



Patricia Rahmé

Open data locale : acteurs, pratiques et dispositifs

RAHME Patricia. *Open data locale : acteurs, pratiques et dispositifs*, sous la direction de Fidelia Ibekwe-SanJuan. - Lyon : Université Jean Moulin (Lyon 3), 2016.

Disponible sur : www.theses.fr/2016LYSE3045



Document diffusé sous le contrat Creative Commons « Paternité – pas d'utilisation commerciale - pas de modification » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.



N°d'ordre NNT : 2016LYSE3045

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
Université Jean Moulin Lyon 3

Ecole Doctorale 485
Education - Psychologie - Information et Communication

Discipline de doctorat : Sciences de l'Information et de la
Communication

Soutenue publiquement le 16/09/2016, par :
Patricia Rahmé

**Open data locale : acteurs, pratiques et
dispositifs**

Devant le jury composé de :

Françoise Albertini, Professeur des Universités, Université de Corse Pasquale Paoli,
Rapporteure
Fidelia Ibekwe-SanJuan, Professeur des Universités, Aix-Marseille Université, Directrice
Jean-Fabrice Lebraty, Professeur des Universités, Université Jean Moulin Lyon 3
Vincent Meyer, Professeur des Universités, Université de Nice Sophia-Antipolis, Rapporteur

*À mes parents,
leur support n'a d'égal que mon amour pour eux.*

REMERCIEMENTS

Tout d'abord, je tiens à remercier Madame le professeur Fidelia Ibekwe-SanJuan qui a encadré cette thèse. Je la remercie de tous ces conseils avisés et de m'avoir intégré dans le projet de recherche BOLD. Je remercie également Mme Françoise Albertini, M. Jean-Fabrice Lebraty et M. Vincent Meyer d'avoir accepté d'être membres de mon jury de thèse.

Je remercie également toutes les personnes qui ont accepté de me donner du temps pour réaliser mes entretiens, de leur intérêt, leurs conseils et leur aide.

Un grand merci à l'Université Libanaise qui m'a donné, dans le cadre de ma bourse universitaire, la chance et les moyens pour poursuivre mes études supérieures de Master et de Doctorat en France. Je remercie en particulier mes professeurs à l'Université Libanaise, Dolla Issa et Gladys Saadé pour leur soutien moral et leur encouragement permanent ainsi que la confiance qu'elles m'ont accordée.

Depuis mon arrivée en France fin 2010, j'ai passé à Lyon des années d'études qui m'ont apporté bien plus qu'un diplôme. J'y ai rencontré certains de mes amis les plus proches pour lesquels j'ai, à la fin de cette thèse, une pensée particulière à Marie Claire, Bassel, Salem, Alexa, Christie, François, Marie Rose, Gilbert, Julie et tous mes amis Libanais/Lyonnais. Un grand merci à tous les pères et les frères de la Communauté Antonine Maronite qui ont contribué à faire de mon séjour en France un cadre chaleureux et accueillant dans lequel je me suis toujours senti bien.

Les derniers remerciements vont à ma famille au Liban dont l'affection et le soutien m'ont donné le courage et la volonté d'avancer. Une pensée particulière à mon père qui a cru en moi et qui me porte dans ses prières au ciel.

TABLE DES MATIÈRES

REMERCIEMENTS	3
TABLE DES MATIÈRES.....	4
LISTE DES TABLEAUX.....	10
LISTE DES FIGURES.....	11
LISTE DES ACRONYMES	12
INTRODUCTION ET MISE EN CONTEXTE	15
1. Open data locale.....	16
2. Acteurs, pratiques et dispositifs au cœur d'un processus.....	18
2.1 Un processus open data.....	19
2.2 Des pratiques d'open data	19
2.3 Des stratégies	20
2.4 Des dispositifs numériques.....	20
2.5 Typologie des acteurs de l'open data	21
3. Présentation de la problématique.....	24
4. Recherche à double dimensions : théorique et socio-technique...25	
4.1 Dimension théorique	25
4.2 Dimension sociotechnique.....	27
5. Démarche hypothético-inductive.....	28
6. Hypothèses.....	29
7. Nécessité d'une approche interdisciplinaire	31
8. Le plan de la thèse.....	33

PREMIÈRE PARTIE : CADRAGE THÉORIQUE ET CONTEXTUEL DE LA RECHERCHE.....35

Chapitre 1. La donnée comme « objet » de production, de diffusion et de traitement	36
1.1 Notions de base : Donnée, Information, Connaissance et Sagesse	36

1.2 Données numérique, document numérique et données numériques structurées	41
1.3 Web de données	45
1.3.1 Web 2.0	46
1.3.2 Web de données liées	48
1.3.3 Open linked data.....	50
1.4 Big Data.....	52
1.4.1 La valeur des Big data	53
1.4.2 L'appropriation des Big data	61
Conclusion du chapitre 1	64
Chapitre 2. Le principe d'ouverture	66
2.1 Culture libre et cyberculture	67
2.1.1 Open source.....	69
2.1.2 Open access.....	71
2.1.3 Open Knowledge	72
2.2 Innovation ouverte	74
2.2.1 Crowdsourcing	76
2.2.2 Open data dans les entreprises	79
2.3 Gouvernement ouvert.....	80
Conclusion du chapitre 2	82
Conclusion de la première partie	84
DEUXIÈME PARTIE : OPEN DATA DANS LE SECTEUR PUBLIC	86
Chapitre 3 : Données publiques ouvertes : définition et intérêt.....	88
3.1 Une définition opérationnelle de l'open data	89
3.2 Le nouveau modèle du gouvernement ouvert.....	94
3.2.1 Open data et transparence	94
3.2.2 Open data, participation citoyenne et innovation.....	97
3.3 Données publiques et croissance économique	100
3.4 Open data et déploiement des TIC dans le secteur public.....	101

Conclusion du chapitre 3.....	105
Chapitre 4. Les initiatives politiques et sociales de l'open data	106
4.1 Expansion mondiale du mouvement Open data	108
4.1.1 Associations internationales.....	110
4.1.2 La Commission européenne et l'Open Data.....	112
4.2 Open data en France : un modèle d'interaction entre les acteurs.....	115
4.2.1 Politiciens et législateurs	115
4.2.2 Agents publics.....	119
4.2.3 Think tanks.....	123
4.2.4 Associations	124
Conclusion du chapitre 4	127
Chapitre 5. La communication et la diffusion des données publiques.....	131
5.1 Documents administratifs, informations publiques et données publiques	133
5.2 Système et acteurs traditionnels de la communication de données publiques	136
5.3 Evolution des modes d'accès : de l'accès sur demande à la diffusion proactive	140
5.4 Évolution des conditions d'usage	143
5.4.1 La réutilisation commerciale	143
5.4.2 De la redevance à la gratuité	145
Conclusion du chapitre 5.....	152
 TROISIEME PARTIE : OPEN DATA LOCALE	 154
Chapitre 6. Méthodologie de recherche.....	156
6.1 Recensement des projets open data locale	157
6.2 Etude de 5 cas en profondeur	161
6.3 Méthodes de collecte des données	165
6.3.1 Corpus de discours.....	165
I. La recherche et veille documentaire.....	166

II. Les conférences	168
III. Les entretiens semi-directifs	168
6.3.2 Corpus de dispositifs sociotechniques	174
6.4 Démarche qualitative.....	177
Conclusion du chapitre 6	180
Chapitre 7. Lancement du projet : réflexion et concertation des acteurs	183
7.1 Des acteurs en mouvement	184
7.1.1 Les acteurs internes : Élus et directeurs de services.....	184
7.1.2 Les acteurs externes : le mouvement associatif.....	187
7.2 Adaptation des acteurs et promotion de la culture de partage	192
7.3 Les risques de l’open data	195
7.4 Le choix de licence.....	198
7.5 Constitution d’une équipe de projet.....	201
Conclusion du chapitre 7	203
Chapitre 8. Identification des données publiques : quelles données ouvrir ?	205
8.1 Un choix stratégique	206
8.2 Un projet transversal	207
8.3 Externalisation des données publiques.....	210
8.4 La diversité des producteurs de données sur un territoire	212
8.4.1 La centralisation de la production et de la diffusion de données publiques au niveau de l’État.....	213
8.4.2 Les entreprises chargées d’un service public	215
8.4.3 Le crowdsourcing	216
Conclusion du chapitre 8	220
Chapitre 9. Dispositifs sociotechniques de l’open data	222
9.1 Structure et fonctionnement du dispositif	224
9.1.1 Un catalogue de jeux de données.....	226
9.1.2 Un outil de valorisation des données et des réutilisations..	229
9.1.3 Un outil de visibilité politique.....	230

9.1.4 Le public visé.....	231
9.2 Modalités de participation en ligne.....	232
9.3 Configuration du lien social	236
Conclusion du chapitre 9	239
Chapitre 10. Promotion et usages des données ouvertes	241
10.1 Les caractéristiques du Marketing de l’open data locale.....	243
10.1.2 Le produit : Les données publiques.....	243
10.1.3 La place : le territoire.....	244
10.1.4 Les techniques de promotion	245
10.2 Les orientations de la promotion	246
10.2.1 La promotion du projet.....	246
10.2.2 La promotion de « la culture des données ».....	246
10.2.3 La promotion de la réutilisation.....	249
10.3 Les usages sociaux des données ouvertes	252
10.3.1 Mesurer l’usage des données publiques ouvertes.....	253
10.3.2 Les logiques d’usage.....	254
10.3.3 Éléments fondamentaux d’usage.....	257
Conclusion du chapitre 10	264
Chapitre 11. La communauté de réutilisateurs et les modèles de réutilisations.....	265
11.1 Les profils des réutilisateurs.....	265
11.1.1 Les développeurs et les graphistes : deux acteurs mis en vedette	266
11.1.2 Les journalistes : réutilisateurs classiques de la donnée publique	268
11.1.3 Les acteurs économiques : une réutilisation encore floue.....	271
11.2 Les modèles de réutilisations	278
11.2.1 La recherche d’une donnée précise	279
11.2.2 La production graphique	280
11.2.3 La production écrite	280

11.2.4 La production d'interface	281
11.2.5 L'extraction ou la combinaison des données.....	281
Conclusion du chapitre 11	285
CONCLUSION GENERALE ET PERSPECTIVES	286
Synthèse de nos contributions.....	286
Les limites de la recherche.....	289
Prolongements possibles de cette recherche	290
BIBLIOGRAPHIE	294
Annexes.....	323

LISTE DES TABLEAUX

Tableau 1. Définitions de donnée, information, connaissance et sagesse synthétisées de Rowley, 2007.....	37
Tableau 2. Système de notation à 5 étoiles de Tim Berners-Lee (2010).	51
Tableau 3. Critères de l'open data	92
Tableau 4. L'intérêt de l'open data dans le secteur public	105
Tableau 5. Typologie des acteurs dans l'implémentation de l'open data	128
Tableau 6. Système traditionnel Vs Open data.....	152
Tableau 7. Les 10 premières initiatives open data locale	160
Tableau 8. Les projets open data étudiés	163
Tableau 9. Catégories des personnes interviewées et les structures auxquelles elles appartiennent	170
Tableau 10. Grille d'observation du dispositif sociotechnique (Jouët et Le Caroff, 2013).....	176
Tableau 11. Notre grille d'analyse.....	181
Tableau 12. Les avantages et les craintes de l'open data	197
Tableau 13. Architecture et fonctionnement des plateformes open data	224
Tableau 14. Modalités de participation sur les plateformes open data	232
Tableau 15. La configuration du lien social sur les plateformes open data	236
Tableau 16. Les fichiers les plus téléchargés sur les plateformes des collectivités	258
Tableau 17. L'écosystème des réutilisateurs des données publiques	276
Tableau 18. L'apport des acteurs dans le processus open data	291

LISTE DES FIGURES

Figure 1. L'hiérarchie ou pyramide DIKW (Rowley, 2007)	37
Figure 2. Capture d'écran d'un document en format CSV	45
Figure 3. The Data Science Venn Diagram (Conway, 2010)	61
Figure 4. Les trois champs du principe d'ouverture	66
Figure 5. Relation entre Open data et Open government.....	90
Figure 6. Les 66 pays membres du Partenariat pour un Gouvernement ouvert (Open Government Partnership, 2015).....	109
Figure 7. Évolution du nombre de textes de niveau législatif (Balcou, Cottin et Iteanu, 2011)	117
Figure 8. Influence des acteurs open data en France (Rahmé, 2015).....	130
Figure 9. Les phases du processus d'ouverture	154
Figure 10. L'évolution des initiatives locales (Rahmé, 2015).....	160
Figure 11. Les acteurs internes : Elus et fonctionnaires (Rahmé, 2015)	185
Figure 12. Page d'accueil de la plateforme open data de la ville de Montpellier	225
Figure 13. Grand Lyon SmartData V2 - Décembre 2013.....	226
Figure 14. V2 du portail open data 71 - Novembre 2012.....	226
Figure 15. Les compétences acquises par les lycéens (Projet #dataviz_rra, 2014).....	261
Figure 16. Article de la Gazette des Communes basé sur des données fournies par data.gouv.fr	269
Figure 17. Article paru dans la Gazette des Communes basé sur des données ouvertes par Direction Générale des Collectivités Locales	270
Figure 18. Article du Monde.fr basé sur des données ouvertes par le Ministère de l'Économie et des Finances.....	270
Figure 19. Exemple de Tableaux de bord des territoires (TBT) (Data Publica, 2014).....	273
Figure 20. Capture d'écran du mashup « les musées en France » prise le 15 août 2014 du site web labs.antidot.net.	284

LISTE DES ACRONYMES¹

AAF	Association des Archivistes Français
AEC	Aquitaine Europe Communication
API	Application Programming Interface
APIE	Agence du patrimoine de l'État
ASNE	Société Américaine des Editeurs de Journaux
BOLD	Big Open Linked Data
BnF	Bibliothèque National de France
CADA	Commission d'Accès aux Documents Administratifs
CC0	Creative Commons Zero
CERTU	Centre d'études sur les réseaux, les transports, l'urbanisme et les constructions publiques
CIL	Correspondant informatique et libertés
CIMAP	Comités interministériels pour la modernisation de l'action publique
CIO	Chief Information Officer
CNU	Conseil National des Universités
CNIL	Commission nationale de l'informatique et des libertés
CNNUM	Conseil National du Numérique
CNRS	Centre national de la recherche scientifique
COEPIA	Conseil d'orientation de l'édition publique et de l'information administrative
CTO	Chief Technology Officer
DaaS	Data as a service
DIKW	Data, Information, Knowledge, Wisdome
DILA	Direction de l'information légale et administrative
DSIIG	Direction des Systèmes d'Information et de l'Information

¹ Dans la thèse, nous ne développerons chaque acronyme que lors de sa première apparition

	Géographique
DUI	Délégation aux Usages de l'Internet
Elico	Équipe de recherche de Lyon en sciences de l'Information et de la Communication
Enssib	École nationale supérieure des sciences de l'information et des bibliothèques
FAQ	Frequently asked questions
Fing	Fondation Internet Nouvelle Génération
FSF	Free Software Foundation
GFII	Groupement Français des Industries de l'Information
GMF	Garantie Mutuelle des Fonctionnaires
HTML	Hypertext Markup Language
IGN	L'Institut géographique national
IJBA	Institut de Journalisme Bordeaux Aquitaine
INIST	Institut de l'Information Scientifique et Technique
Insee	Institut National de la Statistique et des Études Économiques
INSPIRE	Infrastructure d'information géographique dans la Communauté européenne Infrastructure for Spatial Information in Europe
LO	Licence Ouverte
MGI	McKinsey Global Institute
Modra	Matinées de l'open data en Rhône-Alpes
NMP	Nouveau Management Public
OCDE	Organisation de coopération et de développement économiques
ODbL	Open Database License
OKF	Open Knowledge Foundation
ONEMA	Office National de l'Eau et des Milieux Aquatiques
ONG	Organisations Non Gouvernementales

OSM	OpenStreetMap
OWNI	Objet Web Non Identifié
PACA	Région Provence-Alpes-Côte d'Azur
PAGSI	Le programme d'action gouvernemental pour la société de l'information
PDDL	Public Domain Dedication and License
PRADA ou PRADAQRIP	Personne responsable de l'accès aux documents administratifs et des questions relatives à la réutilisation des informations publiques.
RDF	Resource Description Framework
SGML	Standard Generalized Markup
SIC	Sciences de l'Information et de la Communication
SIG	Système d'Information Géographique
SISPEA	Système d'Information sur les Services Publics d'Eau et d'Assainissement
SNIIRAM	Système national d'information inter-régimes de l'assurance maladie
SPIC	Service Public à caractère Industriel et Commercial
TIC	Technologie de l'information et de la communication
UE	Union européenne
UIT	Union Internationale des Télécommunications
URL	Uniform Resource Locator
XML	Extensible Markup Language

INTRODUCTION ET MISE EN CONTEXTE

Le concept de l'« Open data » a connu une attention mondiale avec le lancement en Mars 2009 du portail open data « data.gov » du gouvernement américain. Cette première initiative prise au niveau international est devenue un catalyseur pour d'autres pays à se joindre au mouvement de données ouvertes. Le Royaume Uni a lancé en Janvier 2010 le portail Data.gov.uk, suivi par le portail data.gouv.fr ouvert en décembre 2011 en France. Selon Data.gov, en 2012, 31 pays dans le monde ont lancé leurs propres sites d'open data. Ce nombre a augmenté jusqu'au 41 pays en Mai 2013. Il s'agit certainement d'un mouvement mondial récent en rapide évolution.

Au sein de chaque pays, au niveau des villes ou des collectivités, des projets open data sont déjà lancés ou en cours de préparation. Un nombre croissant de collectivités régionales et territoriales lance d'ailleurs leurs propres initiatives. Dans un premier bilan en France, l'association nantaise LiberTIC qui a comme mission de promouvoir l'ouverture des données publiques, le gouvernement 2.0 et la démocratie électronique (e-démocratie), a recensé en Janvier 2012, 11 villes sur les 15 plus grandes villes de France qui ont ouvert leurs données. Ce chiffre est en évolution rapide et plus de collectivités se sont engagées dans des projets Open data. L'attention croissante portée à la valorisation des données publiques ouvertes se manifeste par le lancement de plusieurs portails open data par les collectivités locales. Cette dynamique rencontre un mouvement de fond qui traverse la France depuis fin 2009.

L'histoire du mouvement est relativement courte. En effet l'open data, considéré comme « phénomène *politico-socio-culturel* d'une ampleur sans précédent » (Frochot, 2014), est devenu un sujet d'actualité au cœur des débats politiques, économiques, culturels et juridiques. L'open data a fait « irruption » (Rauzier, 2011, p.4) dans plusieurs grandes villes françaises, suivies par quelques Conseils généraux. Les porteurs des projets open data sont convaincus de l'apport de l'open data, des associations citoyennes revendiquent et militent pour l'ouverture des données publiques et des entreprises s'empressent d'exploiter les opportunités de l'open data (Bluenove, 2011). Le Monde.fr (2011) décrit l'open data comme : « le "must" de l'administration "new look" et de l'open governance ». Une administration moderne et intelligente mettra à disposition ses données

publiques non accessibles et permettra une meilleure compréhension des politiques publiques par les citoyens.

L'émergence de ce concept dans les gouvernements, rempli de promesses et de valeurs, est basée sur des principes fondamentaux² comme la divulgation proactive et gratuite des données dans des formats ouverts. Ces principes sont porteurs de plusieurs significations et interprétations et leur application pose de nouveaux défis en matière de gestion et de diffusion des données publiques. Toutes les nouvelles pratiques qui se développent très vite et les transformations apportées sur le système traditionnel d'accès à l'information publique, indépendamment de leur caractère durable ou non, méritent d'être soigneusement analysées.

1. Open data locale

Le mouvement le plus large de l'open data recouvre les données détenues par des particuliers ou des organisations qu'ils soient publiques ou privés. Les revendications pour un open data portent sur les données du secteur publique comme celles du secteur privé. En effet, dans son discours à la conférence « The next web » en 2009, en Californie, Tim Berners-Lee a demandé de mettre en ligne des données brutes provenant de divers sources : « Il y a les données gouvernementales, les données d'entreprises, qui sont vraiment essentielles, les données scientifiques, les données personnelles, les données météorologiques, les données concernant les événements, les données des discussions, les informations, toutes sortes de données ».

Dans le travail de recherche que nous restituons ici, nous nous intéressons en particulier à l'ouverture des données portées par une institution publique, donc aux données publiques dans le sens des données produites ou reçues par une autorité administrative dans le cadre de sa mission de service public. Le secteur public est particulièrement concerné à cet égard, tant en raison de la quantité des données qu'il produit, mais surtout à cause d'un devoir d'informer et d'une responsabilité de transparence prévus par la loi d'accès aux informations publiques. Plus particulièrement, notre champ

² Que nous présentons plus précisément dans les chapitres 1 et 2.

d'étude se concentre essentiellement sur l'open data locale en France qui repose sur les projets Open data des collectivités territoriales.

Le choix de centrer notre étude sur les acteurs de l'open data au niveau local n'est pas anodin. L'analyse du processus de diffusion et d'usage des données publiques au niveau des collectivités est un chantier très récent et encore peu exploré. En effet, dans le secteur public, la décentralisation des compétences des différentes collectivités territoriales définies par les lois de décentralisation de 1982 et 2002 a un double impact sur la politique locale. D'un côté, la question de diffusion des données reste une initiative volontaire prise par les collectivités car la législation actuelle repose sur une logique de demande d'accès³ aux données publiques. Les collectivités qui adoptent la logique du mouvement Open data vise à diffuser en ligne les données sans demande préalable par le citoyen. De l'autre, les collectivités ont acquis de nouvelles responsabilités et doivent posséder davantage l'esprit d'ouverture⁴, le savoir faire, l'expertise et le financement pour gérer l'ouverture des données. Le choix apparaît comme une nouvelle épreuve pour les collectivités. Cette étude montrera les enjeux et les bonnes pratiques pour que les collectivités territoriales puissent relever ce défi.

Une autre raison qui nous a incité à étudier l'open data locale est la multiplication des initiatives et des acteurs impliqués dans l'Open data en France depuis la première initiative lancée par Rennes Métropole en 2010. D'après le site Keolis.Renne.com (2013), Rennes Métropole a permis de déclencher un mouvement citoyen de grande ampleur sur le territoire Français. La première version de la carte Open data en France en Aout 2010 montre 9 collectivités dont 3 ont ouvert leurs données et 6 envisagent de le faire. En Août 2013, le nombre a augmenté à 43 collectivités dont 38 collectivités territoriales ont lancé leurs propres plateformes Open data ou hébergé leurs données soit sur la plateforme nationale data.gouv.fr ou sur les plateformes des régions et 5 collectivités sont en cours d'ouvrir leurs données⁵. Cette même année, le site Data.gouv.fr de la plateforme française d'ouverture des données publiques a fait référence seulement à 21 collectivités considérées comme des initiatives d'ouverture des données

³ Que nous présenterons dans le chapitre 2.

⁴ Que nous expliquons dans le chapitre 4.

⁵ <http://www.opendata-map.org/map>

publiques en France⁶. Dans le dernier état des lieux de l'open data réalisé par Joumana Boustani (2013), elle recense au total, entre hébergement dédié et mutualisé, 54 collectivités qui ont ouvert leurs données en Janvier 2013.

La période entre 2011 et 2014⁷ couvre une période critique et un temps crucial pour le mouvement Open data en France : plus de 60 collectivités ont ouvert leurs portails open data. Etalab, la mission gouvernementale chargée de l'ouverture des données publiques à l'échelon national, a vu le jour en Décembre 2011. Un ensemble d'acteurs territoriaux s'y sont associés et ont créé en 2013 l'association Open Data France.

Bien que le contexte local représente un cadre spécifique, il nous semble intéressant de mettre en relief les initiatives prises par les territoires locaux par rapport à d'autres initiatives aux échelles nationale, européenne et internationale. En effet, le mouvement d'ouverture des données publiques est un mouvement mondial qui influence les actions des acteurs à tous les niveaux et le contexte local de l'open data est fortement dépendant du contexte national, européen et international. Ce postulat conduit notre recherche vers une étude du terrain français en le comparant avec d'autres études menées dans d'autres pays. Ceci permet d'avoir une vision globale en s'intéressant aux dynamiques qui organisent le local avec le national et l'international.

2. Acteurs, pratiques et dispositifs au cœur d'un processus

Nous considérons l'open data comme un processus suffisamment large qui recouvre l'ensemble de pratiques, de stratégies, et de dispositifs techniques mis en œuvre par les différents acteurs impliqués.

Il nous semble essentiel de présenter dès maintenant les notions centrales que nous emploierons couramment dans notre étude. Ces notions sont, de notre point de vue, fondamentales pour étudier l'ouverture des données publiques dans la discipline de sciences d'information et de communication (SIC).

