs	0.104*** [0.009]	0.104*** [0.009]	0.103*** [0.009]	0.104*** [0.009]	0.126*** [0.008]	0.126*** [0.008]	0.126*** [0.008]	0.126*** [0.008]
Taille de la famille	0.020** [0.009]	0.019** [0.009]	0.018** [0.009]	0.018** [0.009]	0.007 [0.007]	0.007 [0.007]	0.008 [0.007]	0.008 [0.007]
Nombre de citations antérieures	0.022*** [0.005]	0.022*** [0.005]	0.022*** [0.005]	0.022*** [0.005]	0.010** [0.004]	0.010** [0.004]	0.010** [0.004]	0.010** [0.004]
Nombre de citation hors brevets	0.058*** [0.004]	0.059*** [0.004]	0.058*** [0.004]	0.058*** [0.004]	0.070*** [0.003]	0.070*** [0.003]	0.070*** [0.003]	0.071*** [0.003]
Nombre de revendications	0.116*** [0.007]	0.116*** [0.007]	0.117*** [0.007]	0.117*** [0.007]	0.043*** [0.006]	0.043*** [0.006]	0.043*** [0.006]	0.043*** [0.006]
Constant	0.734*** [0.028]	0.734*** [0.028]	0.734*** [0.028]	0.734*** [0.028]	0.058** [0.023]	0.057** [0.023]	0.057** [0.023]	0.057** [0.023]
Observations	52507	52507	52507	52507	52507	52507	52507	52507
R-Squared	.28	.28	.28	.28	.11	.11	.11	.11

Effets fixes année de priorité, technologie Wipo et entreprise. Entre parenthèses les écarts-types de coefficients avec un cluster sur les technologies et les entreprises.

+ 0.10 ** 0.05 *** 0.01

En résumé, ces résultats indiquent que :

- Les brevets pledge reçoivent moins de citations que les autres brevets de leur portefeuille,
- Les brevets pledge non déposés initialement par ces entreprises ont une valeur privée plus importante en revanche,
- Les brevets pledge non déposés initialement par ces entreprises ont une valeur sociale clairement plus faible.

3.6.2.3 Impact des pledge : détails par entreprise (Google, IBM et Microsoft)

Les régressions suivantes précisent les premiers résultats en spécifiant **l'entreprise propriétaire des pledge (Google, IBM ou Microsoft) ainsi que leur origine**. Il s'agit d'étudier si les premiers résultats ne cachent pas **des différences entre les trois entreprises**. Pour Google, il est spécifié si le pledge fait partie de MapReduce ou non.

Le Tableau 3.12 détaille les pledge pour chaque entreprise et laisse entrevoir des résultats plus détaillés. Ainsi, si on considère la valeur privée, seuls les brevets pledge de Google ont un impact positif, ceux de Microsoft ont un impact négatif. En détaillant l'origine des brevets, il apparaît que les pledge non déposés initialement par Google et IBM augmentent la valeur privée. Concernant la valeur sociale, les pledge reçoivent moins de citations des autres acteurs du secteur que les brevets non pledge du portefeuille et seuls les brevets de l'entreprises Google ont un impact. Les pledge non déposés par Google ainsi que MapReduce réduisent la valeur sociale en revanche, les pledge non initialement déposés par Google hors MapReduce ont une valeur sociale positive.

Ces résultats confirment que :

- Les pledge non déposés initialement par Google et IBM ont permis de développer leur propriété intellectuelle et d'accroître leur valeur privée.
- Seuls les pledge déposés par Google hors MapReduce accroissent leur valeur sociale.

Tableau 3.12 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft
Détail des types de pledge par entreprise