⁶ <http://www.data.gouv.fr/>

⁷ Qui coïncide avec le début de notre thèse.

2.1 Un processus open data

Le terme processus désigne bien un enchaînement ordonné de faits, de phénomènes, d'opérations ou d'actions, répondant à un certain schéma et aboutissant à quelque chose⁸. Par conséquent, s'intéresser à l'ouverture des données comme processus suppose de prendre en considération en premier les modes de production et de collecte de données, ensuite les méthodes de diffusion et les dispositifs utilisés, et enfin les modèles d'usage des données publiques numériques.

Dans notre étude, l'open data ne se résume pas à la simple publication des données des administrations publiques et nous ne séparons pas l'action de production de celle de diffusion et d'usages des données. Dans cette optique, pour désigner le mot anglo-saxon « open data », nous nous référons au terme français « ouverture des données » plutôt que « données ouvertes » pour signaler évidemment ce processus d'ouverture. Cette diversité de perspectives est une richesse qui nous pousse à étudier tous les enjeux interdépendants tant au niveau de la production des données dans les collectivités, les méthodes de collecte, les modèles d'accès, de diffusion et de présentation des données par les dispositifs techniques d'open data, que les modèles d'usage des données. Les notions de production, collecte, accès, diffusion et usage des données seront discutées par la suite dans les chapitres qui leurs seront dédiés.

2.2 Des pratiques d'open data

Comme le rappelle Gardiès, Fabre et Couzinet (2010), le terme « pratique » vient du verbe grec *prattein* signifiant « agir » qui donne praxis, « action ». Le concept de « pratique informationnelle » est employé en SIC, que ce soit pour désigner ce que les acteurs font avec l'information ou pour signifier la manière dont ils utilisent les dispositifs techniques intermédiaires. Ce que nous restituerons dans ce manuscrit, est donc principalement les « pratiques informationnelles » définies par Chaudiron et Ihadjadene (2010) comme « la manière dont un ensemble de dispositifs, de sources formelles ou non, d'outils, de compétences cognitives sont effectivement mobilisés, par un individu ou un groupe d'individus, dans les différentes situations de production, de recherche, d'organisation, de traitement, d'usage, de partage

⁸ Une définition donnée par le dictionnaire de Français Larousse

et de communication de l'information». Pour Souchier, Jeanneret et Le Marec (2003), étudier les pratiques informationnelles devrait « prendre en compte l'épaisseur sociale de la pratique en construction » pour intégrer la question des contextes et saisir la complexité de ce qui est étudié. Pour eux les pratiques résultent de personnes qui assument à la fois des styles, des enjeux individuels et des engagements sociaux et professionnels.

En nous appuyant sur la définition de Chaudiron et Ihadjadene et la réflexion de Souchier, Jeanneret et Le Marec, nous souhaitons étudier les pratiques des acteurs dans les différentes situations du processus open data et dans son contexte social. Dans le cadre de notre recherche, ces pratiques sont imbriquées plus précisément à la notion de données - que ce soit au niveau de la production, de la recherche, de la diffusion, du traitement ou d'usage - avec lesquelles elle entretient des liens très spécifiques. En plus, à ces pratiques, nous ajoutons des pratiques complémentaires comme les pratiques de promotion des données reposant sur nos observations de terrain et discutées dans les chapitres 10 et 11. Enfin, il paraît important de préciser que les actions autour de l'open data peuvent faire référence selon nous à la notion de « nouvelle pratique » du fait qu'elles s'insèrent dans un phénomène récent et nous ne pouvons pas les considérer en tant qu'actions habituelles ou nées d'une tradition.

2.3 Des stratégies

Au sens propre du terme, une stratégie est définie comme l'art de coordonner un ensemble d'actions en vue d'atteindre un but précis⁹. Dans notre recherche, nous proposons de définir les stratégies mises en place par des acteurs opérant des actions en relation avec l'open data. Nous nous intéressons en particulier au développement d'une stratégie-réseau définie par Marcon et Moinet (2010) comme une stratégie qui « consiste à créer ou, le plus souvent, à activer et orienter les liens tissés entre des acteurs dans le cadre d'un projet plus ou moins défini ». Si le réseau est un mode d'organisation et d'action, cette stratégie amène par ailleurs à la création d'un réseau d'acteurs autour du projet open data.

2.4 Des dispositifs numériques

⁹ Dictionnaire de Français Larousse.

Nous donnons ici de très brèves définitions du terme « dispositifs numériques » employé dans cette recherche¹⁰. Lorsque nous parlons de « dispositifs », nous faisons référence à la notion de dispositif étudié en SIC comme objet matériel médiateur qui désigne « l'ensemble des substrats matériels de la communication » (Jeanneret, 2007). Or l'ouverture des données dans un contexte numérique sur le web est fortement liée aux nouveaux dispositifs numériques de communication et d'accès à l'information. L'ouverture met en place, grâce à ces dispositifs numériques sur le web, des interfaces de diffusion et d'accès aux données. Ainsi, dans notre recherche, les dispositifs numériques désignent les plateformes Open data. Sur le plan de la réalisation technique, les plateformes intègrent souvent des briques logicielles spécialisées de stockages, de recherche et de visualisation de données. Bullich et Guignard (2012) définissent une plateforme numérique comme « *un système de distribution de biens et services immatériels qui ne trouvent leur existence que sur les réseaux, nécessitant de fait une connexion. Ce dispositif est fondamentalement lié aux avancées techniques dans les domaines du traitement de l'information numérique (processeur et logiciel), des capacités graphiques et de stockage des terminaux (...)* ».

Il s'agit également d'intégrer dans notre réflexion sur les dispositifs socio-techniques la technologie et les principes du web 2.0 qui, selon O'Reilly (2005), offrent intrinsèquement plus d'avantages aux plates-formes en « délivrant des logiciels comme des services continuellement mis à jour et qui s'améliorent plus ils sont utilisés, consommant et remixant des données en provenance de multiples sources, y compris des utilisateurs, tout en fournissant leurs propres données et services sous une forme qui permet le remix par d'autres, créant des effets de réseau à travers une « architecture de participation », et allant au-delà de la métaphore de la page Web 1.0 pour offrir des expériences utilisateurs riches ». Nous accordons alors dans notre étude une attention particulière aux dispositifs étudiés considérés comme des « dispositifs socio-techniques » émergents du web social dans leurs aspects de partage, collaboration et de co-production de données, sur lesquelles nous reviendrons par la suite.

2.5 Typologie des acteurs de l'open data

¹⁰ Le chapitre 9 sera ensuite dédié à une analyse plus précise des dispositifs numériques de l'open data.

L'open data locale prend forme au sein d'un écosystème dynamique composé de plusieurs acteurs. En faire une typologie primaire des acteurs potentiels relève plus d'une volonté de clarification quant aux termes que nous emploierons par la suite pour présenter notre problématique. Nous présentons donc ici les différentes catégories d'acteurs ainsi qu'un bref descriptif de leurs spécificités.

2.5.1 Producteurs et diffuseurs de données

Les producteurs de données publiques représentent et façonnent l'offre de données à ouvrir. Le secteur public est évidemment un acteur dominant et stratégique dans l'ouverture des données publiques. Les établissements publics, au niveau national et local, forment une source première de données publiques. Ils produisent et détiennent des données qui sont des traces numériques générées suite à des opérations, projets, services, études, ou réunions. Il faut ensuite admettre le rôle de « diffusion de données publiques » assurée par les établissements publics. En tant que source, producteur et diffuseur, un établissement public devient alors un porteur du projet open data définissant ses objectifs, sa réalisation, ses règles, sa durée et le budget consacré au projet. Au niveau de l'Etat, la mission Etalab est chargée de la diffusion des données des administrations par la gestion de la plateforme française data.gouv.fr. Au niveau local, des collectivités territoriales, se transforment en diffuseurs de données en libérant leurs données sur leurs portails open data locaux.

2.5.2 Utilisateurs des données

Précisons tout d'abord les notions d'« usager » et d'« utilisateur » de données publiques. Le Coadic (1997, p.19) relève une distinction entre ces deux notions en précisant que « usage a de fait un sens plus général. L'utilisation est synonyme d'usage pratique. C'est l'action, la manière de faire servir une chose à une fin précise ». Lorsque nous explorons les études consacrées aux usagers de systèmes d'information, nous relevons l'adoption de la notion d'utilisateur qui se rapporte à une appropriation des dispositifs techniques de communication et d'accès à l'information. Ainsi, dans un article sur l'usage et ses modèles, Joëlle Le Marec (2001) précise que l'individu est d'abord utilisateur d'un système et la maîtrise du dispositif technique lui donnera alors le statut d'usager.

Louise Merzeau (2010, p.2) montre bien la responsabilité croissante de « l'utilisateur numérique », qui selon lui, doit devenir un « acteur » dans le réseau informationnel en ligne. Cette responsabilité découle de la multitude de statut d'utilisateur : « *Tantôt récepteur (de messages), utilisateur (d'outils), consommateur (de services et de produits) ou producteur (de ressources et de valeurs), le digital user semble de fait capitaliser toutes les postures que les systèmes d'information et de communication peuvent assigner aux usagers* ». Le terme usager est donc assez large pour englober une pluralité de postures info-communicationnelles. Dans cette thèse, nous choisirons de garder la notion « d'utilisateur » afin de désigner les utilisateurs des plateformes open data. Nous distinguons par la suite entre les différentes postures qu'un utilisateur open data pourrait adopter.

Cette deuxième catégorie d'acteurs open data comprend tout individu ayant accès aux données en ligne. Ce dernier possède d'avantage le statut de citoyen, non seulement par ce qu'il est membre d'une ville ou plus généralement d'un Etat, mais surtout par ce qu'il jouisse du droit d'accès aux données publiques protégé par la loi. À partir justement de cette globalisation, les utilisateurs peuvent avoir des profils différents. Ces profils d'utilisateurs seront discutés dans le chapitre 11.

2.5.3 Militants pour l'open data

Cette dernière catégorie comprend les acteurs militants pour l'ouverture des données publiques. Ces derniers ne se placent ni dans des situations de production ni d'utilisation de données publiques mais assurent une promotion du concept et de la philosophie de l'open data auprès des producteurs/diffuseurs et des utilisateurs. Sont admissibles dans cette catégorie, des activistes et des associations s'étant engagés dans la lutte pour l'ouverture des données publiques. Vu que l'open data est un mouvement mondial, cette catégorie comprend des acteurs nationaux et internationaux qui accompagnent l'émergence de nouvelles pratiques.

Ces quelques notions ayant été posées, nous présentons maintenant notre problématique.

3. Présentation de la problématique

Les différents éléments de contexte présentés dans cette introduction font partie intégrante de notre projet de recherche et de son élaboration. La question centrale qui s'est imposée et qui a guidé notre réflexion sur le sujet de l'Open data locale est :

Quel est le positionnement des acteurs dans le processus d'ouverture des données publiques locales ?

Ce questionnement a été initié par la caractéristique du mouvement open data qui fonctionne dans un « système social » avec un tissu riche en acteurs avec de fortes ambitions : les citoyens, les activistes, les associations, les entreprises, les développeurs, les spécialistes des données, les fonctionnaires du secteur public, les politiciens. L'expression « système social » désigne un ensemble complexe de parties et de pièces en interaction, qui forment ensemble un système Open data. L'association Fondation Internet Nouvelle Génération (Fing), dans son guide pratique sur l'ouverture des données publiques territoriales, publié en 2011, a dressé le panorama des acteurs impliqués sur un territoire : « la collectivité publique invite d'autres acteurs à intervenir : des associations locales qui connaissent bien les besoins d'une population, des médias régionaux, des jeunes entreprises inventives, des citoyens ». Open Society Foundations fait référence à la société civile, en particulier les « civil hackers », les fonctionnaires gouvernementaux et les politiciens comme étant les composantes fondamentales ou les acteurs clés pour une ouverture des données réussie (Hogge, 2010).

Dans ce contexte, la thèse se propose d'examiner l'organisation, l'influence, la stratégie et le rôle de chaque acteur impliqué dans l'ouverture des données publiques considérée comme un processus et un ensemble d'activités de production, de diffusion et d'usage. Elle présentera les interactions entre l'objet, qui représente les données publiques ouvertes, et les acteurs ainsi que les interactions entre ces derniers dans leurs expériences de l'ouverture des données publiques. Nous nous intéressons aux stratégies d'adoption et d'implémentation de ce nouveau concept en représentant les différentes interprétations mais aussi aux stratégies de résistance des acteurs du système traditionnel de diffusion publique. Nous

essayons surtout d'identifier les enjeux, les opportunités, les freins et les limites de l'ouverture des données au niveau local.

4. Recherche à double dimensions : théorique et socio-technique

En cohérence avec nos objectifs, notre recherche s'inscrit dans une perspective qui regroupe deux axes nécessairement corrélés l'un à l'autre : théorique et socio-technique. Ces deux dimensions posent plusieurs questions qui découlent directement de notre problématique de recherche.

4.1 Dimension théorique

Pour comprendre les comportements des acteurs, il faut non seulement étudier les pratiques qu'ils développent mais aussi les idéologies sociétales et techniques sous-jacentes à leurs pratiques. Le concept d'ouverture des données apparaît comme une extension d'autres idéaux et les enjeux théoriques de notre étude sont multiples. En premier, il s'agit de définir ce concept considéré comme flou ou ambiguë (Chignard, 2012) et de préciser ces principes de bases.

Notre étude vise de même à expliciter l'open data et montrer ses origines et ses effets. En fait, le discours sur les données émanant du secteur public est à placer dans un contexte dans lequel baignent déjà plusieurs mouvements et utopies : « la société de l'information », « l'administration électronique », « le gouvernement ouvert », « la gestion des connaissances dans le gouvernement », « la transparence », « la participation citoyenne ». L'ouverture des données s'inscrit également dans le mouvement du « libre accès » (en anglais Open Access) émanant de l'Open Access Initiative de Budapest en 2001. L'open data s'inspire du mouvement libre accès qui défend, surtout dans le domaine de la recherche scientifique, la mise à disposition gratuitement en ligne de contenus numériques sous une licence d'utilisation dite libre. De même, la notion de « logiciel libre » s'applique aux logiciels avec la possibilité de libre réutilisation et d'accès au code source. Au sens plus large, le mouvement « open data » s'intègre dans le mouvement du « savoir libre » (Open Knowledge). D'autres discours, plus politique, sur l'Open data, le rattachent à la conception de « transparence ». Dans le

secteur de management, l'Open data est considéré comme outil d'« innovation ». Notre cadrage théorique, cherche à conceptualiser la notion d'Open data et à éprouver son rapport à ces autres concepts.

Diversité des notions mobilisées

La question de l'open data donne lieu à des discours qui emploient une terminologie variée qui reflète, d'une part, la diversité des concepts et donc des notions qui leur sont afférentes et, d'autre part, elle met en évidence une fluctuation terminologique des termes mobilisés. Nous relevons ainsi de notre corpus analysé un recours aux notions de « données », « informations », et « documents ». Nous trouvons ainsi des expressions telles que : « données brutes » « données structurées », « données non structurées », « données textuelles », « web de données » et « big data ». Ces notions nécessitent que l'on s'y arrête pour les expliciter.

En outre, le mot « open », selon le discours des acteurs, renvoie aux notions comme d'« ouverture », de « diffusion », d'« accessibilité », ou de « réutilisation ». Nous mettons en exergue l'évolution, au fil du temps, de la terminologie utilisée, de celle de l'« accès », la « diffusion » et la « réutilisation ». Le terme anglophone « open data » est le plus fréquemment emprunté et communément partagé par tous les acteurs. Pour autant, sa définition et les contours de ce qu'il recouvre ne sont pas précisément posés. Il est tellement utilisé qu'il est, finalement, assez peu explicité ou défini. Il est mis en œuvre pour désigner un concept ou une action sans qu'il n'y ait nécessairement d'ancrage théorique précis qui accompagne les pratiques des acteurs. Notons que la Commission générale de terminologie et de néologie, en réaction un peu tardive à l'usage courant de l'« open data » dans les discours et les publications officielles, a adopté et publié en Mai 2014, au Journal Officiel¹¹, le terme « données ouvertes » comme équivalent du terme anglais « Open data ». Cette variation et ce flou terminologique nous incitent à préciser ces notions et les inscrire dans notre champ disciplinaire. Nous cherchons également à étudier l'« ouverture des données » en la considérant comme un processus et un ensemble d'actions. En pratique, cette activité recouvre-t-elle quels types de « réalités » ?

¹¹ Voir : <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000028890784&dateTexte=&categorieLien=id>.

4.2 Dimension sociotechnique

Bien que le mouvement open data a été largement accepté au niveau local et international, sa mise en pratique semble complexe et réclame l'intervention de plusieurs acteurs. Il demeure un concept malléable, qui sous-entend plusieurs aspects et applications. Tout en considérant que l'open data est un processus qui recouvre l'ensemble d'actions ou de stratégies, de pratiques et de dispositifs qui peuvent être mis en œuvre par les différents acteurs, nous étudions l'open data par la réunion des deux dimensions qui la composent : une dimension sociale et une dimension technique.

Une approche socio-technique demeure alors la mise en correspondance de la technologie avec celle de stratégies d'acteurs sociaux. Selon Flichy (1995), l'approche socio-technique permet d'intégrer, dans une même analyse, technique et société et voir comment interfèrent de nombreux mondes sociaux. De plus, les recherches s'appuyant sur cette approche portent sur l'action technique, sur les intentions, les projets, les délibérations qui précèdent l'action, sur le déroulement de l'action et surtout, sur l'interaction des différents acteurs et de l'objet technique. L'approche socio-technique sert de fondement à notre recherche car elle permet de suivre de façon fine tous les enjeux en relation avec le processus d'ouverture des données publiques numériques sur des portails open data. Ce n'est pas uniquement les outils techniques, mais aussi les actions sociales, les démarches, les interactions, les usages et les partenariats entre les acteurs, qui sont étudiés.

La dimension sociale nous paraît donc indispensable à l'analyse des activités des acteurs en lien avec l'ouverture des données. Ce type de recherche suppose une analyse sociale à plusieurs niveaux. Il s'agit, en premier lieu, de définir les acteurs impliqués et d'étudier *la stratégie de chaque acteur* et ses activités en lien avec l'ouverture des données, qu'il soit impliqué dans des activités de production, de promotion, de diffusion ou d'usages. Ensuite, il est pertinent d'élargir notre champ d'investigation au *contexte englobant et structurant les activités de tous les acteurs en relation*. Les activités des acteurs sont ainsi appréhendées de manière globale en relation avec l'environnement social. Les activités d'ouverture des données sont considérées comme des activités sociales, influencées par divers facteurs inhérents au contexte et à la situation dans lesquels elles prennent place. Un

environnement qui contraint, influence et structure les stratégies, les pratiques et les besoins des acteurs. Vu que l'open data est un mouvement mondial, les activités des acteurs sont alors étudiées en considérant les facteurs d'influence et le rapport entre le local, le national et l'international. Ainsi, nous cherchons à étudier la nature des relations établies entre les acteurs de différent ou du même caractère. Existe-t-il des objectifs communs qui les incitent à mutualiser leurs expériences et leurs expertises par rapport à un problème ou à un besoin clairement identifié ? Les enjeux concurrentiels sont-ils aussi importants entre les acteurs de même nature ? Quel type de partenariat se développe-t-il entre les acteurs ?

Si la dimension sociale représente les motivations, les objectifs plus larges et les interactions entre les acteurs, au sein du contexte et du terrain auxquels ils appartiennent, la dimension technique recouvre, quant à elle, des questions techniques en liaison avec la mise en œuvre d'outils de diffusion, les formats, la conception de systèmes de traitement et de présentation de données. Les activités techniques se matérialisent dans des projets et des actions qui varient d'une simple création d'une page web vers la mise en œuvre d'applications, des bases de données et des plateformes de recherche. Cette approche nous permet d'identifier les dispositifs existants en relation avec l'ouverture des données et de présenter leurs spécificités et leurs caractéristiques. Nous nous interrogeons sur les modalités d'accès, de structuration et d'organisation des données (dans les collections de fichiers de données ou à l'intérieur des fichiers eux-mêmes). En plus, vu que l'open data est incarné dans la logique du web 2.0, nous accordons, également, une attention particulière aux dispositifs étudiés considérés comme des « dispositifs socio-techniques » émergents du web social dans leurs aspects de partage, collaboration et de co-production des données. Enfin, il reste à faire configurer les liens entre les outils, qui rendent opérationnels certains modèles de stratégies d'acteur et les modes d'appropriation de ces outils par les acteurs concernés.

5. Démarche hypothético-inductive

Glaser et Strauss (1967), fondateurs de Grounded Theory, proposent l'approche inductive pour faire la recherche en sciences sociales. Ils soulignent l'importance de la mise en relation de données issues du terrain afin de proposer par la suite des théories issues de la pratique et de

l'observation. La démarche inductive semble adéquate à notre recherche du fait que l'open data est un mouvement récent ce qui nous conduit de partir des constats de terrain pour produire des modèles d'analyse du processus d'implémentation de l'open data locale en déterminant la position de chaque acteur.

Cependant, pour répondre à notre problématique, nous n'employons pas qu'un raisonnement purement inductif qui néglige toute déduction. En effet, nous procédons déductivement en utilisant les principes de l'open data comme cadre d'analyse pour mener à bien le processus inductif et pouvoir analyser les phénomènes et pratiques observés. De même, comme nous l'avons précisé dans l'introduction, notre attention s'est portée sur l'open data suite à la multiplication des portails open data et parce qu'il s'agit d'un sujet d'actualité et d'intérêt à plusieurs acteurs. Avant de commencer notre recherche, nous avons donc déjà abordé le terrain de l'open data et effectué quelques recherches bibliographiques sur le sujet ou des sujets connexes.

Dans notre étude, les approches inductives et déductives ne sont pas incompatibles puisque comme le précisent Anadón et Guillemette (2007, p.33) « *En somme, toute approche généralement inductive implique nécessairement des moments de déduction sans perdre son caractère essentiellement inductif, celui-ci provenant de l'orientation fondamentale qui consiste à étudier les phénomènes à partir de l'expérience propre des acteurs* ». Ainsi nous avons opté à une démarche hypothético-inductive définie notamment par Crozier et Friedberg (1992), comme une « *démarche en somme qui se sert de l'expérience vécue des participants pour proposer et vérifier des hypothèses de plus en plus générales sur les caractéristiques de l'ensemble* ». L'intérêt premier de cette démarche a donc été pour nous la possibilité d'alterner ou de superposer induction et déduction. Le choix de cette démarche inductive modérée, nous permet donc de mettre en avant certaines hypothèses dans le paragraphe qui suit.

6. Hypothèses

Comme nous l'avons souligné, une approche hypothético-inductive nous permet de vérifier certaines hypothèses préalablement posées. Les hypothèses que nous allons présenter ont été formulées au début de la recherche et constituent de véritables points de départ que notre recherche

permettra de valider ou d'infirmer. Ainsi, « *pour valider une hypothèse, il faut traduire ce que l'on cherche en indicateurs qui peuvent être des données, des préférences, des opinions, des jugements, et/ou des pratiques...* » (Olivesi, 2007). Nous avons formulé trois hypothèses. Celles-ci guident en quelque sorte le lecteur au travers des chapitres de la thèse.

1. La première hypothèse suppose que l'open data donne lieu à une nouvelle distribution des rôles des acteurs impliqués tout au long du processus d'ouverture. Les pratiques de chaque acteur ne se limitent pas à une étape du processus d'ouverture des données - que ce soit de production, de diffusion, de promotion ou d'usage - ce qui crée un système mixte d'open data. Cette hypothèse est issue de nos premières observations d'une dynamique complexe dans laquelle jouent plusieurs acteurs. Cette hypothèse sera traitée dans les chapitres de la troisième partie qui restitue les principales phases du processus d'ouverture et la place de chaque acteur.
2. Nous émettons également l'hypothèse que l'offre des données par les producteurs de données publiques est limitée de sorte qu'elle ne réponde pas à plusieurs besoins différents. Une première observation des usages, met en avant le fait que les opérations et les actions réalisées sont concentrées autour des « geeks » pour développer essentiellement des applications web et mobiles. Cette hypothèse nous permettra de mener une double réflexion sur le système d'offre des collectivités et de son adéquation avec les besoins des utilisateurs qui seront par ailleurs traités dans les chapitres 10 et 11.
3. La troisième hypothèse s'intéresse particulièrement aux dispositifs techniques utilisés pour la diffusion des données publiques. Ces derniers sont considérés comme des « dispositifs socio-techniques » émergents du web social dans leurs aspects de médiations, partage, collaboration et de coproduction de données. Cette hypothèse est issue des réflexions théoriques menées sur le web 2.0 dans le chapitre 1 et sera par ailleurs discutée dans le chapitre 9 en étudiant les plateformes open data.