	# Citations			# Citations sans autocitations		
	(1)	(2)	(3)	(4)	(5)	(6)
Brevets Pledge Google	0.25**			-0.50***		
	[0.11]			[0.07]		
Brevets Pledge IBM	0.05			0.02		
	[0.06]			[0.05]		
Brevets Pledge Microsoft	-0.18***	-0.18***	-0.18***	-0.01	-0.01	-0.01
	[0.06]	[0.06]	[0.06]	[0.04]	[0.04]	[0.04]
		[0.13]	[0.13]		[0.11]	[0.11]
Brevets Pledge non déposés initialement par Google		0.32**	0.32**		-0.60***	-0.60***
		[0.13]	[0.13]		[0.11]	[0.11]
Brevets Pledge déposés initialement par IBM		-0.01	-0.01		0.01	0.01
		[0.05]	[0.05]		[0.04]	[0.04]
Brevets non déposés initialement par IBM		0.76***	0.76***		0.19	0.19
		[0.13]	[0.13]		[0.13]	[0.13]
Brevets Pledge MapReduce Google			-0.05			-0.31***
			[0.04]			[0.03]
Brevets Pledge déposés initialement par Google hors MapReduce			-0.04			0.19***
			[0.16]			[0.02]
Co-brevets	0.03	0.03	0.03	-0.06***	-0.06***	-0.06***
	[0.03]	[0.03]	[0.03]	[0.02]	[0.02]	[0.02]
Nombre de sous-groupes CPC	0.11***	0.11***	0.11***	0.06***	0.06***	0.06***
	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]	[0.02]
Nombre d'inventeurs	0.08***	0.08***	0.08***	0.09***	0.09***	0.09***
	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]	[0.02]
Taille de la famille	0.04**	0.04**	0.04**	0.02**	0.02**	0.02**
	[0.02]	[0.02]	[0.02]	[0.01]	[0.01]	[0.01]
Nombre de citations antérieures	0.01	0.01	0.01	0.00	0.00	0.00
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de citation hors brevets	0.05***	0.05***	0.05***	0.06***	0.06***	0.06***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de revendications	0.13***	0.12***	0.12***	0.06***	0.06***	0.06***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Constant	0.77***	0.77***	0.77***	0.08+	0.09+	0.09+
	[0.04]	[0.04]	[0.04]	[0.05]	[0.05]	[0.05]
Observations	101878	101878	101878	101878	101878	101878
R-Squared	.26	.26	.26	.086	.086	.086

Effets fixes année de priorité, technologie Wipo et entreprise. Entre parenthèses les écarts-types de coefficients avec un cluster sur les technologies et les entreprises. + 0.10 ** 0.05 *** 0.01

Les tableaux suivants précisent les résultats sur les sous-échantillons des TIC et de l'informatique. Il semble que les estimations précédentes reflètent surtout ce qui se passe dans le domaine de « l'informatique ».

- Le rôle des pledge pour la valeur privée apparaît dans le domaine de l'informatique (Tableau 3.14) confirmant l'impact positif des pledge de Google et celui des pledge non initialement déposés par Google et IBM, les pledge de Microsoft ayant un impact négatif.
- Le rôle des pledge pour la valeur sociale apparaît dans le domaine des TIC (Tableau 3.13) et uniquement pour Google. L'impact est négatif dans l'ensemble sauf pour les pledge déposés initialement par Google hors MapReduce.

Tableau 3.13 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft
Détail des types de pledge par entreprise – Sous échantillon TIC

	# Citations					
	Toutes			Sans autocitations		
	(1)	(2)	(3)	(4)	(5)	(6)
Brevets Pledge Google	0.22			-0.58***		
	[0.13]			[0.03]		
Brevets Pledge IBM	-0.03			-0.04		
	[0.02]			[0.04]		
Brevets Pledge Microsoft	-0.17	-0.17	-0.17	-0.00	-0.00	-0.00
	[0.08]	[0.08]	[0.08]	[0.05]	[0.05]	[0.05]
Brevets Pledge déposés initialement par Google		-0.03			-0.01	
		[0.10]			[0.08]	
Brevets Pledge non déposés initialement par Google		0.28	0.28		-0.72***	-0.72***
		[0.16]	[0.16]		[0.08]	[0.08]
Brevets Pledge déposés initialement par IBM		-0.10**	-0.10**		-0.05	-0.05
		[0.02]	[0.02]		[0.02]	[0.02]
Brevets non déposés initialement par IBM		0.72**	0.72**		0.14	0.14
		[0.13]	[0.13]		[0.14]	[0.14]
Brevets Pledge MapReduce Google			-0.04			-0.33***
			[0.04]			[0.02]
Brevets Pledge déposés initialement par Google hors MapReduce			-0.02			0.20***
			[0.18]			[0.01]
Co-brevets	-0.04	-0.04	-0.04	-0.10+	-0.10+	-0.10+
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
Nombre de sous-groupes CPC	0.12***	0.12***	0.12***	0.08***	0.08***	0.08***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre d'inventeurs	0.10***	0.10***	0.10***	0.11**	0.11**	0.11**
	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]	[0.02]
Taille de la famille	0.02**	0.02**	0.02**	0.01	0.02	0.02
	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]	[0.01]
Nombre de citations antérieures	0.01	0.01	0.01	0.01**	0.01**	0.01**
	[0.01]	[0.01]	[0.01]	[0.00]	[0.00]	[0.00]
Nombre de citation hors brevets	0.05***	0.05***	0.05***	0.06**	0.06**	0.06**
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de revendications	0.11***	0.11***	0.11***	0.04***	0.04***	0.04***
	[0.01]	[0.01]	[0.01]	[0.00]	[0.00]	[0.00]
Constant	0.78***	0.78***	0.78***	0.06	0.06	0.06
	[0.06]	[0.06]	[0.06]	[0.05]	[0.05]	[0.05]
Observations	72253	72253	72253	72253	72253	72253
R-Squared	.28	.28	.28	.11	.11	.11