Ces différentes hypothèses seront discutées dans la conclusion générale en les articulant avec les principaux apports de cette thèse.

7. Nécessité d'une approche interdisciplinaire

Notre recherche s'inscrit dans le cadre des Sciences de l'Information et de la Communication (SIC). Comme nous l'avons montré tout au long de cette introduction, nous nous intéressons à la place qu'occupent les données dans le processus d'ouverture. La 71e Section du Conseil National des Universités (CNU) définit l'équipement intellectuel spécifique aux SIC qui recouvrent particulièrement¹² :

- A. « Les études sur les notions d'information et de communication, sur leurs relations, sur la nature des phénomènes et des pratiques ainsi désignés, de même que les différentes approches scientifiques qui s'y appliquent ». Dans notre recherche, nous définissons plus précisément les notions « données », « accès » et « diffusion ».
- B. « L'étude, d'une part, des processus, des productions et des usages de l'information et de la communication, d'autre part, de la conception et de la réception de celles-ci. Ainsi que l'étude des processus de médiation et de médiatisation ». Notre recherche porte effectivement sur l'étude de l'open data comme un processus de production, de collecte, de diffusion et d'usages des données publiques.
- C. « L'étude des acteurs, individuels et institutionnels, de l'information et de la communication, l'étude des professionnels (dont notamment les journalistes) et de leurs pratiques ». Notre étude s'intéresse particulièrement à l'ensemble des acteurs impliqués dans l'open data.
- D. « L'étude de l'information, de son contenu, de ses systèmes sous l'angle des représentations, des significations ou des pratiques associées ». Notre recherche porte en partie sur l'étude des données ouvertes et la présentation de ces propres caractéristiques.

Pour être plus précis, l'ouverture des données publiques est un champ de recherche relativement récent à exploiter et étudier en sciences sociales¹³. Ces transformations récentes peuvent être incluses sous l'axe de recherche sur les informations publiques qui est, de manière générale, peu traité dans les travaux de recherche en sciences de l'information et de la communication

¹² Les extraits suivants sont issus du site web du CNU : <http://www.cpcnu.fr/web/section-71>

¹³ Nous distinguons ici l'Open data au sens informatique, attaché au mouvement Open source initié dans les années 1980, de l'Open data dans le cadre d'ouverture des données publiques.

surtout en le comparant avec les études sur la publication des recherches scientifiques.

Au-delà des SIC, l'étude de l'open data s'insère dans une approche interdisciplinaire. Une telle approche interdisciplinaire, selon Miège (2000), « permet de relier, autour d'axes de recherche, si possible bien spécifiés, des méthodologies provenant de disciplines différentes et de les faire interagir ». Pour Apter (2010, p.8), « la première étape vers l'interdisciplinarité intervient lorsqu'il y a emprunt d'idées à plusieurs contextes d'analyse, chacun se référant à une discipline spécifique. La seconde intervient lorsque ce qui a commencé comme la recherche de solution à un problème précis se prolonge par une exploration théorique d'où émergeront de nouvelles combinaisons qui vont modifier le corpus théorique et conceptuel de la discipline concernée ».

Dans notre cas, notre réflexion s'appuie sur une exploration des champs disciplinaires afin d'articuler les apports théoriques ou les réflexions d'auteurs ayant rapport avec le sujet de l'ouverture des données publiques. En effet, l'ouverture des données publiques soulève des problèmes juridiques, politiques, économiques, et techniques. Par exemple, les pratiques d'ouverture sont dictées par des textes de lois. La littérature juridique traite plusieurs problématiques en relation avec les données publiques : le droit de l'information¹⁴, la propriété intellectuelle et le droit d'auteur, la protection des données personnelles, les licences, la commercialisation des données publiques¹⁵, ainsi que des critiques et commentaires sur les textes de loi et les décrets sur l'accès et la réutilisation des informations publiques¹⁶. La science politique, est également sollicitée, à travers différents concepts et théories tel que la transparence de l'état, la démocratie participative. Nous faisons appel aussi aux Sciences de gestion car l'open data, comme tout autre projet au sein d'une collectivité, est issu d'une stratégie et d'un mode de gouvernance et de gestion dans le secteur public. Un projet qui a des objectifs précis en précisant sa dimension financière, ces moyens humains et techniques et qui nécessite une

¹⁴ Voir les travaux de Marie-Anne Delègue, Jean-Michel Bruguière, Jean-Pierre Balcou, Stéphane Cottin, et Olivier Iteanu sur le droit de l'information.

¹⁵ Voir l'étude de Philippe Gaudrat sur la Commercialisation des données publiques.

¹⁶ Voir l'étude de Bertrand Du Marais sur la Diffusion des données publiques et révolution numérique.

évaluation dans le cadre de la politique publique. Une autre discipline concernée est l'informatique notamment pour comprendre le concept de réutilisation par machine, le traitement informatique des données, les systèmes d'information et les applications informatiques.

Notre étude permet de montrer que cette interdisciplinarité constatée dans la littérature est effectivement transposée sur le terrain. Cette interdisciplinarité se transpose au terrain d'action quand il s'agit de la mise en œuvre des projets open data. Des acteurs divers ont travaillé pour la mise en place de la plateforme des données publiques. Nous repérons des profils différents : chef de projet, élus, développeurs, géographes, informaticiens, juristes, graphistes. De plus, les acteurs intéressés par l'usage des données ouvertes forment un ensemble aux compétences et intérêts très différents.

Nous avons essayé de réunir tous ces champs de recherche pour donner lieu à une philosophie non restrictive sur l'ouverture des données publiques et analyser la dynamique de cette philosophie sur le terrain en prenant la France comme terrain d'étude. Cette approche interdisciplinaire permet une appréhension globale de l'activité d'ouverture des données et une mise en évidence des six dimensions qui la structurent : sociale, économique, juridique, organisationnelle, politique et technique. Notre objectif est de traiter la problématique dans son ensemble, en identifiant et en intégrant toutes les relations entre les différents éléments impliqués.

8. Le plan de la thèse

Après cette introduction générale, notre étude comporte onze chapitres articulés autour de trois grandes parties.

Dans une première partie, nous définirons le cadre théorique et conceptuel de la notion d'open data. Pour ce faire, nous étudions les deux notions « Open » et « Data » qui composent ce néologisme. Ainsi, dans un premier chapitre, nous définirons la notion « d'Open » et, dans un deuxième chapitre, nous interrogerons la notion de « Data ». Dans cette première partie, nous définirons d'autres éléments contextuels (Open source, Open government, Open innovation), conceptuels et théoriques (comme l'information) ou encore notionnels (Big data et web de données).

Dans une deuxième partie, nous présenterons l'open data dans le secteur public. Le chapitre 3 visera à délimiter l'open data dans le secteur public tel qu'il est abordé par les différents acteurs impliqués dans l'ouverture des données publiques. Puis, nous présenterons, dans le chapitre 4, plusieurs initiatives mettant en œuvre des dynamiques en évolution afin d'identifier les acteurs aux niveaux international et national et par la suite au contexte local. Notre intérêt dans le cinquième chapitre est de découvrir le système de communication et de diffusion des documents publics mis en place avant la montée du mouvement open data. Nous proposons alors, de montrer, d'une part les modalités d'accès et de diffusion et d'autre part, les conditions d'usage des documents publics en distinguant les notions « d'accès » et de « diffusion ».

Enfin, dans une troisième et dernière partie, nous présenterons notre méthodologie de recherche et en détaillerons les résultats. Ces résultats nous offriront l'opportunité de présenter et analyser, d'une part, la stratégie des collectivités territoriales, producteurs et diffuseurs de données publiques, et de l'autre l'utilisation des données publiques par les utilisateurs. Ainsi, dans le chapitre 6, nous présenterons notre méthodologie et terrain de recherche. Puis, sur la base de notre étude de terrain, nous décrirons les étapes de mise en œuvre d'un projet open data pour pouvoir identifier le déroulement du processus et ressortir les enjeux de son implémentation. Nous discuterons de la phase de réflexion et de concertation des acteurs impliqués dans le lancement de l'initiative open data dans un territoire en analysant les attentes, les appréhensions et les craintes des parties concernées par le projet, ses opportunités et limites (Chapitre 7). Et, nous mettrons en évidence les facteurs qui influent sur le choix des données à ouvrir en s'intéressant aux modalités de la production et de la gestion des données publiques locales (Chapitre 8). Ensuite, nous présenterons notre analyse des dispositifs de diffusion des données publiques considérées comme dispositifs de nature socio-technique (Chapitre 9). Après, nous présenterons notre analyse sur les stratégies de promotion des données publiques adoptées par les collectivités territoriales par rapport aux usages sociaux des données ouvertes (Chapitre 10). Pour finir, nous présenterons, dans un dernier chapitre (Chapitre 11) le profil des réutilisateurs et les modèles de réutilisation repérés.

PREMIÈRE PARTIE : CADRAGE THÉORIQUE ET CONTEXTUEL DE LA RECHERCHE

Cette première partie de notre travail vise à définir plus précisément le cadre théorique et contextuel de notre recherche. Notre objectif est d'éclaircir et d'étayer la notion d'open data à travers une identification des principaux concepts associés aux deux termes « data » et « open ».

Le chapitre 1 présentera les notions et bases théoriques relatives à l'étude de l'open data en mettant l'accent sur les notions de « donnée », « information » et « connaissance ». Ainsi que les notions de « données numériques » et « données structurées » que nous articulons avec les concepts de « Big data » et « Web de données » qui définissent le contexte dans lequel évolue la notion de donnée. Le chapitre 2 définira le terme « Open » en mettant l'open data au cœur de trois concepts clés : Open Source, Open innovation et Open government. Ainsi nous constituons un point de rencontre de trois communautés bien distinctes en montrant leurs articulations avec l'open data.

Nous signalons que cette partie fait la synthèse des principales définitions des concepts connexes trouvées, excluant celles de « documents administratifs », « informations publiques », « données publiques » qui feront l'objet d'une description plus contextuelle quand nous abordons l'open data dans le secteur public discuté dans la deuxième partie.

Chapitre 1. La donnée comme « objet » de production, de diffusion et de traitement

La « donnée » est un terme devenu si courant dans nos sociétés qu'il paraît aujourd'hui difficile de le définir. Elle est aussi floue et arbitraire, de sorte que nous nous retrouvons avec un champ de terminologie variée qui, dans certains cas, la remplace par des notions plus larges (information, connaissance) ou dans d'autres cas, la décompose en termes plus étroits et restrictifs (données brutes, données numériques, données structurées). Il apparaît important de différencier ces termes pour en avoir une meilleure compréhension et distinguer les différences qu'il y a entre eux. C'est ainsi comme point de départ de nos réflexions que nous souhaitons dans ce premier chapitre définir la notion de donnée mais surtout en la situant par rapport à d'autres concepts et notions. L'hétérogénéité des significations de la notion de donnée nous invite à l'aborder avec une approche qui articule les multiples champs disciplinaires ayant abordés les données.

Nous présenterons, tout d'abord, la notion de donnée par rapport aux deux concepts d'information et de connaissance, ainsi que d'autres dans lequel cette notion de donnée s'insère. Ensuite, nous mettons en perspective la notion de données structurées et réutilisables qui s'inscrit dans l'approche informatique en s'appuyant sur les notions de « document numérique » et de « document structuré ». Enfin, nous nous intéresserons au web de données et au Big data qui définissent le contexte dans lequel évolue la notion de donnée en mettant en exergue les enjeux de production, de diffusion et de traitement des données.

1.1 Notions de base : Donnée, Information, Connaissance et Sagesse

Tout d'abord, le concept de donnée dans la littérature est pensé en complémentarité avec ceux d'information, de connaissance et de sagesse ; il prend sens dans ce jeu de différences qui constituent le socle fondateur de la pensée du Knowledge Management¹⁷. Les relations entre ces concepts sont

¹⁷ Le concept de gestion des connaissances – ou knowledge management selon la forme anglaise beaucoup plus usitée - est une discipline académique en émergence qui s'est développé dans les entreprises dans une préoccupation

souvent représentées sous forme de classement hiérarchique. En effet, l'hiérarchie DIKW (Data - Information - Knowledge - Wisdom) désigne une classe de modèles pour représenter les relations structurelles et / ou fonctionnelles supposées entre les données, informations, connaissances, et la sagesse (Voir figure 1). Comme l'écrit Rowley (2007), l'hiérarchie DIKW, appelée aussi le pyramide de la connaissance, est la plus utilisée et citée pour distinguer les notions de donnée, d'information, de la connaissance et de la sagesse dans la littérature sur la gestion d'information, la gestion des connaissances et les systèmes d'information.

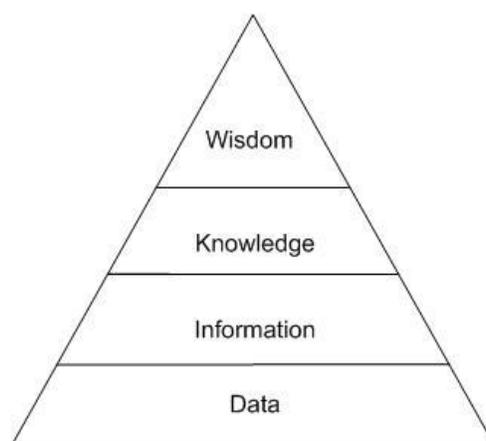


Figure 1. L'hiérarchie ou pyramide DIKW (Rowley, 2007)

La sagesse est située au sommet de la hiérarchie et les données forment l'élément de base de la pyramide. Chaque catégorie de la pyramide s'inscrit dans celle au-dessus. Rowley s'interroge précisément sur les distinctions et les relations à établir entre donnée, information, connaissance et sagesse ; il identifie dans la littérature une série de définitions que nous les synthétisons dans le tableau 1 ci-dessous.

Tableau 1. Définitions de donnée, information, connaissance et sagesse synthétisées de Rowley, 2007.

Sagesse
<ul style="list-style-type: none"> - La sagesse est des connaissances cumulées qui nous permet de comprendre comment appliquer des concepts d'un domaine à des nouvelles situations ou des problèmes (Jessup et Valacich,

de sauvegarder la mémoire et les connaissances collectives et de capitaliser les compétences de ses membres. Voir : Stankosky, Mickel. Creating the discipline of Knowledge Management. Elsevier, 2005.

2003).

- La sagesse est la capacité d'agir de manière critique dans une situation donnée. Elle est basé sur le jugement éthique liées au système de croyance d'un individu (Jashapara , 2005, pp. 17-18).

Connaissance

- La connaissance est le savoir-faire et est ce qui rend possible la transformation de l'information en instructions (Ackoff., 1989)
- La connaissance est la combinaison de données et d'informations, à laquelle est ajouté l'opinion d'experts, les compétences, et de l'expérience, pour aboutir à un atout précieux qui peut être utilisé pour faciliter la prise de décision (Chaffey et Wood, 2005, p. 223)
- La connaissance est une donnée et / ou information qui ont été organisées et traitées et qui se transforme en compréhension, expérience, apprentissage accumulé, et en expertise telles qu'elles s'appliquent à un problème ou une activité actuelles (Turban et al., 2005, p . 38)

Information

- Les informations sont des données qui ajoutent une valeur pour la compréhension d'un sujet (Chaffey et Wood, 2005, p . 233).
- Les informations sont des données qui ont été mise en forme d'une façon significative et utile aux êtres humains (Laudon et Laudon , 2006, p . 13)

Donnée

- Les données sont dépourvues de sens ou de valeur, car elle sont sans contexte et interprétation (Jessup et Valacich 2003 ; Bocij et al., 2003 ; Groff et Jones, 2003).
- Les données sont des symboles qui représentent les propriétés d'objets, d'événements et de leurs environnements. Elles sont produites suite à une observation due à une sensation (Ackoff, 1989)
- Les données sont des faits ou des observations objectives

discrètes, qui ne sont pas transformés et ne transmettent aucune signification particulière (Awad et Ghaziri, 2004 ; Chaffey et Wood, 2005 ; Pearlson et Saunders, 2004 ; Bocij et al., 2003).

- Les données sont une description élémentaire et enregistré des choses, des événements, des activités et transactions (Laudon et Laudon, 2006 ; Turban et al., 2005 ; Lenoir et al., 2005)

Pour comprendre les différences entre donnée, information, connaissance et sagesse, le modèle à deux dimensions de Blamisse (voir figure 2) montre le passage d'une donnée à une information, d'une information à une connaissance puis d'une connaissance à la sagesse en tenant en compte les deux facteurs de compréhension et d'indépendance au contexte. Blamisse met en avant la différence entre une information et une donnée en précisant qu'une information est par définition une donnée interprétée. En d'autres termes, la mise en contexte d'une donnée crée de la valeur ajoutée pour constituer une information.

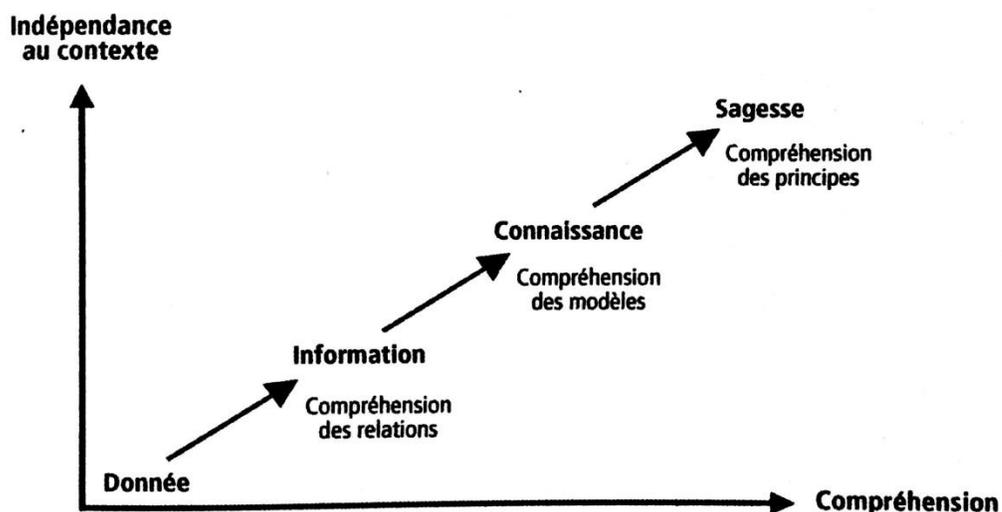


Figure 2. Différences entre donnée, information et connaissance (Blamisse, 2005, p. 15)

Les différentes définitions de la notion de donnée, présentées ci-dessus, recouvrent parfois des significations contradictoires. Selon Henry (1974), le premier auteur à distinguer parmi les données, informations et connaissances, les données sont simplement des faits bruts, cependant, pour Blamisse un fait relève plus d'une information qui met en relation

différentes données. Pour Rosenberg (2013), la « donnée » sert une fonction rhétorique : alors qu'un fait ou une preuve perdent leur qualité lorsqu'ils se révèlent faux, une donnée reste une donnée même lorsqu'elle est contredite.

De même, la définition de donnée de Ackoff (1989) (voir tableau 1) pose la problématique de la méthode de production et de collecte de donnée. Selon Ackoff, les données sont produites suite à une observation due à une sensation, ce qui contredit la définition de donnée comme des observations objectives discrètes, qui ne sont pas transformés et ne transmettent aucune signification. Au sens de Zins (2007), toute observation est subjective d'où la notion de données subjectives. Ainsi, les données sont conçus comme des stimulants sensoriels parce qu'elles sont perçues par nos sens.

La notion de données « brutes » souvent présumées objectives est aussi à interroger. Les données brutes appelées également données primaires sont considérées comme des données non interprétées émanant d'une source primaire et qui n'ont été soumises à aucun traitement ou toute autre manipulation. En effet, selon Bowker (2000), la notion de données brutes reste empreinte d'un certain flou où comme le souligne Gitelman (2013), un terme ambigu qui, lorsque les pratiques qui y sont associées sont étudiées de près, s'apparente à un oxymore. Cela pour dire que les données brutes n'existent pas en tant que telles car elles relèvent d'un traitement et d'une transformation en données cuisinées¹⁸. Bruno Latour (THATCamp Saint-Malo, 2013) aborde dans le même sens lorsqu'il affirme que les données « brutes » sont un mythe : les données, comme les « faits » scientifiques, sont fabriquées. Comme l'explique A. Walford (2013), « les données brutes en sciences sont des données en devenir, des entités ambiguës, multiples, qui attendent la série des traitements qui vont assurer la consolidation progressive de résultats scientifiques par leur inscription dans un réseau sociotechnique stabilisé »¹⁹. Ces traitements sont vus comme des opérateurs de mutation qui remette en cause la notion de données brutes. Même si le recours à la notion de données brutes pour représenter les données collectées de sources primaires recueillies à partir d'un capteur technique, le

¹⁸ Pour désigner le terme anglais « cooked data »

¹⁹ Cité par : Denis, Jérôme, and Samuel Goëta. "La Fabrique Des Données Brutes. Le Travail En Couloirs de L'open Data." In *Penser L'écosystème Des Données. Les Enjeux Scientifiques et Politiques Des Données Numériques*. Paris, 2013. <https://halshs.archives-ouvertes.fr/halshs-00990771/document>.

caractère changeant et fluide des données rend difficile aux auteurs à distinguer une temporalité ou une forme qu'on pourrait qualifier de « brute » dans le cycle de vie des données.

Au-delà de l'approche contextuelle liée au sens et à l'interprétation humaine des données, la notion de donnée est selon nous inhérente à la théorie mathématique de l'information de Shannon²⁰ visant à quantifier l'information présente dans un ensemble de données. S'intéresser à l'information au sens de Shannon, c'est donc prendre en compte l'approche physique, au-delà du sens de l'ensemble des symboles et signes qu'elle véhicule, et d'autre part considérer le champ technologique dans lequel elle surgit. Cette approche physique perçoit les données comme unités binaires dites bits²¹ et constituent à ce titre l'unité de mesure de base de l'information en informatique.

Pour conclure, nous nous appuyerons de manière égale sur les deux approches de données: celle du sens dans un contexte d'interprétation humaine, donc une donnée porteuse de sens, et l'approche physique qui perçoit les données comme unités binaires, ce qui nous permet d'aborder dans le paragraphe qui suit les données dans leur champ numérique.

1.2 Données numérique, document numérique et données numériques structurées

Si l'on considère l'approche informatique, la notion de données sera strictement liée à son format numérique : dans ce cas, une donnée désigne une donnée numérique définie comme « la représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement »²². La donnée est donc de l'information formatée pour pouvoir être traitée par un système informatique. Dans le champ de recherche sur les systèmes d'information, Baskarada et Koronios (2013) précise que les données sont des signes physiques, elles n'ont pas de sens parce qu'elles résident à

²⁰ Shannon, C. E. "A Mathematical Theory of Communications." *The Bell System Technical Journal* 27 (1948).

²¹ Le bit est l'unité la plus simple dans un système de numération, ne pouvant prendre que deux valeurs, désignées le plus souvent par les chiffres 0 et 1.

²²Instruction générale interministérielle sur la protection du secret et des informations concernant la défense nationale et la sûreté de l'État n° 1300/SGDN/ PSE/SSD du 25 août 2003

l'extérieur d'un esprit humain. Dans ce cas, l'information (ou le sens) émerge seulement à travers le traitement cognitif des données.

Pour étudier la notion de donnée numérique, nous souhaitons mettre en avant certains apports des réflexions théoriques menées sur les notions de « document » et « document numérique ». Ces apports théoriques nous permettront de définir ce que c'est par la suite une donnée structurée en signalant en quoi la structure d'un document numérique influe sur la nature même des données inscrites dans ce document. Selon Le Coadic (1994. p.6) le « Document » est le terme générique désignant les objets porteurs d'information. Un document est tout artefact qui représente ou exprime à l'aide de signes graphiques et iconiques (mots, images, diagrammes, cartes, figures, symboles), audio et vidéo (inscrits sur un support papier ou électronique), un objet, une idée ou une information. On le décline en document papier et en document électronique selon le type de support ». A partir de cette définition, le document que ce soit sur papier ou numérique, permet donc l'inscription d'un ensemble de signes que nous définissons ici comme données.

La définition de la notion de données numériques structurées nous amène notamment à présenter la théorie du document numérique de Pédaque²³ (2003) qui propose une définition du document numérique en articulant trois dimensions : le document comme *forme, signe et medium*.

1. Dimension « forme » : « *Un document numérique est un ensemble de données organisées selon une structure stable associée à des règles de mise en forme permettant une lisibilité partagée entre son concepteur et ses lecteurs* ».
2. Dimension « signe » : « *un document numérique est un texte dont les éléments sont potentiellement analysable par un système de connaissance en vue de son exploitation par un lecteur compétent.* ».
3. Dimension « medium » : « *Un document numérique est la trace de relations sociales reconstruite par les dispositifs informatiques* ».

²³ Un travail collectif signé du pseudonyme Roger T. Pédaque pour Réseau Thématique Prioritaire-DOC

Dans notre recherche, nous nous intéresserons à la « forme » ainsi qu'à la dimension « signe » du document numérique en asseyant d'assurer un équilibre entre le contenant et le contenu.

Comme un ensemble de données organisées selon une structure stable associée à des règles de mise en forme²⁴. Pédauque (2003) propose alors l'équation suivante :

$$\text{Document numérique} = \text{structure} + \text{données}$$

Nous adoptons cette équation pour mettre en relief la notion de « données structurées ». En effet, au sens informatique, la notion de « données structurées » est fortement liée à la notion de « document structuré » et aux différents formats informatiques qui organisent les données selon une structure précise. Selon Rekik (2001, p.19) un document structuré est « une représentation d'un document sous la forme d'une structure de données informatiques que l'on peut stocker, manipuler et échanger ». La structure de données adoptée pour la représentation d'un document dépend de sa nature et de son organisation interne. Ce sont les relations entre les éléments d'un document qui définissent la structure logique du document.

Autour de la notion de document structuré, une ingénierie s'est développée, proposant des langages, des techniques et des outils appropriés dans le but de faciliter la réutilisation des données par un traitement automatique et efficace du contenu des documents par un logiciel. En effet, la réutilisation des données a fait déjà l'objet des recherches en ingénierie des documents. Nous citons entre autres (Chikh, 2003) et (Rekik, 2001) qui proposent des méthodologies de réutilisation et des techniques de modélisation, de production et de manipulation des documents structurés qui favorisent leur réutilisation. Selon Rekik (2001, p.13), l'utilisation des technologies standards et non propriétaires basées sur le concept de document structuré a apporté des solutions pour pallier aux inconvénients des formats propriétaires non-structurés qui limitent considérablement les possibilités de traitement, de réutilisation et d'échange de documents.

²⁴ Il convient de préciser que la forme du document numérique n'est pas toujours lisible pour l'humain. D'autres formes de documents sont seulement utilisables et lisibles par les machines.

Les trois propriétés des données numériques réutilisables sont alors les suivants :

- 1) des technologies standards qui désignent des formats de données interopérables.
- 2) des formats non propriétaires (ou format libre ou ouvert) dont les spécifications techniques ne sont pas contrôlées par une entité privée ce qui permet de créer des logiciels qui puissent lire ou écrire ce type de format, par opposition à un format propriétaire ou fermé)
- 3) des formats structurés que nous les présentons ci-dessous selon leurs degrés de structuration : structurés, non structurés et semi-structurés.
 - Les données structurées sont conformes à la structure formelle de modèles de données²⁵ associés à des bases de données relationnelles ou d'autres formes de tableaux de données.
 - Les données non structurées ne sont pas des données mal organisées mais ce sont des données qui ne peuvent pas être traitées facilement par un ordinateur. Un document pdf est un exemple de documents propriétaires non structurés dans le mesure où il est difficile pour un ordinateur d'en traiter le contenu.
 - La technologie des documents structurés s'est principalement accélérée en 1998 dans le cadre de la technologie *Extensible Markup Language* (XML). Le XML est une extension du format *Standard Generalized Markup Language* (SGML) normalisé par l'ISO en 1986. Le format XML est considéré comme un format semi-structuré car il n'est pas conforme à la structure formelle de modèles de données associés à des bases de données relationnelles ou d'autres formes de tableaux de données, mais contient néanmoins des balises ou d'autres marqueurs pour séparer les éléments sémantiques et de faire respecter les hiérarchies des documents et de champs dans les données. Par ailleurs, les documents XML sont devenus largement utilisés comme format de publication de données sur le web que nous discuterons par la suite. Parmi d'autres formats semi-structurés, nous citons JSON (JavaScript Object Notation) et CSV (comma-separated values).

²⁵ En informatique, un modèle de données est un modèle qui décrit de façon conceptuelle comment sont représentées les données dans une organisation ou une base de données.

La mise en forme des données structurées associe, selon nous, les dimensions « forme » et « signe » des documents numériques en signalant que les données structurées sont potentiellement analysable par un système informatique donc une machine que directement par l'être humain. La figure ci-dessous montre une capture d'écran d'un document en format CSV. Ceci nous permet de dire qu'un document numérique n'est pas toujours porteur d'informations donc ayant un sens et lisible par les êtres humains. La figure 3 qui est une capture d'écran d'un document en format CSV est un bon exemple.

```
Communes;Codes_Insee;total_allocataires;total_ALF;total_ALS;tot
l_locataire;total_proprietaire;total_allocataires_logement
HORS TERRITOIRE DE COMPETENCE DE LA CAF;99999;33348;2275;8015;2
COMMUNES INCONNUES INSEE;XXXXX;7275;595;1713;1393;489;1678;1345
L'ABERGEMENT-CLEMENCIAT;01001;103;;;11;;;;;;7;12;19
L'ABERGEMENT-DE-VAREY;01002;31;;;;;;
AMBERIEU-EN-BUGEY;01004;2351;274;288;870;;;816;;;54;1326;106;14
AMBERIEUX-EN-DOBES;01005;220;20;17;20;;;14;;;6;43;14;57
AMBLEON;01006;16;;;;;;
AMBRONAY;01007;317;34;13;46;;;39;;;7;77;16;93
AMBUTRIX;01008;82;;;;;;15
ANDERT-ET-CONDON;01009;34;;;;;;8
```

Figure 2. Capture d'écran d'un document en format CSV

Enfin, cette présentation de la notion de données numériques met en relief sa structuration comme facteur central qui permet une réutilisation et un traitement plus facile par les ordinateurs. Dans ce contexte, nous proposons une première définition de la réutilisation des données qui recouvre « l'utilisation des données numériques structurées en vue d'un traitement informatique plus facile ».

1.3 Web de données

Ce paragraphe vise à définir plus précisément le terrain de production et de diffusion des données numériques, spécifique à notre recherche, qui est le web. Nous proposons de revenir brièvement sur l'évolution du web et ses technologies, et de présenter globalement ce que sous-tend les notions de « web 2.0 ou social » et « web de données liées ou web sémantiques ». Nous mettons en relief d'une part la manière dont les dispositifs sociotechniques du web 2.0 redéfinissent les pratiques informationnelles des usagers du web, et de l'autre, les caractéristiques des données produites et partager sur le web.

1.3.1 Web 2.0

Depuis vingt ans, le Web et ses technologies ont profondément transformé les pratiques informationnelles et médiatiques en introduisant de nouveaux modes de production et d'échange de l'information. La version 2.0 de celui-ci, apparue au début des années 2000 et que l'éditeur Tim O'Reilly considère comme une évolution technique, repose sur un ensemble de plates-formes continuellement mises à jour en s'appuyant sur les données produites par leurs utilisateurs grâce à des « architectures participatives ». Dans la perspective du web 2.0 ou social, l'utilisateur est pleinement engagé dans des processus de production et de diffusion de contenus en ligne. Cette capacité des utilisateurs se traduit concrètement par l'appropriation des techniques considérées comme accessibles et faciles à utiliser pour offrir aux internautes les possibilités de « créer, remixer, et partager des contenus » (Millerand, Proulx et Rueff, 2010, p.15). Plus qu'un simple dispositif technique de production et de publication, le web 2.0 revêt un aspect « participatif ». Grâce à des plateformes participatives, les utilisateurs peuvent coproduire ou modifier un contenu existant sur une page web. L'internaute n'est donc plus un simple consommateur d'information, il devient un coproducteur de contenus interagissant avec d'autres internautes. Nous reviendrons dans le chapitre 9 sur le concept du web social pour son apport à notre recherche dans l'étude du design des plateformes open data.

3.1.1 Les communautés en ligne

Ces nouvelles formes de collaboration nous amènent à questionner la notion de « communauté en ligne » (online community) ou « communauté virtuelle ». Comme nous l'avons vu, le web social repose sur l'idée que ses utilisateurs vont y développer des actions collectives en créant des regroupements d'internautes au sein de dispositifs sociotechniques du web 2.0 (comme les plateformes collaboratives) qualifiés de « communautés en ligne ».

Les communautés en ligne ont été caractérisées par Licklider et Taylor comme des communautés d'« intérêt commun ». Comme le rappelle Rieder (2010), la communauté repose sur un *commun* (des idées, valeurs et signes partagés qui établissent la particularité du groupe) qui vient alors alimenter l'identité collective du groupe. Latham et Sassen (2005) argumentent qu'une communauté consiste en « une identité commune autour d'objectifs

communs et des relations réciproques entre les participants, et que cette identité, ces buts et une telle réciprocité représente un aspect important et de fond de la vie professionnelle ou personnelle de chacun des participants »²⁶. Selon Rieder (2010), des membres cherchaient effectivement à établir une telle communauté en ligne dans le sens de la définition donnée plus haut, cependant très peu des groupes observés en ligne affichent un tel degré d'intégration. C'est le cas, pour l'auteur, « dans les plates-formes du web social qui accueillent souvent des publics très hétérogènes dont la majorité des relations sont faibles, éphémères et dispersées. L'émergence d'une véritable identité commune est de loin l'exception et non la règle ». Ainsi que le note Serge Proulx (2006), il est inadéquat de faire recours à la notion de « communauté virtuelle » pour désigner un collectif d'internautes en ligne étant constitués « par défaut » comme la communauté de Facebook par exemple, ou le fait de s'inscrire sur un forum. De ces différents constats, nous proposons aussi l'approche de Gensollen (2004) sur les communautés virtuelles qui s'organisent autour de la production ou la consommation de biens informationnels. Comme le note Lev-On et Manin (2006), la principale raison de rejoindre une communauté virtuelle étant le désir d'obtenir aisément des informations. Nous prenons, comme exemple d'une communauté de production d'information, la communauté de Wikipedia. Ainsi ses membres, appelés wikipédiens, contribuent à la conception et la création d'une encyclopédie en ligne par des pratiques d'édition et de production d'informations. Une étude (Glott, et. al, 2010) sur les membres de cette communauté a mis en avant deux raisons pour contribuer à Wikipedia : le souhait de partager les connaissances et le désir de corriger les erreurs.

De cette première approche du web, nous pouvons donc retenir que le web 2.0 permet à l'internaute de devenir un acteur de production de données. Cette production de données inhérente aux fonctionnalités des dispositifs sociotechniques prend la forme d'une collaboration collective qui crée de la valeur en contribuant à la construction de connaissances. Cette collaboration a permis la création de communautés en ligne qui sont des regroupements d'internautes motivés par des intérêts communs et s'organisent autour de la production ou la consommation de biens informationnels. Cette définition,

²⁶ Cité dans Rieder, 2010.

pour notre part nous semble suffisante dans le contexte de notre recherche, vu les communautés en ligne que nous avons observées.

1.3.2 Web de données liées

Au-delà de la dimension collaborative et participative du web, se développait en 2001 l'idée d'un « web sémantique » venue de pouvoir enrichir les traitements automatique et de faire ressortir du sens de ces contenus, afin de leur permettre d'être plus précis et performants. Le web sémantique, formulé par Tim Berners-Lee, le « père du web », était d'emblée défini comme une extension du web où « *l'information reçoit une signification bien définie, permettant ainsi une meilleure collaboration entre hommes et machines* », qui allait « *ajouter de la structure aux contenus significatifs des pages Web* » (Berners-Lee, et al., 2001). Aujourd'hui, le W3C le voit comme un « *cadre permettant aux données d'être partagées et réutilisées par delà les limites applicatives, organisationnelles ou communautaires* »²⁷ (W3C, 2015). Lors de son discours « The next web » à la conférence TED 2009, Tim Berners-Lee a lancé la campagne de son nouveau projet : construire le web de données liées. Les appellations « web sémantiques » (*semantic web*), « web de données » (web of data) ou « données liées » (linked data) désignent l'ensemble de technologies que représente la quatrième version du web « web 3.0 » pourtant chacune de ses appellations différentes présente un aspect particulier.

Le « web sémantique » était une vision du web qui met en avant le sens associé aux contenus. Les tendances prévues aux premiers stades du développement du web incluaient l'évolution des objets du web d'un document principalement lisible par l'humain à un contenu d'informations sémantiques conçu pour les machines²⁸ (Berners-Lee et al., 1994). Cette vision initiale s'est évoluée vers un « web de données » qui privilégie les données sur les documents. Comme l'indique Tim Berners-Lee, le web se basait sur les documents reliés par des hyperliens à lire et analyser. Il interconnecte désormais des données manipulables par la machine (Berners-Lee, 2009). Le web de données représente, selon Bachimont et al. (2011),

²⁷ Traduction de : “The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries”.

²⁸ Traduction de “Evolution of objects from being principally human-readable documents to contain more machine-oriented semantic information”.

une mutation du web documentaire. Il repose sur l'architecture du Web pour interconnecter, non plus des bases documentaires, mais des bases de données. Le « web de données liées » est vu quand à lui comme un instrument concret pour atteindre le web de données (Bizer et al., 2009), ce dernier étant plus réalisable que la vision initiale du web sémantique. Le web de données liées fait référence à un ensemble de techniques afin de connecter des données structurées publiées sur le web. On construit donc un web de données liées grâce aux URI (Uniform Resource Identifier) pour identifier les ressources, HTTP (Hypertext Transfer Protocol) qui est un protocole universel de recherche de ressources, et RDF (Resource Description Framework) qui fournit un modèle pour décrire les objets et leurs relations avec d'autres objets.

Pour mieux cerner l'évolution du web, il convient, selon nous, de souligner les caractéristiques des données du web. En effet, dans le web 1.0, les données consistaient en des pages HTML (*Hypertext Markup Language*) et des bases de données (Florescu, et al., 1998). Le HTML qui dérive du format *Standard Generalized Markup Language* (SGML) normalisé par l'ISO en 1986, est le format de données le moins structuré conçu pour représenter les pages web. Dans ce contexte, le principal besoin était la communication entre les données donc établir un pont entre deux ensembles de données homogènes exprimant un même concept²⁹. La tendance à la structuration va réellement s'incarner et prendre son essor en 1998 avec le format XML. Ce format permet de définir des langages de balisage pour échanger, notamment sur le Web, des données semi-structurées en arbres. Cependant, le format HTML est encore le format de diffusion le plus utilisé pour la diffusion de données sur le Web mais compte tenu de ses limites, il n'est définitivement pas adéquat pour l'échange et la réutilisation des données sur le web. La dernière technologie qui représente un degré élevé de structuration des données est le modèle RDF (Resource Description Framework). Développé par le W3C en 1999, RDF est le langage de base du Web sémantique.

²⁹ Ce besoin relève du domaine de manipulation de données du web (MDW) qui s'attache tout d'abord à l'interopérabilité entre des bases de données et par conséquent entre les systèmes qui les utilisent. Voir : Blouin Arnaud, Beaudoux Olivier, Loiseau Stéphane, « Un tour d'horizon des approches pour la manipulation des données du web. », *Document numérique* 1/2008 (Vol. 11), p. 63-83.

Le passage vers le web 2.0 a renforcé ces besoins vu le volume et la variété de données publiées. Rebillard (2007) s'accorde à dire que le web social amène à une abondance de données, phénomène aujourd'hui appelé « Big data » comme nous allons le voir ensuite. En effet, Tim O'Reilly et John Batelle ont proposé, en 2009, le terme « web² » (web au carré) pour désigner un web dont la profusion de données semble être une caractéristique première. Il s'agit d'un web plus riche de données - publiques, privées, de bien commun, statistiques, géographiques, personnelles - qui se développent, selon O'Reilly et Batelle (2009) plus de manière arithmétique mais exponentielle. Le web au carré repose sur le web 2.0, en s'appuyant sur l'ensemble de plateformes techniques permettant à tout utilisateur de produire et partager des contenus variés en ligne.

1.3.3 Open linked data

Dans le même contexte, Tim Berners-Lee a lancé, en 2009, l'appel à des « données ouvertes liées » (open linked data) en demandant aux détenteurs de données de mettre leurs données brutes sur le web dans des formats accessibles et ouverts, plutôt que de les garder enfermées dans des systèmes internes, ou de les publier sous formes de résumés ou de rapports. Selon Tim Berners-Lee (2012), les principaux problèmes de l'exploitation des données sont de l'ordre technique car il est impossible de traiter humainement des milliers de données sur le web. « Les données sont des termes, carrés, ennuyeux car ils ne peuvent pas être utilisés naturellement tel quel ». Le web connectait au départ des documents et des pages à lire et analyser. Désormais, il interconnecte des données manipulables par la machine. Il serait programmes - et pas seulement les gens - qui traitent les données. « Pour les documents, c'est la capacité de suivre les liens, mais pour les données ouvertes, il est la capacité d'interconnecter mais également de rejoindre, de résumer et de comparer, de surveiller, d'extrapoler, de déduire » (Berners-Lee, 2012). L'expression « données ouvertes liées » met l'accent, selon Bachimont et al. (2011, p.31), sur l'opportunité d'exploiter des données ouvertes par les applications et rappelle qu' « une grande valeur ajoutée réside dans l'utilisation et la réutilisation des URI pour joindre des assertions de différentes provenances mais portant sur un même sujet ».

Cette ouverture technique correspond alors aux données accessibles réutilisables dans des formats ouverts et structurés. Pour mesurer le niveau

d'ouverture des données sur le web, Tim Berners-Lee, a défini en 2010 lors de Gov2.0 Expo, un système de notation à 5 étoiles :

Tableau 2. Système de notation à 5 étoiles de Tim Berners-Lee (2010).

1	Données accessibles sur le web quel que soit le format (par exemple, fichiers PDF), mais avec une licence ouverte
2	Données accessibles dans des formats structurés, lisibles à la machine (par exemple, fichiers Excel à la place de PDF ou image numérisée)
3	Données dans des formats accessibles, structurés et non-propriétaires (par exemple fichier CSV au lieu d'Excel développé par l'éditeur Microsoft)
4	Données liées (Linked open data) en utilisant les standards ouverts du W3C (RDF et SPARQL) pour identifier les données
5	Données liées avec d'autres sources de données pour créer un contexte sémantique

Les données publiées sur le web en format PDF ou HTML ne sont pas assez structurées pour être facilement réutilisables. Ce qui nécessitait un travail fastidieux de « data scraping » pour extraire des données de formats non structurés publiés sur le web. En effet, l'extraction des données, à partir des informations, des textes, des fichiers, de la page web publiés sur internet, est devenu une pratique exercée par les informaticiens et les développeurs pour collecter les données et les transformer en formats structurés qui peuvent par la suite les intégrer dans leur système de traitement. Malgré ce paysage, des initiatives ont été lancées pour un web plus structuré. DBpedia est un exemple de la transformation de données libres non structurées, sur les pages de wikipedia, en données structurées dans une base de données libre. DBpedia est « un projet universitaire et communautaire d'exploration et d'extraction automatique de données dérivées de Wikipédia. Son principe est de proposer une version structurée et sous forme de données normalisées au format du web sémantique des contenus de chaque fiche encyclopédique » (Wikipedia, 2014). Les acteurs du web cherchent d'ailleurs à transformer le web en web sémantique quand il y aura davantage des données structurées et liées : ils revendiquent des formats qui facilitent le partage, la réutilisation, le traitement et le mixage des données pour développer un réseau plus riche de données liées.

1.4 Big Data

Big data ou mégadonnées³⁰ ou encore données massives n'est pas totalement nouveau concept. Cependant il n'a pas une définition claire (Zaslavsky et al., 2013). La première utilisation du terme « Big Data » date d'un article présenté au congrès de l'Econometric Society en 2000 (Delort, 2015). Le concept de Big Data est décrit comme un phénomène défini par l'accélération rapide du volume de données. Les défis de la gestion des données dans les entreprises, et surtout celles qui exercent des activités commerciales en ligne (e-commerce) ont été initialement étudiés par Doug Laney en 2001, sous trois dimensions : le volume (quantité des données), la vitesse (vitesse des données entrant et sortant), et la variété de données (différents types et sources de données). Le concept de Big Data s'est premièrement forgé autour de ces trois caractéristiques connues par les 3V du Big data. En plus, le Big Data sert surtout à désigner de nouvelles technologies informatiques capables de traiter ces données massives. Selon Dumbill (2013), Big Data dépasse la capacité de traitement des systèmes de bases de données classiques, les données sont trop volumineuses, qui se déplacent trop rapidement et ne correspondent pas à l'architecture basique des bases de données. Pour gagner de la valeur à partir de ces données, il faut choisir une autre façon de les traiter.

Viktor Mayer-Schönberger, professeur de gouvernance et de règlement d'internet à l'Université d'Oxford et Kenneth Cukier « data editor » de « The Economist » parlent dans leur livre *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, publié en 2013 du phénomène de « datafication » du monde, caractérisé par la capacité à traduire en données de nombreux aspects du monde qui n'ont jamais été marqués avant par le fait de tourner de nombreux aspects de notre vie en données informatisées. Par exemple, l'emplacement a été mis en données avec les systèmes de satellites GPS, les mots sont traités comme des données lorsque les ordinateurs fouillent les textes, même les amitiés sont transformées en données via Facebook et LinkedIn a converti les réseaux professionnels en

³⁰ C'est le 22 août 2014 que ce terme a été officiellement publié dans le Journal Officiel de la République française pour désigner le terme Big data. Les mégadonnées sont donc des « données structurées ou non dont le très grand volume requiert des outils d'analyses adaptés ». Voir : « Ne dites plus "big data", mais "mégadonnées" », *Le Point*, 22 août 2014.

données. Ainsi, le volume des données stockées est en pleine extension par le fait qu'en 2010 la quantité totale de données sur la terre a dépassé un zettaoctet et il est prévu que ce nombre atteindra 35 zettaoctet en 2020 (Zaslavsky et al., 2012). Non seulement les données s'accumulent mais elles proviennent de sources de plus en plus variées, et sont composées de données structurées dans des bases de données et de données non structurées issues des pages web et des réseaux sociaux.

1.4.1 La valeur des Big data

Outre les trois caractéristiques – volume, vitesse, variété – d'autres chercheurs ajoutent une quatrième caractéristique du Big data, la « valeur », qui selon nous, montre l'aspect le plus important du Big data. Les Big data comme « valeur ajoutée » que les organisations peuvent analyser automatiquement afin de générer des avantages sociaux et économiques est fortement discutée dans la littérature anglo-saxonne, et spécifiquement dans les champs de marketing, business, management et santé qui l'abordent comme une ressource que les organisations ne peuvent ignorer. En effet, Big data est intrinsèquement lié au terme Analytics qui recouvre « l'utilisation intensive de Data, de techniques d'analyses statistiques et quantitatives, de modèles explicatifs et prédictifs pour influencer la prise de décision, transformer les méthodes de management et revoir l'approche de la création de valeur en entreprise » (Data-Business.fr, 2014). Ainsi les différentes applications de Big data que nous pouvons distinguer sur les marchés visent une meilleure prise de décisions, l'élaboration des stratégies de prévention et l'amélioration des services et produits dans plusieurs domaines tels que le secteur bancaire, la santé (grippe³¹, épidémiologie³²), le commerce et la gestion des risques, la production industrielle³³, le marketing (par l'analyse des données sur les clients issues des bases de données des entreprises ou des réseaux sociaux³⁴), la gouvernance locale³⁵, le climat³⁶ et le trafic

³¹ L'exemple des recherches sur la grippe basées sur les données de « Google flu ».

³² Voir : Young, John Mark. *An Epidemiology of Big Data*. Syracuse University, 2014.

³³ Voir : McKinsey, C. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey & Company, 2011.

³⁴ Par exemple le marketing politique de Barak Obama lors de l'élection présidentielle de 2012 consistait à traiter les données collectées des réseaux sociaux pour l'analyse prédictive des préférences des électeurs. Voir :

routier³⁷. Selon un rapport de McKinsey Global Institut publiée en 2011, l'analyse de grands ensembles de données constitue une base essentielle pour la concurrence, qui sous-tend de nouvelles vagues de croissance de la productivité, l'innovation et le surplus du consommateur³⁸. Ce rapport a estimé que les commerçants qui exploitent et analysent les données massives pourraient augmenter leurs marges bénéficiaires de plus de 60 % et que le secteur de santé des États-Unis pourraient réduire ses coûts de 8%. Selon le rapport, il existe cinq manières susceptibles d'être mises en pratique pour aider les organisations à créer de la valeur, que nous exposons ici synthétiquement :

1. Favoriser l'expérimentation pour découvrir les besoins, exposer la variabilité, et améliorer les performances

Les gros volumes de données peuvent créer de la valeur par le stockage des données transactionnelles sous forme numérique afin de recueillir des informations plus précises et détaillées comme par exemple sur la performance des employés (inventaires, jours d'absences). Ainsi, l'analyse des données permet de comprendre les causes et les conséquences de la variabilité de la performance.