Effets fixes année de priorité, technologie Wipo et entreprise. Entre parenthèses les écarts-types de coefficients avec un cluster sur les technologies et les entreprises. + 0.10 ** 0.05 *** 0.01

Tableau 3.14 Impact des pledge sur les citations des brevets de Google, IBM et Microsoft
Détail des types de pledge par entreprise – Sous échantillon « Informatique »

	# Citations			# Citations sans autocitations		
	(1)	(2)	(3)	(4)	(5)	(6)
Brevets Pledge Google	0.33**			-0.61***		
	[0.16]			[0.08]		
Brevets Pledge IBM	-0.05			-0.06		
	[0.05]			[0.05]		
Brevets Pledge Microsoft	-0.26***	-0.26***	-0.26***	-0.07	-0.07	-0.07
	[0.08]	[0.08]	[0.08]	[0.06]	[0.06]	[0.06]
Brevets Pledge déposés initialement par Google		0.03			-0.07	
		[0.19]			[0.21]	
Brevets Pledge non déposés initialement par Google		0.45**	0.45**		-0.81***	-0.81***
		[0.20]	[0.20]		[0.05]	[0.05]
Brevets Pledge déposés initialement par IBM		-0.09+	-0.09+		-0.07	-0.07
		[0.05]	[0.05]		[0.05]	[0.05]
Brevets non déposés initialement par IBM		0.62***	0.62***		0.03	0.03
		[0.21]	[0.21]		[0.22]	[0.22]
Brevets Pledge MapReduce Google			-0.08			-0.35
			[0.24]			[0.21]
Brevets Pledge déposés initialement par Google hors MapReduce			0.14			0.21
			[0.28]			[0.34]
Co-brevets	-0.07+	-0.07+	-0.07+	-0.12***	-0.12***	-0.12***
	[0.04]	[0.04]	[0.04]	[0.03]	[0.03]	[0.03]
Nombre de sous-groupes CPC	0.12***	0.12***	0.12***	0.09***	0.09***	0.09***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre d'inventeurs	0.10***	0.10***	0.10***	0.13***	0.13***	0.13***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Taille de la famille	0.02**	0.02**	0.02**	0.01	0.01	0.01
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Nombre de citations antérieures	0.02***	0.02***	0.02***	0.01**	0.01**	0.01**
	[0.01]	[0.01]	[0.01]	[0.00]	[0.00]	[0.00]
Nombre de citation hors brevets	0.06***	0.06***	0.06***	0.07***	0.07***	0.07***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Nombre de revendications	0.12***	0.12***	0.12***	0.04***	0.04***	0.04***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Constant	0.73***	0.73***	0.73***	0.06**	0.06**	0.06**
	[0.03]	[0.03]	[0.03]	[0.02]	[0.02]	[0.02]
Observations	52507	52507	52507	52507	52507	52507
R-Squared	.28	.28	.28	.11	.11	.11

Effets fixes année de priorité et entreprise. Entre parenthèses les écarts-types de coefficients avec un cluster sur les entreprises. + 0.10 ** 0.05 *** 0.01