2. Favoriser l'innovation de nouveaux modèles d'entreprise, de produits et de services

L'étude de McKinsey Global Institut (2011) souligne que les données

<http://bosmol.com/2013/02/how-big-data-analysis-helped-president-obama-defeat-romney-in-2012-elections.html#.VaNpi5PQM6k>

³⁵ Un exemple de l'approche Big data de la ville de New York pour améliorer les services publics et réduire les coûts. Pour adopter une nouvelle stratégie de prévention des incendies, une équipe a créé une base de données de tous les 900 000 bâtiments de la ville et augmentée avec des données recueillies par 19 agences de la ville: les dossiers de privilèges fiscaux, des anomalies dans l'utilisation de l'utilitaire, les coupures de services, les paiements manqués, des visites d'ambulance, les taux de criminalité locale, les plaintes de rongeurs, et d'autres. Voir : <http://spotfire.tibco.com/blog/?p=28727>

³⁶ Voir : Edwards, P. N. *A vast machine: Computer models, climate data, and the politics of global warming*. Cambridge, MA: MIT Press, 2010.

³⁷ Comme exemple le projet MoCoPo (IFSTTAR) qui vise, par la collecter dans une base de données des données sur les trajectoires et la pollution, à maximiser l'efficacité, la sobriété énergétique et la sécurité de la circulation routière. Voir : <http://serres.ifsttar.fr/en/welcome/>

³⁸ Le surplus du consommateur est la différence entre ce qu'un consommateur est prêt à payer pour un bien et le montant effectivement payé.

massives peuvent être utilisées pour améliorer le développement de nouvelles générations de produits et services³⁹. Par exemple, les fabricants utilisent des données obtenues à partir de capteurs embarqués dans les produits pour créer des offres innovantes dans le service après-vente tel que la maintenance proactive (des mesures de prévention qui ont lieu avant qu'une panne se produit ou est même remarqué).

L'émergence des données de localisation en temps réel a créé un ensemble de nouveaux services comme la navigation et la tarification d'assurance vie et d'assurance IARD⁴⁰ basés sur l'analyse de données sur la localisation et la manière dont les gens conduisent leurs voitures.

3. Remplacer ou soutenir la prise de décision humaine par des algorithmes automatisés

L'analyse sophistiquée peut, selon le rapport, considérablement améliorer la prise de décision, minimiser les risques et dénicher des informations précieuses qui autrement resteraient cachés. Ces analyses ont des applications automatiques, par exemple les commerçants peuvent utiliser des algorithmes pour optimiser les processus de décision tel que le réglage automatique des stocks et la fixation des prix en réponse aux ventes à temps réel en ligne.

4. Segmenter la population afin de personnaliser l'action

Le Big Data Analytics permet une segmentation plus étroite des clients et par conséquent une offre des produits ou services beaucoup plus adaptée. Cette approche est bien connue dans le marketing et la gestion des risques mais les entreprises qui ont utilisé la segmentation commencent à déployer les techniques du Big data, toujours plus sophistiquées, comme le microsegmentation⁴¹ pour cibler les clients par les promotions et la publicité en temps réel.

³⁹ D'un point de vue des organisations, les différentes capacités offertes par les techniques de web 2.0 à l'internaute ont rapidement été prises en compte pour capter les opinions de leurs clients et de s'appuyer sur leurs données de navigation sur le web pour leur créer des profils et améliorer certains de ses services. Dans une approche marketing, le web 2.0 est devenu, pour ces organisations, un outil de publicité et de promotion gérer leur e-réputation et une source de données sur leurs clients.

⁴⁰ Dans le monde de l'assurance, l'abréviation IARD signifie « incendie, accidents et risques divers ».

⁴¹ La généralisation des techniques de micro-segmentation est issue d'un nouveau paradigme dans le marketing qui est le marketing 1 to 1. Les marketers s'intéressent à des micro-campagnes taillées spécifiquement pour une audience de niche. Voir *Dossier Big Data & Analytics : la Micro-Segmentation* <http://www.data-business.fr/dossier-big-data-analytics-micro-segmentation/> consulté le 18 octobre 2014.

5. Favoriser la transparence

Selon le rapport de McKinsey Global Institut (2011), rendre accessible les masses de données facilement et de manière opportune peut créer une valeur considérable dans le secteur public, par exemple, l'échange de données pertinentes entre les ministères peut réduire le temps de la recherche et le traitement des données publiques.

Face au discours emphatique sur l'hyperpuissance du Big Data et les Analytics, plusieurs constats et questionnements surgissent. D'une part, de nombreux chercheurs en sciences humaines et sociales émettent des doutes sur les capacités des nouvelles technologies dans le traitement des données massives mais surtout ils critiquent l'usage de Big data dans la science par un passage de méthodes hypothético-déductives à une logique inductive et quantitative comme étant en lien direct avec la connaissance objective. Par exemple la proclamation de la fin de la théorie scientifique de Anderson (2008), basée sur une science produite d'une façon inductive, à partir des corrélations extraites de grandes masses de données est fortement critiquée par des auteurs comme Geoffrey Bowker (2014), Antonio Casilli (2012), Danah Boyd et Kate Crawford (2012) qui dénoncent les limites épistémologiques, méthodologiques, et éthiques d'une approche Big data (traitement automatique d'énormes bases de données natives du Web) dans des études sur l'ethnographie des usages des technologies de communication. Comme le note Antonio Casilli (2012) « les Big data modifient la théorie mais ne doivent en aucun cas l'évacuer. Pour ne pas toujours céder à une recherche dirigée par les données (data-driven), il est important de préserver le rôle essentiel de la théorie dans le travail des chercheurs (**theory-driven**) ». Comme alternative aux Big data, Antonio Casilli et Paola Tubaro (2010) ont proposé la méthode ethno-computationnelle⁴². Le principal avantage de cette méthode est de collecter des ensembles de données ethnographiques denses et dirigées et par conséquent adopter une approche Small data. De plus, Bowker (2014) précise que la corrélation ne remplace pas souvent la causalité car les individus ne sont pas des catégories stables que les corrélations - tels les algorithmes qui peuvent extraire des *patterns* dans le comportement humain - peuvent toujours expliquer ou prédire leurs actions. Et juste parce que

⁴² Voir : Tubaro, P., and A. Casilli. "An Ethnographic Seduction?: How Qualitative Research and Agent-Based Models Can Benefit Each Other." *Bulletin de Méthodologie Sociologique* 106, no. 1 (2010): 59–74.

nous avons une grande quantité de données ne signifie pas, selon Bowker, que les bases de données ne sont pas théoriquement structurées de manière à permettre certaines perspectives et désactiver les autres. En effet, et comme le souligne Danah Boyd et Kate Crawford (2012): « les Big Data risquent de remettre à l'ordre du jour les divisions qui organisent les éternelles querelles sur les méthodes scientifiques ». Elles interrogent, comme nous allons le voir ensuite, de manière critique ce phénomène, ses hypothèses, ses partis-pris et rejettent l'idée que d'autres méthodes seront marginalisées au profit du volume des répétitions numériques et de vastes modélisations.

D'autre part, il apparaît que les applications du Big data dans le secteur économique posent aussi de nombreux défis. Quatre ans après la publication en 2011 du premier rapport de McKinsey Global Institut sur l'impact du Big data, David Court (2015) affirme qu'il était difficile d'atteindre le grand impact prévu du Big data sans une transformation stratégique vers une gestion centrée plus sur les données. Par ailleurs, de nouveaux défis ont surgi dans les organisations depuis l'adoption du Big data. Premièrement, les tendances issues de l'analyse de données ne donnent pas lieu à des mesures d'exploitation ou n'influencent pas les résultats d'affaires d'une manière significative. Deuxièmement, les gestionnaires de première ligne ne comprennent pas les techniques ou les modèles d'analyses et ne sont pas convaincus du pouvoir des Analytics pour les aider dans la prise de décision. Et enfin, les organisations ne possèdent pas la technologie ou les logiciels de traitement des données ce qui empêche l'exploitation efficace des Analytics sophistiqués.

De même, d'autres difficultés ont conduit plusieurs auteurs à s'interroger sur ce que peut apporter réellement les masses de données. Des préoccupations ont été soulevées au sujet de l'application du Big data en posant une série de challenges en ce qui concernent la qualité des données, les modules de traitement, l'articulation des données, la puissance des algorithmes et les problèmes éthiques que nous synthétisons ci-dessous.

a) Qualité des données

Pour Alan Mitchell, le Big data est une grande illusion s'il ne vise pas à « résoudre des problèmes, prendre des décisions, organiser et mettre en

pratique des actions sans perdre de temps et d'énergie à chercher la bonne donnée, encore moins à trier et jeter des masses d'informations non pertinentes »⁴³. Il faut certainement poser le problème des erreurs dans les données elles-mêmes surtout que les grands jeux de données récoltés sur le web sont, selon l'auteur, souvent peu fiables. Alors le principal défi et le plus important, selon Mitchell est de reconnaître la valeur des « Toutes Petites Données » (*Very Small Data*) dans la masse des données. À la question de la fiabilité des énormes masses de données s'est attaché également Antonio Casilli (2012) en affirmant que la qualité des données n'est pas toujours vérifiable surtout que les données venant de grandes bases propriétaires (produites, par exemple, par les géants du Web et les médias sociaux) ne sont pas suffisamment documentées et finissent par générer un effet de boîte. Une autre limitation des Big data à prédire le comportement humain est que « les Big Data portent sur le présent exclusivement, sans le contexte historique qui est un facteur prédictif » (Bollier, 2010). Comme l'expliquent Boyd et Crawford (2012), Facebook et Twitter sont des exemples de sources de Big Data qui n'offrent que des fonctions limitées d'archivage et de recherche. Il sera difficile, voire impossible, pour les chercheurs d'accéder à des données plus ancienne sur les réseaux sociaux comme les réactions à une élection, une finale télévisée ou un désastre naturel. Dans ce cas, les données massives ne suffisent jamais si le Big data se concentre sur des données présentes ou sur le passé immédiat.

b) Structuration des données

Comme nous l'avons déjà mentionné, Big Data est à la fois massive en volume et de nature diverse. Afin d'exploiter efficacement les données massives, il faut en premier transformer toutes les données non structurées en données structurées pour un traitement automatique et efficace des données par un logiciel. En effet, un des défis de management et de traitement des données est que la plus grande quantité des données reste largement inexploitée car elle est diffusée dans des formats non structurés. La société Merrill Lynch estime que plus de 85% de toutes les informations d'une entreprise existe en données non structurées, souvent apparaissant dans les e-mails, les notes, les rapports, les lettres, les enquêtes, les livres blancs, les présentations et les pages Web (Blumberg et Atre, 2003). Alors

⁴³ Cité par: Kaplan, Daniel. "Big Data, Grande Illusion?" *InternetActu.net*, November 4, 2012. <http://www.internetactu.net/2012/04/11/big-data-grande-illusion/>.

que les technologies et les techniques du Big Data sont bien adaptées à l'exploration des données non structurées, selon Stewart (2013), ce «grand contenu» reste grossièrement sous-utilisé et son potentiel largement inexploré.

c) Articulation des données

La capacité du Big data réside dans l'articulation des données provenant de sources différentes. Comme le précisent Boyd et Crawford (2012), « les *big data* sont remarquables, non en raison de leur taille, mais de leur capacité à être articulées à d'autres données. En raison des efforts pour exploiter et agréger les données, les *big data* sont fondamentalement liées aux réseaux. Leurs valeurs viennent des *patterns* qui peuvent être tirés du fait de connecter entre eux des jeux de données, concernant un individu, des individus liés à d'autres, des groupes de gens, ou simplement concernant la structure de l'information elle-même ». Cependant, le fait de combiner des données issues de multiples sources soulève des défis particuliers. Les erreurs et lacunes sont décuplées quand on croise de multiples jeux de données. Pour eux, l'articulation est une des questions critiques du Big data et selon eux, il est impératif que de telles combinaisons de données se fassent avec rigueur méthodologique et transparence. Notons tout de même les problèmes d'extraction des données. En effet, l'extraction est un phénomène fréquent dans le traitement du Big data mais qui dilue la valeur de données incomplètes transférées entre différents systèmes. Comme le souligne Paul Bertrand (2012), la mise en relation de données aboutit à la création d'une donnée partielle et souvent non définie détachée de son contexte de production.

d) Puissance algorithmique

La puissance algorithmique est critiquée par plusieurs auteurs qui mettent en exergue les limites techniques et analytique du Big data. Boyd et Crawford (2012) parlent d'une mythologie qui entoure l'émergence du phénomène de Big data fondé sur la croyance répandue que de grands ensembles de données offrent une forme supérieure de l'intelligence et des connaissances qui peuvent générer des idées qui étaient auparavant impossibles et affirment que les revendications d'objectivité et d'exactitude sont trompeuses. Dans le même sens, Rouvroy et Berns (2013), expriment que dans le cas de machine learning, « nous nous trouvons à nouveau face à

l'idée d'un savoir dont l'objectivité pourrait paraître absolue, puisqu'il serait éloigné de toute intervention subjective (de toute formulation d'hypothèse, de tout tri entre ce qui est pertinent et ce qui ne serait que du « bruit », etc.) ». En outre, pour Bollier (2010), la valeur d'objectivité des données est critiquée par le fait que la collecte et le nettoyage des données demeurent une affaire subjective. Le fait de modifier un jeu de données dans le but de le nettoyer (le choix des attributs et des variables à supprimer ou à garder) est un processus subjectif qui élimine l'objectivité du jeu de donnée. Les problèmes d'avoir une interprétation objective des données sont plus graves lorsque l'information provient de différentes sources.

e) Problèmes éthiques

Nous abordons la question d'éthique du Big data sous deux angles. Le premier consiste à rendre compte des sources de données pendant la collecte ce qui soulève le problème de la propriété intellectuelle. Selon Boyd et Marwick, il n'est pas éthique, de la part de chercheurs, de légitimer leurs actions par le simple fait que les données sont accessibles. Les données déjà publiées sur les réseaux sociaux sont collectées, par les chercheurs ou les marketeurs, pour des objectifs précis alors que des internautes n'ont pas conscience de la multiplicité d'agents et d'algorithmes qui collectent et stockent leurs données pour des usages ultérieurs. Le second angle souligne la variété d'utilisateurs potentiels avec accès aux données recueillies pour une grande variété de fins représentant les préoccupations de sécurité et de vie privée. Zimmer (2008) donne l'exemple d'un projet de recherche basé à Harvard qui a commencé par rassembler les profils de 1700 étudiants usagers de Facebook afin d'étudier comment leurs centres d'intérêts et leurs amitiés évoluaient avec le temps. Ces données prétendument anonymes ont été rendues accessibles à tous, permettant à d'autres chercheurs de rapidement découvrir qu'il était possible de désanonymiser certaines parties de ce jeu de données, compromettant ainsi la vie privée des étudiants, dont aucun ne savait que ces données avaient été collectées. Ainsi, le croisement de données peut révéler de très sérieuses violations de la vie privée.

Nous pouvons ajouter à ces défis la question des compétences associées au Big data. Dans le monde des entreprises, le rapport de McKinsey Global Institute (2013) souligne la pénurie imminente de « data scientist ». Nous présentons le profile du data scientist et ses capacités dans la partie qui suit sur l'appropriation des Big data.

1.4.2 L'appropriation des Big data

Selon l'étude de McKinsey (2013), en 2018, les États-Unis connaîtront une pénurie de 190 000 spécialistes qualifiés de la donnée, et 1,5 million de gestionnaires et d'analystes capables de tirer des pistes d'actions à partir du grand déluge de données. Le besoin de spécialistes de la donnée doit être ainsi confronté par la formation spécifique, l'accroissement du nombre de diplômés dans les domaines de la science, technologie, ingénierie et mathématiques qui permet de créer une main d'œuvre plus qualifiée et plus compétitive. Au niveau K-12, l'amélioration de l'enseignement en classe et l'introduction d'outils d'apprentissage numériques peuvent améliorer le rendement des élèves.

Selon Conway (2010), un « data scientist » compétent est la personne qui possède toutes les compétences techniques d'un Hacker, analytique d'un statisticien et l'expertise dans un contexte métier particulier. (Voir figure 4).

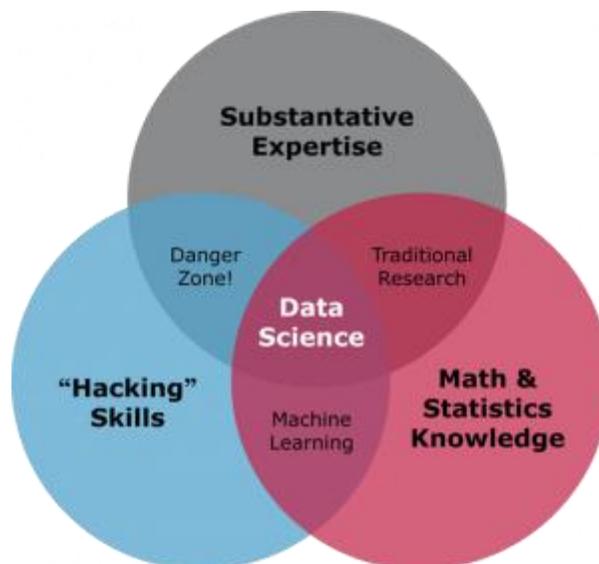


Figure 3. The Data Science Venn Diagram (Conway, 2010)

Data Science Venn Diagram (Conway, 2010) fusionne trois compétences de nature interdisciplinaire:

- Les compétences d'un pirate informatique : La manipulation des fichiers numérique (y compris la collecte, le nettoyage), la compréhension des opérations vectorielles et algorithmiques.

- Les compétences d'un statisticien : L'extraction des informations clés à partir des données en appliquant les méthodes statistiques appropriées.
- Les connaissances d'un expert dans un domaine du savoir : la création et la découverte de nouveau savoir à partir des données en posant des questions réflexives et émettant des hypothèses testées par des méthodes statistiques.

Pour Conway (2010), le dernier élément est crucial pour être un data scientist. L'absence de l'expertise personnelle, dans un milieu qui combine uniquement des compétences informatiques et statistiques n'apporte que des « résultats techniques » de ce qu'on appelle le *machine learning*. En revanche, quand l'expertise est réunie aux compétences statistiques, nous nous trouvons dans la catégorie des « chercheurs traditionnels » experts dans leurs domaines mais qui connaissent insuffisamment sur la technologie. Enfin, la « zone de danger » représente pour lui la catégorie des gens qui possèdent des compétences techniques d'extraction et de structuration des données dans un domaine qu'ils maîtrisent mais qu'ils n'acquièrent pas les connaissances statistiques nécessaires. Le danger réside d'une incompréhension des méthodes statistiques qui entraînera par la suite de faux résultats.

Dans le contexte des entreprises, Tam Harbert (2012), dans son article sur les compétences et les métiers associés aux données, fait référence à cinq profils de spécialistes :

- Les scientifiques de données ou Data scientists: ils jouent un rôle principal dans le processus de traitement des données brutes à cause de leurs qualités d'analyses profondes. Ils déterminent quel type d'analyse sur les données apportera les meilleurs résultats. Les data scientists, tel que cité par Harbert, sont issus notamment des domaines des mathématiques et statistiques, ainsi que l'intelligence artificielle et le traitement du langage naturel.
- Les architectes de données: ils sont des programmeurs ayant des qualifications élevées en mesure de construire un modèle qui articule des données en désordre de plusieurs sources. Ils ont besoin de la créativité et de persévérance pour être en mesure d'exploiter les données dans de nouvelles façons et de créer de nouvelles connaissances.

- Les visualiseurs de données: ils transforment et présentent les analyses de données dans un format très visuel et compréhensible par les décideurs de la société. Ils exploitent les données et les mettent en contexte, en langage profane, pour donner un sens aux données et saisir leur impact sur la société.
- les Agents de changement: les personnes qui conduisent des changements dans les opérations et les processus internes basée sur l'analyse de données. Ils sont des évangélistes et mobilisateurs des ressources nécessaires pour innover.
- Ingénieur de données / opérateurs: ils sont les concepteurs, constructeurs et gestionnaires de l'infrastructure des données. Ils développent l'architecture pour satisfaire au mieux les besoin de l'entreprise.

Cette présentation des compétences et profils des spécialistes des données nous servira dans nos réflexions sur le profil des utilisateurs des données publiques ouvertes, mais aussi notamment sur leurs capacités d'appropriation des données.

Cette description synthétique des enjeux du Big data dans la littérature montre un environnement d'incertitudes face à la valeur réelle des données gigantesques omniprésentes sur le web ou dans des bases de données. Plusieurs applications démontrent l'intérêt du Big data mais signalent de sérieux défis méthodologiques et techniques. Le phénomène de Big data soulèvent des questions sur la manière dont les chercheurs (incluant les chercheurs en humanités et en sciences sociales), les marketeurs, les organisations, s'emparent des données massives pour générer des bénéfices. Cette question n'aura pas une réponse claire tant que les données et les algorithmes qui les traitent font parti du secret professionnel, sauf dans le cas des recherches ou les méthodes de collecte et de traitement sont connues.

Conclusion du chapitre 1

Dans ce chapitre, nous avons présenté les principales notions conceptuelles en relation avec la notion de « données », nécessaires à l'appréhension des observations effectuées durant notre recherche. Nous avons montré, en nous appuyant sur la littérature adéquate, l'émergence de la notion de données comme objet de production, de diffusion et de traitement.

Etant des symboles enregistrés par les outils informatiques, les données numériques manipulables par les machines sont devenues massives vu la capacité à traduire en données de nombreux aspects du monde. Ce nouveau monde de données est caractérisé par une évolution des modes de production, de mise en réseau et d'exploitation de données.

Le web 2.0 et ses technologies ont profondément transformé les pratiques informationnelles et médiatiques en introduisant de nouveaux modes de production et d'échange de données. En effet, dans une société numérique, tout individu constitue une source de données en laissant des traces de ses interactions avec et sur un territoire. Les outils qui enregistrent et qui captent ces traces permettent une production et une matérialisation des données. Le choix de mettre à disposition et de partager leurs données constitue une étape supplémentaire rendu possible par les applications du web 2.0. La création de plateformes de données ouvertes collaborative dans un « esprit wiki » transforme le citoyen en producteur et diffuseurs collaboratif de données. L'exploitation des données (ou les données massives avec le Big data) afin de générer de la valeur nécessite de repenser la structure même des données, voire même la manière dont les données sont publiées sur le web. Le mouvement open data, n'est donc qu'un facilitateur pour le développement d'un web plus riche en données structurées. L'open data, en tant que donnée structurée, ouverte techniquement est une matière première réutilisable facilement dans le processus de traitement du Big data. Comme intermédiaire entre les deux pôles, le web et le Big data, l'open data est un ensemble de caractéristiques d'une donnée publiée sur le web qui facilite son exploitation et traitement. Des langages, techniques et outils sont évidemment créés dans le but de faciliter la réutilisation des données par un traitement automatique et efficace mais des entraves, non techniques, subsistent encore empêchant l'exploitation des données. L'ensemble des contraintes et les mouvements favorisant l'exploitation des données sont traités dans le chapitre suivant.

Chapitre 2. Le principe d'ouverture

Suite à la présentation de la notion de données et ses enjeux de production, de diffusion et de traitement, nous souhaitons aborder dans ce chapitre la notion « open » ou « ouvert ». Plus précisément, nous allons ici poser les caractéristiques d'une donnée, dite ouverte, et plus globalement présenter le principe d'ouverture qui régit dans plusieurs domaines. Comme pour le chapitre précédent, nous s'appuyons sur la littérature existante dans multiples champs disciplinaires ayant abordé le principe d'ouverture.

L'adjectif «open» est devenu un préfixe puissant qui véhicule plusieurs valeurs et principes. Il présente et précise un ensemble de caractéristiques du mot qui le suit : open source, open access, open knowledge, open hardware, open innovation, open government ... Ces termes sous-entendent des concepts philosophiques et idéologiques, des technologies et des terrains d'application spécifiques. Nous les avons classés en trois grandes catégories dont chacune représente un environnement différent (figure 3):

1. Culture libre
2. Innovation ouverte
3. Gouvernement ouvert

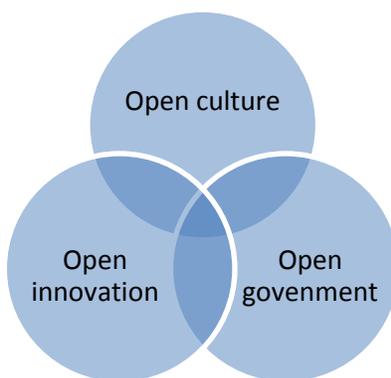


Figure 4. Les trois champs du principe d'ouverture

Nous procédons tout d'abord par une présentation de la notion de « culture libre » sous laquelle nous englobons les notions : Open source, Open access,

Open knowledge. Cette association de termes qui semble propre au web nous offrira l'occasion de questionner la cyberculture, ainsi que l'aspect collaboratif propre au web dont les notions de « réutilisation » et de « remixage » sont l'un des avatars les plus prégnants. Ensuite, nous tendrons vers la notion d'« Open innovation » issue du monde des entrepreneurs. Nous aborderons alors les stratégies d'ouverture au sein des organisations vues comme un nouveau modèle de management en faveur de l'innovation. Enfin, nous reviendrons sur la notion d'« Open government » en mettant en avant les recherches qui interrogent les bases et les idéologies d'un gouvernement ouvert. Ce questionnement nous ouvre la voie, dans la partie suivante, vers l'étude des données ouvertes dans le secteur public, l'environnement propre à notre travail de recherche.

2.1 Culture libre et cyberculture

« Le mouvement dit « *culture libre* » fonctionne comme un mode de pensée visant à défendre et garantir certains droits fondamentaux aux citoyens face à une – prétendue – mainmise de la propriété intellectuelle sur les connaissances et leur transmission » (Oliveri, 2011). Bien que ce phénomène existe avant l'internet⁴⁴, le développement de l'informatique réseautique apparaît clairement comme le support idéal de l'essor de la culture libre. Selon Musso (1997), de la culture de masse centralisatrice, massive et fermée, nous voyons, avec les diverses formes de sociabilité dans les réseaux télématiques, l'expansion d'une culture basée sur la personnalisation, la collaboration et l'ouverture. Ainsi que, et comme l'indique Soufron (2009), les modèles collaboratifs et le Web 2.0, que nous avons déjà évoqués dans le premier chapitre, ne sont que des « formes applicatives rendues possibles par l'irruption du libre et de l'ouvert ». Ces modèles reposent alors sur une dynamique socioculturelle qui se retrouve sur le web par la mise en place des nouvelles formes de distribution, de coopération et de, l'appropriation des biens numériques, tels que les logiciels et les contenus numériques. Ces affirmations rencontrent le discours du courant technophile⁴⁵ particulièrement actif visant à reconnaître que le web

⁴⁴ Voir l'exemple du partage d'information techniques de James J. Flink (1977). *The Car Culture*. MIT Press. [ISBN 0-262-56015-1](#).

⁴⁵ Voir : Lévy, Pierre. *Cyberdémocratie : essai de philosophie politique*. Paris : Éditions Odile Jacob, 2002.

est le lieu privilégié d'émancipation collective et de libre circulation des idées.

D'après Lemos, « les nouvelles technologies d'information et de communication modifient les processus de production, de création et de circulation de biens et de services dans ce début de siècle, en apportant une nouvelle configuration culturelle ». Ainsi, c'est cette nouvelle relation entre les technologies et la sociabilité, configurant la culture contemporaine, que Lemos (2006) définit par la cyberculture. Selon Lemos et Pedro (2005), « la culture des réseaux a été (et en est encore) fruit d'influences mutuelles, de travail coopératif, de création et de circulation d'informations ».

Comme l'a souligné Oliveri (2011), la culture libre défend la transmission des connaissances afin de faire face à la propriété intellectuelle. Mais, pour Soufron (2009), le libre ne se construit pas comme une opposition au droit de propriété, il est plutôt « une solution juridique simple à la question de savoir comment travailler collaborativement tout en respectant le droit d'auteur ». D'ailleurs, afin d'asseoir la culture libre, cinq points ont été établis comme cadre suffisamment souple permettant des activités de partage et de coproduction :

1. Conservation : le droit de produire, posséder, et contrôler des copies d'un contenu (par exemple, télécharger, dupliquer, stocker et gérer).
2. Réutilisation : le droit d'utiliser le contenu dans un large éventail de moyens.
3. Révision : le droit d'adapter, d'ajuster, modifier ou altérer le contenu lui-même.
4. Remix : le droit de combiner le contenu original ou révisé avec d'autres contenus pour créer quelque chose de nouveau (par exemple, incorporer le contenu dans un mashup).
5. Redistribution : le droit de partager des copies du contenu original, ses révisions, ou ses remixes.

Ajoutons à ces points, le droit de commercialisation qui met en exergue la valeur économique des œuvres dérivées qui pourraient être bâties de l'œuvre originale.

Selon Soufron (2009), « le succès du libre repose sur la création d'un écosystème dans lequel chaque membre est poussé à réutiliser au fur et à mesure la création des autres ». Tout comme la réutilisation, le remix, selon

Lemos (2006), provient de plusieurs pratiques sociales produites par les « trois lois de la cyberculture ». La culture de remix, est alors un « ensemble de pratiques sociales et communicationnelles de combinaisons, de collages et d'appropriation des morceaux d'informations à partir des technologies numériques ». Les pratiques de réutilisation et de remixage font alors valoir leur pouvoir créatif quand il s'agit des données numériques.

Historiquement, les premiers modèles de licences libres touchent au domaine du logiciel, que nous aborderons ci-après. Il en existe un très grand nombre dont chacune autorise une activité ou interdit une autre. Les contreparties les plus courantes concernent surtout le respect de la paternité du ou des auteurs⁴⁶. Certaines sont des licences maison et ne concernent qu'un seul projet ; d'autres ont vocation à servir de modèle et rencontrent une popularité particulière. Des associations comme la Free Software Foundation, l'Open Source Initiative, Creative Commons, et l'Open Knowledge Foundation assurent le développement d'une série de licences qui couvrent plusieurs type d'œuvres. Les modèles de contrats de licence les plus connus sont la GNU Public License (GNU-GPL) et les contrats Creative Commons qui ont été traduits en plusieurs langues et qui régissent de nombreux contenus multimédias y compris les articles de Wikipedia.

La cyberculture met alors en synergie des processus d'échange et de modification créative d'œuvres numériques. Ces processus sont à l'origine du développement de nombreux termes, open source, open access et open knowledge, que nous synthétisons ici.

2.1.1 Open source

Les logiciels open source⁴⁷ sont des programmes d'ordinateur qui ont comme base une licence qui permet la modification de leurs codes source afin de

⁴⁶ Nous reviendrons ensuite dans ce chapitre sur les divers licences disponibles, surtout celles développées pour accompagner le mouvement open data.

⁴⁷ Les deux notions « Open source » et « Free software/logiciel libre » représente deux courants de pensée par rapport à la notion de liberté des logiciels. Comme l'explique Oliveri (2011): « un schisme survenu au 1998 entre les défenseurs de la culture libre de l'informatique. Effectivement, à partir de cette période, deux courants de pensées spécifiques vont venir reconfigurer le rapport jusqu'alors entretenu par les informaticiens avec la notion de liberté. La création de l'Open Source Initiative marque l'avènement de l'*open source*. Ce mouvement s'appuiera sur un discours visant à défendre une meilleure adéquation de leurs principes avec les enjeux économiques et commerciaux

développer un logiciel dérivé ; par opposition au logiciel propriétaire qui fournit seulement une licence d'utilisation, sans que l'utilisateur puisse connaître ou modifier le code source. L'histoire du logiciel libre (Free software⁴⁸) remonte aux années 80 quand Richard Stallman, fondateur de Free Software Foundation a annoncé son projet du système d'exploitation libre GNU. Ainsi, le logiciel libre, qui est à la base une méthodologie de développement de logiciel, est devenu un mouvement social concrétisant le concept de culture libre prétendu comme imaginaire. Comme nous l'avons déjà souligné, les premiers modèles de licences libres touchent effectivement au domaine du logiciel.

Au niveau technologique, l'accès au code source du programme permet aux développeurs d'exécuter le programme, d'étudier son fonctionnement, de le modifier et de le redistribuer. Philosophiquement, le mouvement du logiciel libre défend l'idée de la participation, de la collaboration et de l'engagement d'une communauté à partager et réutiliser le code source du programme pour apporter des améliorations, pour en faire profiter toute la communauté (Free Software Foundation, 2013). D'après Oliveri (2011), « les développeurs de logiciels libres sont apolitiques, indépendants et motivés par un esprit d'entraide, de partage et d'échange, qui n'est pas sans évoquer les principes de la mutualisation ».

La collaboration explique l'émergence d'un modèle différent de production en contraste avec les structures rigides et centralisées monopolisées par les logiciels commerciaux et les grands éditeurs scientifiques. Ce modèle, qui n'est plus nouveau et qui date des années 2000 a prouvé sa viabilité. Comme le précise Oliveri (2011), « peut-être victime de son succès, la philosophie du don et de la collaboration, tend parfois à verser dans une utopie communicationnelle. Cependant, cette alternative du partage des savoirs trouve un écho grandissant et indéniable auprès du grand public ». L'auteur donne l'exemple du navigateur libre Firefox lancé par la fondation Mozilla en 2002. Aujourd'hui, l'organisation à but non lucratif revendique pour son navigateur pas moins de 500 millions d'utilisateurs dans le

du capitalisme. L'idée étant de dire que l'*open source* peut être une alternative viable, en termes commerciaux et techniques, au logiciel propriétaire ».

⁴⁸ Plus clairement, le terme anglais *free* ne signifie pas gratuit et un logiciel libre n'est pas forcément gratuit.

monde⁴⁹. Firefox, est classé 2ème dans le monde. Il possède entre 23% et 21% de l'usage des internautes des moteurs de recherche en 2015⁵⁰. Ce chiffre, atteste-il sans doute la viabilité de ce modèle économique et l'engagement d'une communauté dans le partage et la coproduction. .

2.1.2 Open access

Au-delà du monde des logiciels, des explorations parallèles des effets du mouvement du libre ont eu lieu dans d'autres domaines. Le principe de partage est appliqué sur plusieurs types de contenus numériques sur le web. Début des années 90, le mouvement « **open access** » ou « accès libre » visait à mettre à disposition gratuitement en ligne la littérature scientifique. Avec la naissance et l'essor du World Wide Web, les chercheurs scientifiques ont voulu faire face au monopole exercé par les grands éditeurs scientifiques et les coûts excessifs des revues scientifiques. L'appel lancé lors du congrès de 1er et 2 décembre 2001 à Budapest et diffusé à partir du 14 février 2002 et connu sous le nom de *Initiative de Budapest pour l'accès ouvert* entend par « accès libre » à une littérature :

« Sa mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces articles, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celle indissociable de l'accès et l'utilisation d'Internet. La seule contrainte sur la reproduction et la distribution, et le seul rôle du copyright dans ce domaine devrait être de garantir aux auteurs un contrôle sur l'intégrité de leurs travaux et le droit d'être correctement reconnus et cités » (Initiative de Budapest pour l'accès ouvert, 2002).

L'encyclopédie libre Wikipédia, depuis son lancement en 2001, illustre parfaitement le succès du modèle du libre acte basé sur le principe de fonctionnement et de production en collaboration. En Mai 2014, selon la compagnie Alexa Internet du groupe Amazon, fournisseur de statistiques sur

⁴⁹ Voir : "Mozilla at a glance". Mozilla blogs. Accessed Juin 20, 2015. <https://blog.mozilla.org/press/ataglance/>

⁵⁰ Voir : "Browser statistics and trends". W3schools. Accessed May 20, 2015. http://www.w3schools.com/browsers/browsers_stats.asp.

le trafic du Web mondial, le site web Wikipedia est classé # 6 par le nombre de visiteurs par mois, juste après Baidu (le moteur de recherche chinois), Yahoo, Youtube, Facebook, et Google qui est classé premier⁵¹.

Des pratiques de production collaborative d'objets libres sont également observées. **Open Hardware** ou Matériel libre, selon les principes d'open source, permet de créer des machines, dispositifs ou toutes choses physiques de façon que quiconque puisse les fabriquer, modifier, distribuer et utiliser. Ces objets libres sont normalement fabriqués dans des *fab lab* ou laboratoire de fabrication. La première Fabrique d'Objets Libres créée à Lyon en 2012, se décrit comme : « Espace de création et de fabrication numérique ouvert à tous, la Fabrique d'Objets Libres permet à chacun de découvrir, d'inventer et de fabriquer tout type d'objet »⁵². En pharmaceutique, The Tropical Disease Initiative, le projet libre de découvertes de médicaments, « vise à fournir un «noyau» pour permettre aux scientifiques des laboratoires, des universités, des instituts et des sociétés de travailler ensemble pour une cause commune: trouver de nouveaux médicaments contre les maladies tropicales telles que le paludisme ou la tuberculose »⁵³.

2.1.3 Open Knowledge

Cette extension des principes de l'Open source a incité l'association Open knowledge foundation (OKF), créée en 2004 en Royaume-Uni défendant le mouvement libre, à donner une description en essayant de définir le mot « Open » ou « libre⁵⁴ ». L'OKF a ajouté une série de conditions, résumés dans la version 1.1 de l'« Open Definition⁵⁵ », achevée en Novembre 2009, pour qu'une œuvre, y compris les données, soit ouverte. La définition plus large de l'« Open Knowledge » ou « le savoir libre » inclut :

- des contenus tels que la musique, les films ou les livres

⁵¹ Voir : “Top 500 Global Sites”, *Alexa Internet*. Accessed May 28, 2014. <http://www.alexa.com/topsites>.

⁵² Voir : “Fabrique d'Objets Libres,” 2013. <http://www.fablab-lyon.fr/>.

⁵³ Voir : “The Tropical Disease Initiative.” Accessed May 28, 2014. <http://sgt.cnag.cat/TDI/>.

⁵⁴ Dans la version française du site de l'Open Knowledge Foundation, le mot « Open » est traduit en « Libre ».

⁵⁵ Voir Open Knowledge Foundation. “Définition Du Savoir Libre.” *Open Definition*. Accessed February 22, 2013. <http://opendefinition.org/okd/francais/>.

- des données, qu'elles soient scientifiques, historiques, géographiques ou autres
- des données publiques émanant du gouvernement ou d'autres administrations publiques

La version abrégée de l'Open Définition déclare qu' « une donnée ou un contenu est ouvert si quelqu'un est libre de l'utiliser, le réutiliser et le redistribuer – dans la seule limite d'attribuer et / ou partager à l'identique »⁵⁶.

La définition détaillée donne un cadre sur la nature de la donnée elle-même mais surtout elle détaille les conditions d'usages sous forme de licence. Les données doivent être publiées sur internet dans des formats ouverts non propriétaires pour abolir les obstacles techniques de réutilisation. Concernant les conditions d'usages, toute personne peut utiliser (vendre, réaliser des œuvres dérivés) les données sous le principe de « non discrimination » même pour des usages commerciaux. Pourtant le diffuseur peut attribuer un prix « ne dépassant pas un coût raisonnable de reproduction » pour l'usage des données, même si la définition accorde une préférence au principe de gratuité. En plus, il peut imposer l'attribution par la mention des auteurs de la donnée ou la redistribution des données aux mêmes conditions que les données originales.

En accordant une attention particulière aux données, dans le cadre de son projet « open data commons », l'OKF a publié en 2008 les premières licences dédiées aux données : Open Data Commons Public Domain Dedication and License (PDDL), Open Data Commons Attribution License (ODC-BY), Open Data Commons Open Database License (ODbL).

La licence ODC-BY impose d'indiquer le nom de l'auteur de la base de données originale. On retrouve ces principes dans la licence Creative Commons By. La licence (ODbL) donne la possibilité aux utilisateurs de partager, créer, et modifier et produire une création à partir d'une base de données sous réserve de mentionner la source de la base de données et la redistribuer sous les mêmes conditions imposées par la licence originale. La

⁵⁶ Traduction de l'Anglais de : “A piece of data or content is open if anyone is free to use, reuse, and redistribute it - subject only, at most, to the requirement to attribute and/or share-alike”.
<http://opendefinition.org/#sthash.LltuANuK.dpuf>

première de son genre, la PDDL donne la possibilité d'utiliser, de copier, de modifier, de redistribuer une base de données sans aucune restriction. C'est une licence libre de tout droit, de type domaine public. Les données peuvent donc être exploitées de façon totalement libre et l'auteur abandonne son droit moral. Une licence similaire, licence CC0 – zero, a été ensuite développée par Creative Commons et lancée en 2009. La licence CC0 autorise toute personne à réutiliser librement les données, les modifier, quel que soit le but et sans aucune restriction.

A partir de ce rapide panorama de la notion d'Open, nous pouvons dire que la culture libre prend de plus en plus d'ampleur, elle se développe et évolue dans plusieurs contextes comme un modèle de partage et de création collaborative sur le web. Le monde des logiciels libre et l'open access à la littérature scientifique démontrent bien le succès du libre pour être justement appliqué à tout savoir. Cette évolution représente, selon nous, une bonne adéquation avec les besoins ou les envies des communautés virtuelles. Le mouvement du libre a atteint son apogée par la publication des licences de domaine public (PDDL ou CC0) qui n'accordent aucune restriction à l'usage des données. Ainsi, les données ont réussi à outrepasser les limites d'accès et de réutilisation inhérentes à d'autres types de contenus sur le web. Ainsi, cette nouvelle génération de licence sur les données du domaine public, nous invite à redéfinir la définition d'Open Knowledge de L'OKF en distinguant les données ouvertes (ou les bases de données ouvertes), mais surtout de s'interroger sur le nouveau contexte et le terrain d'adoption de ces licences⁵⁷.

Le mouvement du libre donne alors un cadre pratique des principes d'une donnée ouverte engendrés par une philosophie fondée sur la participation, la collaboration, la coproduction et l'engagement d'une communauté à partager le savoir pour en faire profiter toute la communauté.

2.2 Innovation ouverte

Dans nos questionnements sur le principe d'ouverture, nous nous intéressons au concept d'open innovation qui s'attache à un modèle ouvert

⁵⁷ Nous nous interrogeons par la suite sur l'effet du mouvement Open data dans le secteur public sur la publication des licences du domaine public.

de gestion au sein des organisations afin d'améliorer leur performance et créer de la valeur. Aborder ce concept répond à plusieurs motivations : tout d'abord, et comme nous allons le voir, l'open innovation propose des modèles de gestion des flux de savoir entre les acteurs. De plus, nos observations sur le terrain, nous ont amené à nous appuyer sur plusieurs approches issues de l'open innovation, que nous insérons par la suite dans le contexte de notre recherche sur l'open data dans le secteur public.

L'innovation est l'application de nouvelles solutions qui répondent aux nouvelles exigences, aux besoins inarticulés, ou aux besoins existants du marché. Ceci est accompli grâce à des produits plus efficaces, des processus, des services, des technologies ou des idées qui sont à la disposition des marchés, des gouvernements et de la société. La logique de «l'innovation ouverte», issue des écoles de business et de management, et popularisée par Henri W. Chesbrough (2003), suppose que, dans un monde de connaissance, l'innovation se fait maintenant dans les réseaux d'entreprises, plutôt que dans une seule entreprise. Une entreprise ne peut pas compter entièrement sur ses propres recherches, une ouverture est indispensable par le fait accès à des briques technologiques développées à l'extérieur et qui entrent dans la conception de ses propres produits et services (outside – in). Inversement, les inventions internes sont prises en dehors de l'entreprise par des licences en faveur d'autres compagnies pour créer un nouveau marché (Inside-out). Pour Vanhaverbeke⁵⁸, l'ouverture est d'avoir des partenaires externes et la gestion du réseau externe de partenaires est le processus clé pour l'ouverture. Le paradigme de l'innovation ouverte, est également un changement dans l'utilisation, la gestion et l'emploi de la propriété intellectuelle (West et Gallagher, 2006). Selon Chesbrough, Vanhaverbeke et West (2006), l'innovation ouverte est l'usage de flux entrant et sortant de savoir dans le but d'accélérer l'innovation interne et d'élargir les marchés pour un usage externe de l'innovation, respectivement.

Si l'open innovation suppose et encourage la participation de tiers, le terme ne précise pas comment un tiers est-il impliqué, quel rôle joue-t-il et comment est-il invité à participer. Selon Phillips (2010), ce manque de clarté à propos de l'innovation ouverte et comment elle est déployée et les

⁵⁸ Propos tenus par Wim Vanhaverbeke lors de la journée d'étude "L'open Innovation" à laquelle nous avons participé à Lyon en 2013.

avantages qu'elle peut créer provoquent souvent une confusion pour les entreprises. Philips propose alors une typologie qui permettra de distinguer différents types d'innovation ouverte. Les deux attributs définissant cette typologie sont :

- Participative ou sur invitation : l'entreprise invite des personnes précises à soumettre leurs idées ou toute personnes intéressées ?
- Suggestive ou dirigée: les idées devraient être influencés ou dirigés par des sujets ou des besoins spécifiés par l'entreprise, ou devraient être permis aux participants de soumettre des idées sans limites revendiqués ou conditions ?

Cette typologie est particulièrement utile lorsque l'on considère comment déployer l'innovation ouverte dans une entreprise. En considérant qui pourrait participer, les entreprises ont un vaste éventail d'options. : d'un côté, un petit groupe d'individus choisis pour leurs connaissances, de l'autre, le « crowdsourcing », c'est à dire toute personne ayant un intérêt dans le sujet peut soumettre ses idées, selon une logique participative que sur invitation.

Les modèles observés sont très divers, depuis les collaborations simples jusqu'à la mise en place d'écosystèmes complexes. Dans la logique de l'innovation ouverte, telle qu'elle a été définie par Chesbrough, nous nous intéressons aux modèles d'organisation et de gestion de flux de savoir entre les acteurs sous les deux aspects d'internalisation (outside-in) et d'externalisation (inside-out). Dans le cas d'internalisation, nous présentons le modèle de crowdsourcing. Ensuite, nous questionnons l'open data comme un modèle d'externalisation des données internes exercé par les entreprises.

2.2.1 Crowdsourcing

Le fait de s'appuyer sur la foule n'est pas à proprement parler une approche nouvelle dans l'open innovation. Selon Renault (2014), le crowdsourcing « permet à des organisations de toute nature d'externaliser vers la foule des activités réalisées traditionnellement en interne ou par un prestataire identifié ». Pour Burger-Helmchen et Pénin (2011), il représente une forme de collaboration entre une entreprise et des individus extérieurs ou bien encore une forme de production alternative aux autres formes de production en interne.

Tel qu'il est entendu sur le web, le crowdsourcing repose sur la mise en place par une organisation d'une plate-forme dédiée, et sur laquelle les internautes seront incités à y agir (Chanal et Caron-Fasan, 2010). Le crowdsourcing constitue, d'après Amy Shuen (2008) un exemple en matière de prospective des nouveaux usages des technologies de l'information. Dans son livre sur les modèles Web 2.0 dans les entreprises, elle estime que le Web 2.0 adopte une vision fondamentalement différente de la façon dont les entreprises, les clients et les partenaires interagissent, et, ce faisant, il ouvre un éventail de nouveaux modèles d'affaires.

Burger-Helmchen et Pénin (2011) distinguent trois types de crowdsourcing :

1. Le « crowdsourcing d'activités routinières » (réalisation de tâches peu complexes mais demandant du temps à l'organisation).

Un exemple caractéristique du crowdsourcing d'activités routinières est la surveillance de caméras vidéo par la foule sur le site « Internet eyes »⁵⁹. C'est un service dont l'objectif est de permettre aux entreprises de surveillance, qui ont installé des caméras chez des clients, de faire surveiller ces caméras par la foule.

2. le « crowdsourcing d'activités inventives » (développer des solutions ou des connaissances sur un sujet). Des clients fidèles, ou des personnes partageant un intérêt ou une passion dans différents domaines propose des solutions à un problème posté en ligne sur la plateforme de l'entreprise. Dans le cadre d'un concours, les idées gagnantes reçoivent une certaine forme de prime, et l'entreprise rentabilise des connaissances développées en externe pour son propre profit. Le cas de l'entreprise Procter and Gamble, souvent cité⁶⁰, décrit comment l'entreprise publie certains défis techniques sur le web et obtient ainsi des solutions rapides et pertinentes. Nous pouvons citer également le constructeur automobile Audi qui a lancé en 2012 le Audi Production Award⁶¹. C'est un concours qui demande aux participants de réfléchir à la production des voitures du futur. Le vainqueur a reçu un trophée ainsi que la somme de 12,500 euros.

⁵⁹ www.interneteyes.co.uk

⁶⁰ Voir : Sakkab, N. "Connect and Develop Complement Research and Develop at P&G." *ResearchTechnology Management*, no. March-April (2002): 38-46.

⁶¹ Voir : Ennomotive. "Audi's Vision of Future Manufacturing Using Crowdsourcing." *Ennomotive*. Accessed October 3, 2015. <http://www.ennomotive.com/audis-vision-of-future-manufacturing-using-crowdsourcing/>.