Au total, les estimations détaillées des entreprises indiquent plusieurs éléments notables. Tout d'abord pour l'entreprise Google, les brevets pledge acquis par l'entreprise reçoivent significativement plus de citations que les brevets non pledge si on tient compte des autocitations, ces brevets ont permis d'accroître la valeur privée de Google. De plus, les brevets pledge déposés initialement par Google hors technologie MapReduce reçoivent significativement plus de citations que les brevets non pledge si on exclut les autocitations. **Ces brevets ont donc eu un impact sur les autres entreprises du secteur, ils ont permis**

d'accroître la valeur sociale de Google. En somme, les brevets pledge que l'entreprise Google a créés ou acquis ont permis de développer soit la valeur privée (brevet acquis) soit la valeur sociale (brevets créés) de la firme.

Des trois entreprises étudiées, l'entreprise Google est la seule entreprise qui a su profiter d'une valorisation interne et externe à travers ses pledge. Ce résultat est cependant altéré si on restreint le périmètre d'analyse aux TIC. Il apparaît donc que suivant le périmètre technologique considéré, les brevets pledge de Google n'ont pas le même impact. Enfin on note que les brevets pledge sur la technologie MapReduce n'ont eu aucun impact significatif par rapport aux brevets non pledge. Pour l'entreprise IBM, on retient qu'à l'image de Google, les brevets acquis ultérieurement par la firme reçoivent significativement plus de citations que les autres brevets. Ces brevets ont donc permis d'accroître la valeur privée de l'entreprise IBM. Ce résultat reste robuste quand on se limite aux TIC. Enfin pour Microsoft, on constate que les brevets pledge, qui ont été développés en quasi-totalité par la firme, ont plutôt eu un impact négatif sur sa valeur privée et aucun impact sur sa valeur sociale par rapport aux autres brevets du portefeuille. Ce résultat n'est pas modifié quand on se restreint aux TIC.

3.7 Conclusion

Ce chapitre avait pour objectif d'étudier si la technologie MapReduce était devenue une technologie majeure des technologies Big Data grâce à son ouverture par des mécanismes d'innovation ouverte ou si c'est le caractère innovant intrinsèque de la technologie qui avait permis d'impacter le secteur d'activité. Dans une première partie, nous avons étudié les différentes stratégies d'innovation ouverte afin de pouvoir positionner la technologie MapReduce dans le champ de l'innovation ouverte et d'élargir la discussion à l'ensemble des pledge.

L'analyse empirique réalisée dans ce chapitre a exploré l'impact des pledge en termes de valeur privée et sociale en comparaison à deux périmètres d'analyse, d'une part, les brevets des portefeuilles des entreprises Google, IBM et Microsoft et des contrôles ayant les mêmes caractéristiques et d'autre part, uniquement au sein de leurs portefeuilles. L'objectif était in fine de vérifier l'impact des patent pledge incluant la technologie MapReduce sur ce périmètre par rapport à des brevets ne faisant pas l'objet d'un patent pledge.

Quatre résultats majeurs doivent être retenus :

1. Google et IBM se sont appuyés sur leur pledge et en particulier sur les pledge acquis pour asseoir leur développement technologique en interne. Le rôle des pledge a eu un impact positif sur leur valeur privée. Ce résultat est robuste sur l'échantillon avec contrôle et celui incluant uniquement leurs portefeuilles.
2. Microsoft est la seule entreprise dont les inventions ont contribué au développement technologique des autres entreprises du secteur lorsque l'on considère le périmètre avec contrôles. Autrement dit, les pledge augmentent la valeur sociale des brevets comparé aux brevets du secteur. Ce résultat n'est plus vrai au sein du portefeuille des entreprises.
3. Seuls les brevets de Google développés en interne hors MapReduce ont favorisé la valeur sociale, le développement technologique des autres entreprises du secteur, comparé aux brevets maison.
4. La technologie MapReduce n'a eu d'impact ni en termes de valeur privée ni en termes de valeur sociale que ce soit par rapport aux autres brevets du secteur ou au sein du portefeuille de Google.

Ce dernier résultat indique que la technologie MapReduce n'est pas une technologie plus innovante si on la compare aux brevets du secteur des TIC et aux autres pledge. Ce résultat tranche avec le chapitre 2 mais s'explique par les différences de périmètres. Dans le chapitre 2, il s'agissait de MapReduce par rapport à des contrôles limités aux technologies de traitement des Big Data alors que dans le chapitre 3 le périmètre est très large. Au total, il s'avère donc que l'ouverture de la technologie MapReduce a été un élément déterminant pour en faire une technologie majeure mais son rôle de pledge n'est pas déterminant au sein du portefeuille de Google.