3. Le « crowdsourcing de contenus » (production d'information pour l'entreprise). Ce type de crowdsourcing est différent des autres dans le sens où la foule apporte ici non pas simplement de son temps pour accomplir une tâche aisée, ou d'apporter des solutions mais alimente un stock de données et d'informations pour créer une sorte de base de données. L'intérêt de la foule réside donc ici non seulement dans sa taille, mais également dans sa diversité afin d'apporter un contenu riche, naturellement sa construction par une entité simple est très compliquée et coûteuse. Un exemple de crowdsourcing de contenu fonctionnant selon le modèle micropaiement⁶², est celui du stock de photos iStockphoto⁶³. L'entreprise créée en 2000 et gérée par Getty Images met en relation sur son site les photographes (amateurs ou non) et les acheteurs d'images. En s'inscrivant, n'importe quel individu peut afficher ses photos sur le site et espérer ainsi qu'un acheteur en manifeste de l'intérêt.

En abordant la question de l'innovation collective, la plupart des auteurs assimilent les communautés de l'innovation avec les communautés d'open source et d'open access. A cet égard, les exemples les plus connus du modèle de crowdsourcing du contenu sont la base de données cartographique libre Openstreetmap⁶⁴ et l'encyclopédie libre Wikipedia⁶⁵. Cependant, comme le souligne Chesbrough (2006), l'innovation ouverte et l'open source, bien qu'ils sont similaires dans leur démarche d'innovation, sont différents dans la mesure où : *"Open innovation explicitly incorporates the business model as the source of both value creation and value capture... While open source shares the focus on value creation throughout an industry value chain, its proponents usually deny or downplay the importance of value capture."*

Dans le cas de iStock, le crowdsourcing s'inscrit fondamentalement dans une logique entrepreneuriale. Il repose sur un modèle d'affaires explicite pensé le plus souvent par l'entreprise. On ne parle donc pas ici de communautés d'individus s'échangeant des informations, mais d'une entreprise qui décide de s'appuyer stratégiquement sur la foule pour réaliser une tâche dans une

⁶² Un service de paiement permettant d'acheter des services ou des contenus de faible valeur unitaire.

⁶³ <http://www.istockphoto.com/>

⁶⁴ www.openstreetmap.fr

⁶⁵ www.wikipedia.fr

logique de production et de création de richesse. Ce point, met l'accent sur l'intention stratégique et le modèle d'affaires et permet ainsi de distinguer les différentes approches du crowdsourcing.

Suite au phénomène de crowdsourcing et dans un contexte organisationnel avec ses propres modes de gestion, besoins et objectifs, l'Open innovation semble être un cadre d'organisation du processus d'échange des savoirs en offrant un modèle à suivre. L'intérêt du crowdsourcing nous permet au final de nourrir notre réflexion sur la nécessaire prise en compte de la foule pour comprendre comment et pourquoi elle participe dans un processus d'open innovation et que nous souhaitons questionner en fonction de nos observations sur notre terrain de recherche.

2.2.2 Open data dans les entreprises

L'externalisation est le deuxième aspect de l'open innovation qui remet l'entreprise en situation de diffusion de ses sources de l'intérieur à l'extérieur. Ceci nous amène à questionner le modèle d'externalisation des données internes : l'open data est-il un modèle que les entreprises peuvent adopter ?

L'open data s'attache dès ses prémices à une vision étatique dans une volonté politique de transparence et de service public⁶⁶. Quand il s'agit d'ouvrir ses propres données, le projet open data d'une entreprise s'inscrit dans une stratégie d'innovation. La décision de l'entreprise d'adopter une politique d'ouverture est liée à l'analyse d'opportunités commerciales directes par la vente de nouveaux produits et services ou indirectes par la création de services gratuits visant à attirer ou fidéliser les clients. Les résultats de l'étude Bluenove (2011) auprès des grandes entreprises françaises démontrent que le retour sur investissement de projets d'ouverture n'est pas la préoccupation première. Le modèle économique repose donc sur la capacité des données ouvertes par l'entreprise à gagner en productivité, en compétitivité, en notoriété voire en image, autant de retombées indirectes souvent difficilement mesurables.

Les interactions de l'entreprise avec ses clients créent de nombreuses données : données de consommation, de géolocalisation, de fréquence

⁶⁶ Voir la deuxième partie sur l'open data dans le secteur public

d'utilisation. L'une des stratégies d'ouverture des données consiste à offrir à chacun de ses clients l'accès libre aux données qui le concernent. L'open data comme un service permettrait ainsi d'améliorer la satisfaction client : il peut ainsi suivre et optimiser sa consommation, adopter le meilleur contrat ou forfait, profiter au maximum d'un éventuel système de fidélisation. Il ne s'agit pas seulement de mettre ces données à disposition du client, encore faut-il lui permettre de s'en servir en développant pour lui une application dédiée ou en lui fournissant des éléments de comparaison. C'est ce que font différents opérateurs télécoms, fournisseurs d'électricité ou d'eau comme GDF Suez.

Les résultats de cette étude montrent que 43% des entreprises ont entrepris une démarche plus ou moins avancée en matière d'ouverture de leurs données. « Cette volonté d'ouvrir les données de l'entreprise s'inscrit principalement dans une stratégie d'innovation. 62% des entreprises interrogées pensent que l'ouverture de leurs données leur apporterait des bénéfices forts ou importants pour développer un écosystème de partenaires innovants. 60% considèrent que l'ouverture de leurs données les aiderait à accélérer le développement de produits et services innovants » Bluenove (2011). Pourtant, le principal bénéfice attendu de l'ouverture de leurs données, selon l'étude, est l'amélioration de la satisfaction de leurs clients.

Les promesses de l'open innovation peinent cependant à se réaliser sans le développement de modèles efficaces d'une stratégie d'open innovation basée sur les données. Le premier challenge, selon Chesbrough and Appleyard (2007) réside dans l'architecture de base ou la plateforme qui sera capable de connecter les différentes pièces de savoir ou tout composent d'innovation. Ce point associe l'open innovation à la problématique de traitement des données discutée dans le cadre du phénomène du Big data.

2.3 Gouvernement ouvert

Le concept d'origine du gouvernement ouvert (Open government) a été élaboré sur la base des principes philosophiques idéalistes d'avoir des citoyens informés sur les actions des gouvernements. Yu et Robinson (2012) retracent les origines conceptuelles du gouvernement ouvert aux États-Unis. L'un des premiers emplois connus du terme « gouvernement ouvert » était en 1953 quand Harold Cross, l'avocat d'un grand journal, a publié un rapport intitulé : « *The People's Right to Know: Legal Access to Public Records and Proceedings* » (Harold, 1953) ou « Le droit du peuple à savoir: l'accès légal à

des documents publics et procès ». En effet, après la deuxième guerre mondiale, pendant les années 1950 et 1960, ce fut une période cruciale de pression pour une plus grande ouverture accordée à la presse et en partie à la Société Américaine des Editeurs de Journaux (ASNE). En temps de guerre, le gouvernement a pris l'habitude de restreindre l'accès à l'information gouvernementale. Une action pour ne pas divulguer ses opérations et ses activités au grand public et à la presse. En conséquence, en 1955, le Congrès américain a créé le sous-comité spécial sur l'information du gouvernement qui a contribué à la législation qui est devenu, en 1966, le « Freedom of Information Act ». Cette loi a été amendée plusieurs fois et la dernière version de 2007 « *Open Government Act* »⁶⁷ a précisé comme objectif de « promouvoir l'accessibilité, la transparence, et l'ouverture du gouvernement »⁶⁸.

Ainsi, les États-Unis étaient le troisième pays, après la Suède en 1766 et la Finlande en 1951, à adopter une loi sur la liberté d'accès à l'information publique qui prévoit que toute personne ait le droit de demander l'accès aux documents ou informations des organismes fédéraux des États-Unis. Le programme *Justice initiative* de l'Open Society Foundations recense, en février 2014, 99 pays qui disposent des lois sur la liberté et l'accès à l'information publique⁶⁹. Ce droit est déclaré dans l'une des quatre lois fondamentales de la Constitution de Suède, le premier pays à adopter, en 1766, la loi sur la liberté de l'information. Le dernier pays ajouté à la liste est La Côte d'Ivoire qui a voté en décembre 2013 une loi qui instaure la possibilité pour tout citoyen d'accéder à l'information et aux documents publics. Dans l'esprit de cette loi, la culture du secret inhérente à la bureaucratie est ébranlée par la demande de la transparence du travail administratif.

D'après Alasdair Roberts (2006), cet exemple de la transparence gouvernementale a servi de source d'inspiration à plusieurs nations à travers le monde en reproduisant des lois similaires dans cinquante-neuf pays. En France, la liberté d'accès à l'information publique a été incitée

⁶⁷ Voir : *Open Government Act*, 2007. <http://www.justice.gov/oip/amendment-s2488.pdf>.

⁶⁸ «An act to promote accessibility, accountability, and openness of government »

⁶⁹ Voir : Open Society Justice Initiative. "Constitutional Provisions, Laws and Regulations," *Right2Info.org*. Accessed May 22, 2014. <http://www.right2info.org/laws/constitutional-provisions-laws-and-regulations#section-1>.

essentiellement par la loi *De la liberté d'accès aux documents administratifs*, n°78-753 du 17 juillet 1978 qui permet à tout citoyen de consulter ou se faire communiquer un document administratif. Boisnier-Bardou et Isabelle Pailliar (2012) décrivent le contexte des années soixante qui a abouti à des revendications d'accès à l'information publique. « La montée en puissance des revendications est liée à une configuration de la société française dans les années soixante marquée par l'apparition et le développement des classes moyennes dont la formation favorise l'accès aux savoirs techniques et gestionnaires. Un accès pour résister également à la situation de monopole qu'entretiennent les stratégies des groupes de la presse quotidienne régionale ».

Ce bref aperçu historique montre que le droit d'accès à l'information est à la base du concept du gouvernement ouvert. Ainsi, le concept «open government» est une présomption d'ouverture et de transparence en faveur de la divulgation et il est basé en pratique sur l'accès du public aux informations publiques. Ces principaux attributs seront ensuite discutés et mis en articulation, dans la partie suivante, avec l'« Open data dans le secteur public », nous ouvrant alors la voie vers une étude des données publiques ouvertes propre à notre travail de recherche.

Conclusion du chapitre 2

Dans ce chapitre nous avons proposé une analyse de la notion d'Open, et ce par la mise en exergue de trois terrains d'application différents. Nous avons tout d'abord présenté la cyberculture qui met en relief toute la complexité liée au développement d'une culture libre sur le web. Dans cette présentation, nous avons voulu souligner que cette culture est en constante évolution, du fait de l'évolution des conditions de partage et de coproduction réglées par les licences. L'attribuer et / ou le partage à l'identique ont été rapidement rejetés par la publication des nouvelles licences du domaine public. Cette culture a atteint, selon nous, son apogée par la publication de nouvelles licences (comme PDDL ou CC0) qui n'accordent aucune restriction à l'usage des données.

Nous avons par la suite présenté la notion d'open innovation en l'articulant avec l'ouverture des données. Dans ce cadre, il est pertinent de dire que la motivation sous-jacente à l'ouverture et l'exploitation des données dans les

entreprises se trouve dans la logique de l'innovation ouverte et consiste à baigner dans une ambiance d'échange avec l'extérieur. L'innovation ouverte et les données ouvertes se recoupent de plus en plus dans des situations d'internalisation (le cas des données publiques ouvertes) ainsi que dans des situations d'externalisation. Cependant, le premier challenge réside dans le développement de modèles efficaces pour l'intégration de ces données dans l'entreprise pour être valablement exploitée lors d'une prise de décision.

Enfin, nous avons abordé la notion de gouvernement ouvert en retraçant ces origines conceptuelles aux premières revendications à l'accès aux documents publics aux États-Unis dans les années 50. Le concept d'«open government» représente ainsi une présomption d'ouverture et de transparence en faveur de la divulgation et il est basé en pratique sur l'accès du public aux informations publiques.

Conclusion de la première partie

Cette première partie nous a permis de poser les bases théoriques et contextuelles de notre recherche. Dans un premier temps, nous avons présenté la notion de données en s'interrogeant précisément sur les distinctions et les relations établies entre donnée, information, connaissance et sagesse. Ainsi, de l'articulation de diverses définitions, nous considérons les données comme un ensemble de symboles numériques lisibles par machine. Les « données brutes » sont des symboles enregistrés par des capteurs non humains. De même, les données structurées permettent une forme de traitement plus facile par les ordinateurs en vue d'une nouvelle réutilisation. Par la suite, nous inscrivons l'émergence des données dans un nouveau contexte caractérisé par une évolution des modes de production, de mise en réseau et d'exploitation de données. L'évolution du web est ainsi présentée, nous permettant de souligner l'impact du web 2.0 sur la production et la diffusion des données sur les différents dispositifs sociotechniques ainsi que les technologies du web sémantique introduisant les caractéristiques d'une donnée ouverte et liée. Nous avons conclu ce premier chapitre par la présentation de la notion de Big data pour mettre en relief les challenges d'exploitation et de traitement des données. Afin de générer de la valeur, il semble alors nécessaire de repenser la structure même des données, voire même la manière dont les données sont publiées sur le web.

Dans le deuxième chapitre de cette partie, nous avons présenté le principe d'ouverture par la mise en exergue de trois environnements différents. Nous avons tout d'abord abordé la cyberculture pour laquelle nous avons mis en relief son aspect collectif et libéral de la circulation des idées sur internet et le rôle central des mouvements favorisant le partage et la coproduction du savoir sans aucune restriction. Ensuite, nous avons présenté la notion d'« Open innovation » issue du monde des entrepreneurs. Par la mise en relief de l'internalisation et l'externalisation des données, nous avons vu que l'ouverture des données est un élément constitutif de l'Open innovation. Enfin, il était nécessaire d'abordé la notion d'« Open government » pour définir la mise en œuvre de la culture d'ouverture dans le secteur public. Dans ce contexte, le principe d'ouverture représente une transparence en faveur de la divulgation et est basé en pratique sur l'accès du public aux informations publiques.

Cette première partie souligne alors les bases exploratoires de notre thèse. Face à un phénomène que l'on nomme couramment « l'open data », nos observations nous ont amené à faire appel à de nombreuses notions, et dont les parties suivantes de cette thèse vont montrer en quoi leurs définitions nous permettent d'explorer l'open data dans le secteur public.

DEUXIÈME PARTIE : OPEN DATA DANS LE SECTEUR PUBLIC

« L'open data est le "must" de l'administration "new look" et de l'open governance »⁷⁰. Cette énonciation dans un article publié sur le Monde.fr en 2011 est illustrative de la place que l'open data occupe dans nos sociétés où l'ouverture des données publiques devient un élément essentiel pour un Etat moderne et intelligent. Dans cette partie nous traiterons donc l'open data dans le secteur public en particulier.

Notre troisième chapitre visera à délimiter l'open data dans le secteur public tel qu'il est abordé par les différents acteurs impliqués dans l'ouverture des données publiques. La notion d'open data est encore peu traitée par les recherches scientifiques et s'avère tout d'abord modelée par des politiciens, juristes, informaticiens, ou encore des organisations ou des individus favorisant l'ouverture des données publiques. Il conviendra donc de déconstruire cette notion par l'analyse critique de leurs discours, c'est-à-dire des définitions qu'ils donnent à cette notion ou encore de certaines idéologies sous-jacentes présentes dans les dires de ces acteurs. Ce faisant, il conviendra alors de questionner de nombreuses notions proches, qui plus est, lorsque l'open data est associé aux organisations publiques, comme la transparence ou la participation citoyenne si tant que l'open data apparaisse nécessaire pour le gouvernement ouvert.

Dans le quatrième chapitre, plusieurs initiatives mettant en œuvre des dynamiques en évolution feront l'objet d'une présentation afin d'identifier les acteurs aux niveaux international et national et fournir quelques éléments relatifs à l'expansion du mouvement open data et leurs influences directes ou indirectes sur l'implémentation de l'open data au niveau national.

A la fin de cette partie, nous présenterons l'open data comme une extension des pratiques de communication et de diffusion des informations exercées par les institutions publiques et qui ne représentent pas un phénomène nouveau. Au-delà d'une simple diffusion de données publiques, il s'agit désormais de s'interroger sur de nouveaux principes qui identifient ce que

⁷⁰ Voir : "Des statistiques d'Etat à géométrie variable." *Le Monde.fr*, December 12, 2011. http://www.lemonde.fr/idees/article/2011/12/14/des-statistiques-d-etat-a-geometrie-variable_1618346_3232.html.

c'est une donnée ouverte diffusée dans des conditions différentes du système traditionnel de communication et de diffusion des données et des informations publiques existant. Il nous paraît donc utile et nécessaire de revenir sur des notions comme « document administratif » et « informations publiques » et distinguer entre « accès » et « diffusion ». Ensuite, nous découvrons ce système traditionnel pour mettre en avant ses critiques et comprendre les impulsions au changement et à l'adoption de l'open data.

Chapitre 3 : Données publiques ouvertes : définition et intérêt

Dans notre contexte de recherche, nous mettons en avant dans ce chapitre les aspects et définitions de l'open data dans le secteur public. Nous proposons tout d'abord de présenter les caractéristiques d'une donnée publique ouverte en pointant les divergences entre les définitions propres aux acteurs et celles vues dans les chapitres précédents à l'aune des définitions de l'open knowledge et l'open linked data.

Ensuite, nous aborderons l'intérêt de l'open data dans le secteur public par la prise en compte de trois approches, politique, économique et technologique et leurs apports. En le mettant en corrélation avec le gouvernement ouvert, il conviendra d'articuler l'open data avec des notions proches comme la transparence et la participation citoyenne. Nous essayons d'y révéler de même ce que l'open data relève d'innovation ouverte dans le secteur public. L'approche économique nous donne l'occasion d'interroger la valeur économique des données publiques. Enfin, nous présentons l'approche technologique de l'open data en mettant en avant le déploiement des nouvelles technologies d'information et de communication (NTIC) dans le secteur public. Des notions comme « gouvernement électronique », « État plateforme, « ville numérique » nous invitent à s'interroger sur la place et la corrélation de l'open data avec d'autres technologies et outils.

Ce chapitre consiste donc à la fois des discours des acteurs impliqués dans l'ouverture des données publiques, ainsi que d'une construction théorique de cette notion en s'appuyant sur les différents cadres théoriques présentés dans le chapitre 2.

3.1 Une définition opérationnelle de l'open data

La première définition des données publiques ouvertes (*Open Government Data*) nous ramène en Californie aux États-Unis où a lieu, en Décembre 2007, la réunion d'un groupe de travail de 30 personnes coordonnées par Tim O'Reilly de O'Reilly Media et Carl Malamud de Public.Resource.Org, avec le partenariat de l'association Sunlight Foundation, Google, et Yahoo⁷¹. Avant de présenter les caractéristiques d'une donnée publique ouverte, il est intéressant de décrire le contexte social de cette réunion qui explique en grande partie le choix de ces principes. En effet, cette réunion constitue un point de rencontre de deux communautés bien distinctes :

1. la première communauté est celle des avocats du « gouvernement ouvert » classiques dont l'objectif est la publication et le partage de matériel du domaine public représenté par Public.Resource.Org.
2. La deuxième rassemble les partisans des logiciels libres et le mouvement open source représenté par Tim O'Reilly.

Leur objectif était de trouver l'intersection entre le « gouvernement ouvert » et les « données ouvertes » (voir la figure 6) et le débat politique et technique s'est polarisé autour des données publiques.

⁷¹ Voir : Open government working group. "The 8 Principles of Open Government Data." In Open Government Data Workshop. Sebastopol, CA: opengovdata.org, 2007. <http://opengovdata.org/>.

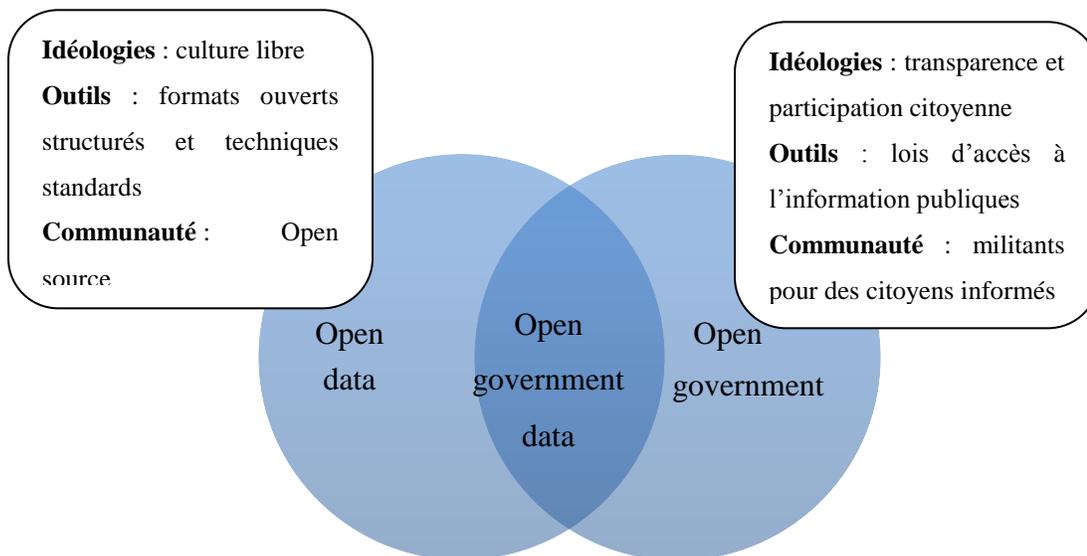


Figure 5. Relation entre Open data et Open government

Huit principes ont été énoncés par l'assemblée pour définir une donnée publique ouverte :

- I. **Complète** : toutes les données publiques stockées dans n'importe quel format y compris mais non limité aux documents, bases de données, transcriptions et enregistrements audio / visuel sauf celles soumises à des restrictions de confidentialité et de sécurité. Une offre complète de données publiques pose le conflit lié aux « données fermées » et l'open data prend la forme d'une injonction à l'ouverture de données qui ne sont pas accessibles. Nous mettons ce point en lien avec la question de l'offre de données publiques présentée dans le chapitre suivant.
- II. **Primaire** : les données sont ouvertes telles que collectées à la source, avec le plus haut niveau possible de granularité, et non dans des formes agrégées ou modifiées. Ce principe est souvent lié au caractère « brut » des données que nous avons déjà évoqué dans le chapitre 1. Ce terme ambigu et complexe met en relief les activités de collecte et de traitement de données qui devraient être perçues et examinées avec attention.
- III. **Opportune (Timely)** : les données sont mises à disposition aussi rapidement que nécessaire pour préserver la valeur des données. La fraîcheur des données impose un dynamisme de l'offre.

- IV. Accessible : les données sont disponibles pour le plus large éventail d'utilisateurs pour la plus large gamme d'applications. C'est un critère essentiel à prendre en compte face par exemple aux usages commerciaux des données publiques discutés dans le chapitre suivant.
- V. Lisible par machine : les données sont assez structurées pour permettre le traitement automatisé par ordinateur. La notion de données structurées a été déjà soulevée dans le chapitre 1. Il s'agit bien de données numériques mais le degré de structuration reste à préciser.
- VI. Non discriminatoire : les données sont disponibles pour n'importe qui, sans obligation d'enregistrement. La non-discrimination peut amener à percevoir les internautes ou utilisateur comme « anonymes », pourtant cela a des conséquences comme nous allons le voir dans le chapitre sur l'usage.
- VII. Non exclusive : les données sont disponibles dans un format sur lequel aucune entité n'exerce un contrôle exclusif. Cela désigne les formats ouverts non propriétaires.
- VIII. Non pas soumises à une licence (Licence-Free) : le diffuseur n'impose pas des termes sur l'usage des données. La non-licence nous invite par la suite à examiner le choix des licences adoptées dans les projets open data.

Six années plus tard, dans une optique de développer des indicateurs d'évaluation des données publiques ouvertes, l'Open Data Index⁷², publié en 2013 par l'OKF a retenu neuf critères pour classer les pays selon « degré d'ouverture » de leurs données. Les critères représentent les niveaux d'ouverture d'ordre croissant : la donnée existe et elle est numérique, accessible, gratuite, en ligne, lisible à la machine, disponible en vrac, sous licence libre et mise à jour.

L'organisation Sunlight Foundation, fondée en Avril 2006 avec un objectif d'accroître la transparence et la reddition de comptes au États-Unis, donne, quant à elle, des recommandations qui ne figurent pas dans les définitions

⁷² Voir : Open Knowledge Foundation. "Open Data Index," October 28, 2013. <https://index.okfn.org/country/>.

précédentes. Dans son guide⁷³ mis à jour toutes les années depuis 2012, l'association précise une série de points à prendre en considération en appliquant une politique open data à l'ouverture des données publiques en s'inspirant de plusieurs applications, politiques et pratiques de plusieurs villes américaines comme Oakland, New York, Louisville et autres. En étudiant l'ouverture des données en tant que processus, le guide ajoute les principes de « permanence » et de la « documentation » des données publiées. Les données permanentes devraient être disponibles à une adresse Internet stable indéfiniment et dans un format de données stable aussi longtemps que possible.