3.8 Références

- Barton, J. H. (2004). Patents and the Transfer of Technology to Developing Countries. *In Patents, Innovation and Economic Performance* (pp. 321–332). OECD. <https://doi.org/10.1787/9789264015272-24-en>
- Baudry, M., & Dumont, B. (2020). Que nous apprennent les brevets de Tesla Motors sur sa stratégie PI ? https://www.openscience.fr/IMG/pdf/iste_techinn20v5n4_6.pdf
- Beky, A. (2012). IBM reste le moteur de la propriété intellectuelle avec 6180 brevets déposés en 2011 aux États-Unis. <https://www.Silicon.Fr/Ibm-Reste-Le-Moteur-de-La-Propriete-Intellectuelle-Avec-6180-Brevets-Deposes-En-2011-Aux-Etats-Unis-70655.Html>.
- Briscoe, G., & Mulligan, C. (2018). Digital Innovation: The Hackathon Phenomenon. <https://core.ac.uk/download/pdf/30697508.pdf>
- Chesbrough, H. (2003a). Open Innovation : The new imperative for creating and profiting from Technology. *Harvard Business School Press*
- Chesbrough, H. (2003b). The logic of Open Innovation: Managing intellectual property. *California Management Review*. 45. 33-58. [10.2307/41166175](https://doi.org/10.2307/41166175).
- Chesbrough, H. (2006). Open Innovation: A New Paradigm for Understanding Industrial Innovation. *Oxford University Press*.
- Contreras, J. L., Hall, B. H., & Helmers, C. (2018). Pledging Patents for the Public Good: Rise and Fall of the Eco-Patent Commons. *Houston Law Review* 57: 61-109
- Euchner, J. (2010). Two Flavors of Open Innovation. *Research-Technology Management*, 53(4), 7–8. <https://doi.org/10.1080/08956308.2010.11657634>
- Euchner, J. (2013). The uses and risks of open innovation. *Research Technology Management*, 56(3), 49–54. <https://doi.org/10.5437/08956308X5603936>
- Galli, P. (2005). IBM Gives 500 Patents to Open-Source Developers. <https://immagic.com/elibrary/archives/general/genpress/e050111g.pdf>
- Hall, B. H., & Helmers, C. (2013). Innovation and diffusion of clean/green technology: Can patent commons help? *Journal of Environmental Economics and Management*, 66(1), 33–51. <https://doi.org/10.1016/j.jeem.2012.12.008>
- Lagane, C. (2011). Google rachète plus de 1000 brevets à IBM. <https://www.Silicon.Fr/Google-Rachete-plus-de-1000-Brevets-a-Ibm-57683.Html>.
- Lerner, J., & Tirole, J. (2001). The open source movement: Key research questions. *European Economic Review*, 45(4–6), 819–826. [https://doi.org/10.1016/S0014-2921\(01\)00124-6](https://doi.org/10.1016/S0014-2921(01)00124-6)

- Maurer, S. M., Rai, A., & Sali, A. (2004). Finding cures for tropical diseases: Is open source an answer? *In PLoS Medicine (Vol. 1, pp. 183–186)*.
<https://doi.org/10.1371/journal.pmed.0010056>
- Pénin, J. (2008). More open than open innovation? Rethinking the concept of openness in innovation studies. *Working Papers of BETA 2008-18, Bureau d'Economie Théorique et Appliquée, UDS, Strasbourg*.
- Sapsalis, E., van Pottelsberghe de la Potterie, B., & Navon, R. (2006). Academic versus industry patenting: An in-depth analysis of what determines patent value. *Research Policy, 35(10), 1631–1645*. <https://doi.org/10.1016/j.respol.2006.09.014>
- Temiz, S. (2021). Open innovation via crowdsourcing: A digital only hackathon case study from Sweden. *Journal of Open Innovation: Technology, Market, and Complexity, 7(1), 1–14*. <https://doi.org/10.3390/joitmc7010039>

CONCLUSION GENERALE

L'objectif principal de cette thèse est d'étudier l'impact des technologies de traitement Big Data sur l'innovation des organisations dans le domaine du traitement de données. Cette analyse a été réalisée à travers l'étude d'une des principales technologies de traitement Big Data : MapReduce, technologie qui a fait l'objet d'un dépôt de brevet par l'entreprise américaine Google en 2004.

Dans le chapitre I, une étude scientométrique sur le brevet MapReduce a été réalisée afin d'étudier l'émergence des technologies de traitement Big Data. Elle a permis de mettre en évidence plusieurs caractéristiques de ce secteur. Tout d'abord, au niveau temporel, on a pu définir que la mise en application des technologies Big Data a été effective à partir des années 2010. En effet, à partir de cette date, les dépôts de brevets sur les technologies ayant une filiation directe ou indirecte avec la technologie MapReduce ont augmenté de +80 % en moyenne. Ensuite, au niveau de la localisation des technologies Big Data, il apparaît que 85 % des brevets ayant une filiation avec la technologie MapReduce sont déposés par des organisations américaines. De plus, on constate, qu'à partir de 2010, période de mise en application des technologies Big Data, des entreprises récentes telles que Google ou Amazon ont été les plus actives sur ces technologies. Enfin, les champs d'application des technologies Big Data sont concentrés dans les domaines de l'informatique et de la communication numérique (selon la classification WIPO).

Le Chapitre II s'est concentré sur l'étude des mécanismes d'accessibilité à la technologie MapReduce. En effet, la technologie MapReduce est accessible à travers deux mécanismes distincts :

- La fondation communautaire Apache et son projet open-source Hadoop
- Le dispositif de patent pledge « *Open Patent Non-Assertion Pledge* » de l'entreprise Google déployé en 2013.

Une analyse empirique a été menée afin d'évaluer si ces deux mécanismes ont eu un impact sur le nombre de brevets déposés ayant un lien de filiation avec la technologie MapReduce et si leur qualité technologique en a été affectée et mesurée par le nombre de citations reçues.

Lors de cette analyse, nous avons systématiquement distingué le nombre de citations avec ou sans autocitation afin d'évaluer, respectivement, si les brevets avaient permis d'accroître la valeur privée des entreprises (par des développements internes) ou la valeur sociale en

favorisant le développement d'innovations externes à l'entreprise. Les estimations agrégées sur le nombre de citations reçues par brevet, mettent en évidence que les brevets qui citent directement la technologie MapReduce ont permis d'augmenter la valeur privée et sociale des sponsors Apache platines.

Les estimations désagrégées sur le nombre de citations reçues par brevet et par année, mettent en évidence que sur la période 2004-2012, période antérieure au dispositif « *Open Patent Non-Assertion Pledge* », les brevets MapReduce ont permis d'accroître la valeur privée et sociale de l'entreprise Google, mais la mise en place du dispositif des pledge après 2013 n'a pas eu d'impact sur la valeur des brevets. En revanche, sur la période ultérieure à 2013 et le déploiement du dispositif d'ouverture directe des technologies MapReduce, ce sont les brevets « *follow-on innovations* » qui ont obtenu le plus de citations.

Le chapitre III est un prolongement du chapitre II avec comme objectif d'évaluer l'impact des patents pledge sur un périmètre plus large que celui du secteur des technologies Big Data. Dans cette optique, l'analyse empirique de ce chapitre a mesuré l'impact des pledge (incluant la technologie MapReduce) sur la valeur privée et sociale des brevets inclus dans le portefeuille de trois acteurs majeurs des TIC que sont Google, Microsoft et IBM. Les estimations incluant des brevets de contrôle ayant des technologies similaires indiquent que les brevets pledge n'ont pas eu le même impact suivant les entreprises. Tout d'abord pour l'entreprise Google, ce sont les brevets pledge dont elle a fait l'acquisition pour développer sa propriété intellectuelle qui a permis d'accroître sa valeur privée. Ensuite, pour IBM, l'augmentation de la valeur privée de la firme provient de l'ensemble de ses brevets pledge. Enfin, Microsoft est la seule entreprise étudiée qui voit sa valeur sociale croître du fait de ses pledge.

Concernant les estimations se concentrant uniquement sur le portefeuille de brevets de Google, IBM et Microsoft, il apparaît que les brevets pledge que Google a acquis contribuent à la valeur privée de la firme, cependant ce résultat n'est plus significatif si on considère un sous échantillon des TIC. Pour IBM, ce sont uniquement les brevets pledge acquis par la firme qui permettent d'accroître sa valeur privée. Enfin pour Microsoft, les brevets pledge n'ont plus d'impact positif significatif.

Sur l'ensemble des estimations (avec brevets de contrôles ou limitées aux portefeuilles des trois entreprises), les brevets sur la technologie MapReduce ne reçoivent pas significativement plus de citations.

Il s'avère que lorsqu'on élargit l'analyse à un périmètre plus large que celui du secteur des technologies Big Data et incluant d'autres brevets pledge, la technologie MapReduce n'est pas une technologie plus innovante par rapport à d'autres technologies.

En croisant les résultats du chapitre 2 et du chapitre 3, il apparaît donc que l'accessibilité aisée à la technologie MapReduce a été un élément déterminant pour en faire une des technologies majeures des technologies de traitement Big Data. Ce résultat n'est pas surprenant car la technologie MapReduce a été créée en 2004, période durant laquelle les entreprises commençaient à prendre conscience du traitement des données Big Data. Pour l'entreprise Google, cet enjeu était déjà majeur dans le cadre de l'amélioration de son moteur de recherche. En effet, les revenus publicitaires à travers son moteur de recherche ont été les premières sources de revenus de l'entreprise, ce qui est toujours le cas actuellement (source : [webrankinfo](http://webrankinfo.com)³⁷). En publiant le document de recherche « *MapReduce: Simplified Data Processing on Large Clusters* » (source : Google³⁸), Google a rapidement capté l'intérêt de nombreuses entreprises du secteur des TIC sur une technologie nouvelle capable de traiter une très grande quantité de données. C'est d'ailleurs à la suite de cette publication que l'ingénieur Doug Cutting a démarré l'élaboration de la technologie Hadoop (source : medium.com³⁹). C'est dans ce contexte que l'accessibilité à la technologie MapReduce a permis à cette technologie de devenir une technologie majeure des Big Data.

Depuis, la technologie MapReduce a été supplantée par des technologies de traitement Big Data plus récentes comme la technologie Spark (Gopalani & Arora, 2015). Il est intéressant de noter que cette technologie développée par un doctorant de l'université de Berkeley en Californie en 2009 a également été transmise à la fondation Apache en 2013. Il apparaît que les technologies de traitement Big Data ne font pas l'objet de droits de propriété intellectuelle fermée et qu'un consensus sur leur accessibilité est admis au sein du secteur. Dans son rapport "Data-driven innovation for growth and well-being" (*Data-Driven Innovation*, 2015), l'OCDE a défini les données comme une « infrastructure ». A travers ce terme, on comprend que les données sont un support, pouvant être les fondations de futurs développements. Dans cette optique, ces technologies sont essentielles pour utiliser cette « infrastructure » qui est la donnée. On pourrait

³⁷ <https://www.webrankinfo.com/dossiers/google/resultats-financiers>

³⁸ <https://static.googleusercontent.com/media/research.google.com/fr//archive/mapreduce-osdi04.pdf>

³⁹ <https://medium.com/@shraddhabag7583/meet-hadoop-d85795c2d587>

donc les assimiler à des biens publics, c'est-à-dire à un bien non rival ou non excluable, ce qui expliquerait ce consensus autour de leur accessibilité.

Au total, les travaux réalisés dans cette thèse vont venir enrichir la littérature économique portant sur les analyses de brevets grâce à l'étude de la technologie MapReduce et de ses brevets associés. A travers cette étude, il a été possible de caractériser un secteur émergent, le secteur des technologies Big Data. Ensuite, deux mécanismes originaux d'innovation ouverte ont pu être analysés : la fondation Apache et les patent pledge. Enfin, notre étude sur les patent pledge a été complétée par une analyse sectorielle sur trois acteurs majeurs du secteur des TIC.

Références

- Gopalani, S., & Arora, R. (2015). Comparing Apache Spark and Map Reduce with Performance Analysis using K-Means. *In International Journal of Computer Applications (Vol. 113, Issue 1)*.
- OECD. (2015). Data-Driven Innovation. <https://doi.org/10.1787/9789264229358-en>