Au vu des différents critères de l'open data soulignés dans de nombreuses définitions issus principalement des discours des associations et présentés dans ce chapitre et le chapitre précédent, nous proposons de les synthétiser et les schématiser dans le tableau suivant.

Tableau 3. Critères de l'open data

Offre de données	Existence des données. Ce point met en question le problème de productivité des données publiques au sein des établissements publics.
	Accessibilité à toutes les données publiques. Une offre complète de données publiques pose le conflit lié aux « données fermées » et l'open data prend la forme d'une injonction à l'ouverture de données qui ne sont pas accessibles
	Données primaires non interprétées
	Fraicheur et mise à jour
	Permanence des données
	Documentation sur les données
Conditions d'usage	À tous les usagers sans discrimination, sans exclusivité
	A des usages commerciaux. Notons que l'open definition de l'OKF laisse le choix au diffuseur de données de soumettre les œuvres dérivées aux mêmes conditions de réutilisation que l'œuvre

⁷³ Voir : Sunlight Foundation. "Open Data Policy Guidelines." *Sunlight Foundation*, 2014. <http://sunlightfoundation.com/opendataguidelines/>.

	originale.
	Non pas soumises à une licence (Licence-Free). Le diffuseur n'impose pas des termes sur l'usage des données pourtant plusieurs licences libres imposent l'attribution.
	Gratuité. Même si l'open definition a introduit le principe de gratuité de la réutilisation, elle n'empêche pas l'établissement de redevances pour couvrir les coûts de communication ou reproduction des données.
Spécificités techniques	Données numériques, donc lisible par machine. Les cinq niveaux d'ouverture de Tim Berners-Lee élargie le spectre des données numériques des moins structurées (pdf) aux plus structurées (Open linked data) ⁷⁴
	Formats non prioritaires et standardisés
	Données disponibles en vrac (Bulk data) c'est à dire un grand volume de données est disponibles et téléchargeables d'un seul fichier. Le Flux de données en vrac (Bulk data feed) représente le mécanisme d'importer les données mises à jour en temps réel par des applications informatiques.

Bien que pour qu'une donnée publique soit qualifiée d'ouverte « elle doit répondre à des critères techniques, juridiques et économiques qui visent à en faciliter la réutilisation » (Chignard, 2012). Les définitions apportées par les différents acteurs évoquent les caractéristiques d'une donnée ouverte mais qui manquent, selon nous, de critères rigoureux. Si l'ouverture est une conquête progressive, il y a des « degrés d'ouverture » selon l'effort et les techniques disponibles. Pour Henri Verdier (Cité par Blanc, 2013), directeur de la mission Etalab, la définition la plus stricte de l'open data est basée sur le format ouvert, la gratuité et la liberté de réutilisation. Comme nous pouvons le conclure en abordant l'open data par ses acteurs, les approches fournies se veulent opérationnelles dans le sens qu'elles permettent à chaque acteur de la définir en considérant ses objectifs et ses compétences.

⁷⁴ Voir partie 1.3

Ce point nous invite par la suite à s'interroger des stratégies des collectivités locales dans le cadre des projets open data locale.

3.2 Le nouveau modèle du gouvernement ouvert

En 2009, un des premiers projets de l'ouverture des données, celui de l'administration Obama aux Etats-Unis a été inscrit sous la *Directive du gouvernement ouvert*⁷⁵. Selon O'Reilly (2010), le gouvernement ouvert est l'aboutissement et la raison d'être du mouvement Open Data et les notions d'Open Data et d'Open Government sont indissociables. Nous devons alors nous demander : Comment l'open data est-il devenu le nouveau modèle du gouvernement ouvert?

La directive pose les trois piliers du gouvernement ouvert :

1. la transparence du gouvernement
2. la participation
3. la collaboration

Si le gouvernement ouvert sous-entend l'open data, il convient alors de poser une série de questions clés sur la place de l'open data dans la réalisation de ces trois objectifs : Comment l'ouverture des données publiques est-t-elle un moyen qui favorise la transparence ? Quelles formes de participation et de collaboration citoyenne permet-elle ?

3.2.1 Open data et transparence

L'idée de base à l'origine de cette question sur la relation entre l'open data et la transparence repose sur le concept même de la transparence. Grimmelikhuijsen et Welch (2012) donnent la définition suivante de la transparence : « La transparence est la disponibilité des informations sur une organisation ou un acteur, qui permettent aux acteurs extérieurs de surveiller les travaux internes ou les résultats de cette organisation ou de cet acteur ». Ce résultat correspond avec ce qu'écrivait Thierry Libaert (2003) dans son livre *La Transparence en trompe-l'oeil*, la transparence est

⁷⁵ Voir : États-Unis. Executive office of the President memorandum for the heads of executive departments and agencies. 2009. http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf

devenue l'argument ultime d'une nouvelle idéologie technique et médiatique ». Elle sert de prétexte ou de justification à des discours et des actions parfois obscures. Et en se référant à l'idée que « le message, c'est le medium » de McLuhan, il pointe que « la transparence dépendra toujours du media considéré et de sa relation avec l'objet. Plutôt qu'une faculté de mise en lumière, la transparence doit donc être définie comme un processus relatif d'accessibilité de données informationnelles ». Notons alors que la divulgation des informations (et les données) publiques est essentielle pour garantir la transparence des organismes publics.

En outre, en étudiant la réforme des managements publics, Pollitt et Bouckaert (2011) indiquent qu'il y a une alternance entre une priorité urgente pour une efficacité accrue des services publics, et ensuite, une insistance forte sur la transparence et la redevabilité d'Etats de manière à justifier ces actions et décisions. Ainsi, la transparence est un des objectifs majeurs des réformes dans le management du secteur public et qui prend de plus en plus d'ampleur et d'importance par rapport à ce qu'elle était il y a vingt ou trente ans, affirment Pollitt et Bouckaert.

En effet, la question sur l'open data, en tant qu'outil de transparence et un moyen pour rétablir la confiance entre l'Etat et les citoyens, nous amène à considérer la corrélation entre les trois phénomènes : la divulgation des informations publiques, la transparence et la confiance. La divulgation des informations publiques produira un effet : « la transparence », qui restaure la confiance du public dans le gouvernement. Alasdair Roberts (2006) souligne le lien entre « la nouvelle culture d'ouverture et l'engagement vigoureux à la liberté de l'information » et le « renouveau de la confiance » dans le discours des politiciens dans les mois avant la mise en œuvre, en janvier 2005, de la loi à la liberté d'information publique. Plusieurs auteurs (Cook, Jacobs et Kim, 2010) considèrent que les « informations factuelles et détaillées » sur les processus et les activités de l'Etat augmentent la confiance chez le public. Ainsi, la problématique de l'open data dépasse le cadre de la transparence garantie par la publication des informations publiques. La transparence ne consiste pas uniquement à rendre des informations accessibles, dans le sens traditionnel hérité du gouvernement ouvert. Le mouvement open data tente de renouveler la question de la transparence par l'élaboration et la communication en ligne, par les administrations, des détails fines de leurs actions et décisions. Si l'accès aux

informations publiques est un indicateur d'évaluation de la transparence des administrations publiques, Van de Walle et Roberts (2008) indiquent que :

« nous vivons une époque de quantification de la performance. Il n'est désormais plus possible de croire qu'un organisme public exerce sa fonction efficacement ou de faire confiance aux professionnels au sein de cet organisme. Nous voulons des données qui nous permettront de juger et de comparer le comportement des prestataires de services. En ce nouveau millénaire, le nombre de régimes conçu pour satisfaire à cette demande de données de performance a considérablement augmenté ».

Dans le concept du Nouveau Management Public (NMP), « la transparence et la mesure chiffrée de la performance sont deux composantes qui semblent constituer une combinaison idéale » (Grimmelikhuisen, 2012). Dans le débat sur l'ouverture des données publiques et la transparence, les partisans de l'open data croient que les données constituent une *valeur d'évidence* et s'imposent comme *preuve objectives* de l'authenticité des administrations. Les données brutes sans interprétation sont perçues comme *impartiales et objectives*. L'ancien président de la mission etalab, Severin Naudet⁷⁶, affirme que l'open data est une démarche qui permet aux citoyens de fonder leurs choix « non plus sur des positions idéologiques à priori, mais sur des analyses informées et nourries de chiffres objectifs » (Naudet, 2011). Le pouvoir des données s'explique par leur précision et leur volume. L'Open data sera « un outil numérique qui peut aider à *rétablir la confiance* entre l'Etat et le citoyen » (Leloup et Numès, 2011), une conviction annoncée en novembre 2011 par l'ancienne ministre chargée de l'économie numérique et la Maire de la ville de Longjumeau Nathalie Kosciusko-Morizet⁷⁷. Elle ajoute : « le moteur de la démocratie, comme de l'économie de marché, c'est la confiance. On donne, et on reçoit. Le numérique peut aider à rétablir cette confiance. Il y a un caractère éruptif de certains sujets comme les places en crèche, l'accès au logement social, le nombre de bénéficiaires des aides sociales octroyées par les collectivités territoriales, les tarifs de cantines, les dérogations pour la carte scolaire » (Leloup et Numès, 2011). Dans ce cas,

⁷⁶ Directeur de la mission Etalab depuis sa création en février 2011 jusqu'au Octobre 2012. Prédécesseur de Henri Verdier

⁷⁷ Nathalie Kosciusko-Morizet était maire de la commune de Longjumeau entre 2008 et mars 2013.

l'open data augmente le niveau « d'authenticité » des politiciens pendant leurs discours en se référant à des données déjà publiées.

Pour Dominique Cardon (2011), sociologue au Laboratoire des usages (SENSE) d'Orange Labs, l'open data est une technique de dévoilement de secrets en s'appuyant sur les « *vérités cachées dans les chiffres des statistiques* ». Il ajoute :

« le mouvement open data refuse les techniques d'échantillonnages au profit d'une visée d'exhaustivité complète et de granularité la plus fine possible des données. L'exigence de complétude et de granularité que porte le mouvement des données ouvertes est nourrie par une revendication de transparence à l'égard des institutions productrices de données, lesquelles cachent parfois dans leurs tiroirs des secrets de fabrication, des techniques de redressement et des opérations de recodage » Cardon (2011).

Dans le même esprit, le rapport sur l'ouverture des données publiques, menée pour la Délégation Aux Usages de l'Internet, considère l'open data comme un outil puissant d'amélioration de la transparence de l'État : « En donnant à tous les citoyens qui le souhaitent accès aux *détails* du fonctionnement des institutions publiques, cette ouverture nourrit et éclaire le débat démocratique. Elle peut contribuer à baser les choix politiques sur des analyses informées et ancrées dans le réel plutôt que des positions idéologiques formulées a priori » (Lacombe et al., 2011). Ainsi pour la communauté qui défend le concept du gouvernement ouvert dont l'objectif est la publication et le partage de matériel du domaine public, l'open data est un outil pour surveiller les acteurs publics en se basant sur le concept de « public scrutiny ». Les données ouvertes deviennent un instrument d'examen le plus minutieux de la part de l'opinion publique.

3.2.2 Open data, participation citoyenne et innovation

Selon l'association Décider Ensemble (2012), l'open data a un « impact démocratique » en matière de participation des citoyens dans la vie publique. Nous pouvons faire une distinction entre deux facettes différentes de la participation citoyenne. D'abord, avoir l'accès à des données précises concernant un problème ou un projet public. Le deuxième modèle de

participation, le plus ambitieux, correspond à la volonté d'associer directement les citoyens pour aboutir à une forme de collaboration citoyenne. Au-delà d'une démarche très circonscrite par un citoyen informé, une politique de gouvernement ouvert cherche alors à faire participer les citoyens par l'usage et la création de nouvelles informations par la collaboration collective. Selon une logique d'« intelligence collective » dans le cyberspace que Pierre Levy (1994, p.29) définit comme « *une intelligence partout distribuée, sans cesse valorisée, coordonnée en temps réel, qui aboutit à une modélisation effective des compétences* ». Ainsi l'État reconnaît qu'il y a une expertise dispersée au sein des citoyens et qui repose sur l'idée issue du web social que ces citoyens vont développer des communautés d'actions collectives, dans un espace public ouvert à tous et en s'appuyant sur les nouveaux liens sociaux qu'ils tissent.

Le public est à l'extérieur des frontières organisationnelles et en dehors du contrôle de la hiérarchie. Dans un système ouvert, le public devient partie intégrante du système de traitement de données et peut traiter, enrichir et les combiner avec d'autres sources et pourrait même recueillir leurs propres données (par exemple grâce à l'utilisation de leurs téléphones mobiles). Cela ressemble à un changement dans les frontières traditionnelles entre les organismes publics et le public dans lequel pratiquement tout le monde a accès aux données. Labelle et Le Corf (2012) décrivent l'ouverture des données dans le secteur publique comme un « détachement » qui donne aux données un nouveau statut plus autonome de son administration d'origine en faisant le rapport avec la notion de « Créativité » : « Il s'agit de faire sortir des données de leur contexte initiale de production pour leur offrir un nouveau cadre d'interprétation et de traitement dans de nouveaux contextes sociaux ».

Le principe d'open data s'insère dans ce contexte en permettant aux citoyens de réutiliser les données ouvertes afin de créer ou améliorer les services sociaux. Prenons un cas de figure exemplaire de ces services sociaux : le développement d'applications web et mobiles créés en totalité ou en partie des données publiques ouvertes dans l'objectif d'améliorer la vie quotidienne des citoyens. En effet, l'explosion de l'usage des appareils mobiles connectés (smartphones, tablettes), est un facteur important dans la transformation des interactions entre le citoyen et son territoire, duquel les administrations publiques sont pleinement conscientes. L'invasion des applications mobiles urbaines influence sur les perspectives des services

publics sur mobiles (Anthes, 2011). Il s'agit d'un défi pour une ville de rendre possible la production de nouveaux services associés à l'expérience d'un utilisateur en mobilité. Le Portail national des applications et de services aux citoyens sur téléphone mobile, Proxima Mobile, créé par la Délégation aux Usages de l'Internet (DUI), promeut l'utilisation, le développement d'applications mobiles d'intérêt général en signalant que « les services sur mobiles modifient la gestion des services urbains et progressivement aussi la vie locale elle-même. Les villes amorcent également des expérimentations de nouvelles formes de services structurants » (Proxima Mobile, 2010). Les perspectives des services mobiles pour les collectivités locales sont alors marquées par un triple choix:

- financer elles-mêmes les applications (en confiant la réalisation à un prestataire),
- nouer des partenariats spécifiques avec des entreprises chargées du développement des applications (en mettant les données municipales à disposition),
- mettre à disposition les données issues de leurs services et permettre à des individus, des associations ou des entreprises de développer des applications utiles.

Confrontés à ces alternatives, les administrations qui choisissent la dernière solution représentent un cas classique d'externalisation vers la foule que nous avons déjà abordé sous l'angle d'open innovation. Il s'agit d'une alternative de production qui repose sur la contribution des citoyens suite à la démarche d'ouverture des données publiques. Sur le principe des innovations permises par l'exploitation des données en tant que matière première, l'ouverture des données devrait permettre la création de nouveaux marchés, produits et services sociaux. L'innovation ouverte et les données ouvertes ne s'excluent pas mutuellement mais se recoupent de plus en plus dans des situations d'externalisation. Après cette articulation entre Open innovation et l'ouverture des données, nous pouvons donc dire que l'ouverture des données est un élément constitutif de l'Open innovation dans le secteur public.

3.3 Données publiques et croissance économique

Après avoir mis en avant l'intérêt de l'open data en s'appuyant sur des promesses de transparence et participation citoyenne, les promoteurs de l'open data diffusent et médiatisent un discours économique sur l'exploitation des données publiques par les entreprises. Ce discours rejoint actuellement les promesses du Big data et la valeur économique des données discutées dans le chapitre 1. Néanmoins, l'engouement pour l'exploitation des informations issues du secteur public de la part des entreprises trouve son origine dans les revendications de l'Union Européenne à ouvrir les informations publiques. Dans « Le Livre vert sur l'information émanant du secteur public dans la société de l'information », publié en 1998, la commission européenne présente l'information du secteur public comme une ressource clef tout en soutenant la richesse économique et en montrant les opportunités qu'elle apporte pour la croissance économique. Une étude sur les marchés de services d'information électroniques montre que l'information du secteur public est une ressource clef pour une large partie des services d'information, en étant soit le sujet central pour l'information administrative/politique et juridique, soit la matière première essentielle pour le profil des entreprises, les informations sur les brevets, information à caractère scientifique, technique ou médical (STM). « L'information émanant du secteur public joue un rôle fondamental dans le bon fonctionnement du marché intérieur et dans la libre circulation des marchandises, des personnes et des services. En l'absence d'une information publique de nature administrative, législative, financière ou autre, disponible dans des conditions faciles et conviviales, les acteurs économiques ne sont pas en mesure de prendre les décisions d'une façon pleinement informée » (Commission Européenne, 1998, p.3). Une mise à disposition sans entrave de l'information publique devient préalable à la compétitivité de l'industrie européenne. Aujourd'hui dans le contexte du développement du discours sur les données numériques, la Commission Européenne (2011), dans le cadre de sa stratégie numérique, a estimée en 2011 que le potentiel économique des données publiques est à hauteur de 40 milliards euro par an.

Cette approche inscrit alors les données publiques dans le cadre globale du marché de données ancré dans des approches économiques où l'open innovation semble une motivation sous-jacente à l'ouverture et l'exploitation des données publiques par les entreprises. Nous pouvons cependant

distinguer cette « approche économique orientée vers les entreprises » de celle de l'Open Government de Obama considérée, selon nous, comme une « approche politique orientée vers le gouvernement ». L'Open Government adopte une autre direction : les données partagées aideront le gouvernement et les administrations publiques à améliorer leurs efficacités. Ainsi, dans une perspective bureautique, la réutilisation des données par le grand public se transforme en de modèles de participation et de collaboration qui permettent d'apporter des idées et de l'expertise afin que leur gouvernement puisse établir des politiques plus avancées et offrir de nouveaux services.

Pour conclure, nous souhaitons souligner que les modèles de business sous-tendus par des informations publiques, ne sont pas bien identifiés. En dehors du modèle de partenariats entre une entreprise et un organisme publique, et par manque de recherches scientifiques, il sera difficile d'appréhender comment les données ouvertes du secteur public seront-elles une source suffisante d'innovation pour les entreprises ou un moyen de réduire les coûts de certaines activités? Comment transformer les données collectées en valeur ajoutée pour l'entreprise?

3.4 Open data et déploiement des TIC dans le secteur public

L'administration publique est un processus intensifié par le numérique, de telle manière que nous parlons de « gouvernement électronique », « gouvernement 2.0 » et « État plateforme » dont le fonctionnement passent par les technologies de l'information et de la communication (TIC). De même, se multiplient des expressions qui indiquent l'ancrage territorial des TIC, comme par exemple la « ville numérique », « ville 2.0 » et la « cité digitale ». L'approche technologique de l'open data nous invite à s'interroger tout d'abord sur les présupposés sous-jacents à ces notions, et par la suite sur la place et la corrélation de l'open data avec le déploiement des technologies et dispositifs numériques dans le secteur public.

La question des TIC occupe une place non négligeable dans l'agenda politique depuis les années 1990. Plusieurs initiatives et programmes ont été lancés dans ce domaine, le dernier celui de l'« État plateforme » qui illustre

ce que les directions entendent par un accès universel aux services publics⁷⁸. En effet, le déploiement des TIC dans le secteur public est marqué par des héritages hétérogènes. Tout d'abord, ces initiatives sont engendrées par le développement de « la société de l'information » qui offre tant de promesses aux plans social et économique. D'après Flichy et Dagiral (2004), les pouvoirs publics investissent de manière accrue dans les solutions technologiques pour deux majeurs raisons : la simplification administrative et la modernité. Ainsi, en complément des objectifs liés à la généralisation de l'usage des technologies de l'information et de la communication, la dématérialisation des documents pour simplifier les procédures administratifs, et le développement de portails sur internet comme forme de guichet unique virtuel à distance, la modernisation apparaît clairement dans la politique des pouvoirs publics. Comme le souligne dans la même lignée Isabelle Paillart (2002): « La modernisation fait partie ici du programme de développement de la société de l'information mais elle accompagne les transformations organisationnelles des services publics (efficacité, rapidité de traitement, restructuration interne, performance individuelle de chaque agent...) ». Selon Boudreau (2009), les organisations publiques ont profité aussi des gains d'efficience et de productivité que procure l'utilisation des nouvelles technologies de l'information⁷⁹. À côté de l'action de l'État, les discours portant sur le rôle des collectivités territoriales dans le développement de la société de l'information mettent en avant une panoplie de pratiques très diverses : création d'un réseau généralement à haut débit avec l'appui financier des collectivités concernées, l'informatisation de l'administration communale, l'électronisation de la vie démocratique (forum en ligne permettant de débattre des problèmes locaux, projets de vote électronique (Flichy et Dagiral, 2004). Ces projets se cristallisent autour des notions labels comme « ville numérique », « cité digitale », « campus numériques » ou « région numérique ».

⁷⁸ Voir : Blanc, Sabine. «L'État Plateforme, Vraie Source de Services Publics Innovants Ou Cache-Misère?» *Lagazette.fr*, 2015. <http://www.lagazettedescommunes.com/323547/letat-plate-forme-vraie-source-de-services-publics-innovants-ou-cache-misere/>.

⁷⁹ Voir, entre autres, Danziger, John N. et Kim V. Andersen. « The Impacts of Information Technology on Public Administration : an Analysis of Empirical Research from the "Golden Age" of Transformation », *International Journal of Public Administration*, 25, 5, 2002, p. 591–627.

La question des TIC dans le secteur public rejoint aussi le discours sur la cyberculture et le gouvernement ouvert que nous avons déjà évoqués dans les chapitres 2 et 3. Les notions de « gouvernement 2.0 » et de « ville 2.0 » sont non seulement pour souligner les évolutions de dispositifs techniques propres à l'Internet et au web, mais aussi un ensemble de pratiques et d'usages induits par le développement de ces dispositifs par les administrations publique. Selon Kaplan et Marcou (2009), la ville 2.0, est une plateforme d'innovation ouverte qui « se focalisent, non pas sur la technique, mais sur les caractéristiques fondamentales qui ont fait le succès du web 2.0 : la construction de services par assemblage de briques ouvertes, l'architecture de participation et le fonctionnement comme plateforme ». Pour Alberede et Eychenne (2012), dans une ville 2.0, « les innovations urbaines et numériques convergent pour créer de nouvelles dynamiques aussi bien portées par les acteurs publics, les entreprises que les citoyens ». La ville, avec l'aide du numérique, devient une plateforme d'innovation ouverte par la contribution collective des habitants pour la construction et l'aménagement du territoire. Selon Paillart (2012), l'enjeu de la mise en place des nouvelles techniques est la réorganisation des structures, des services, des modes de collaboration, bref, l'organisation des acteurs.

Notre volonté de mettre en exergue l'approche technologique des projets open data dans le secteur public repose sur nos observations empiriques qui repère un ensemble de pratiques et d'outils très diverses : diffuser des infrastructures de réseaux, offrir des connexions à « haut débit », lancer des sites web municipaux, créer des comptes sur les réseaux sociaux et installer des dispositifs numériques interactifs⁸⁰. La ville de Longjumeau catégorise l'open data avec d'autres projets de la « ville numérique » comme le déploiement de la fibre optique pour une connexion à internet en très haut débit, les ateliers d'informatique, l'utilisation des tablettes interactives pour accéder aux services numériques et l'école 2.0. Le projet open data de la ville de Montpellier s'inscrit dans le projet « Montpellier Territoire Numérique » lancé en 2010, qui a pour but de « transformer la relation des habitants au territoire, de repenser l'information et les services urbains, d'inventer de nouvelles formes d'espace public grâce aux nouvelles

⁸⁰ Dans le cadre du projet « Montpellier Territoire Numérique », la ville a installé de grands écrans interactifs dans l'espace public

technologies de l'information et de la communication »⁸¹. Pour transformer la ville en « ville numérique », le projet consiste à créer plusieurs outils et services : des points d'accès Internet wifi gratuits dans la ville, des visites virtuelles de la ville en ligne, un réseau fibre optique, des formations au multimédia, etc. Le conseil général de la ville de Bordeaux annonce que la démarche open data entre pleinement dans la dynamique du programme "Bordeaux Cité Digitale" : « Comme vous le savez, la Ville met à votre disposition depuis plusieurs années déjà des services numériques et des informations *via* son portail bordeaux.fr, ses sites internet culturels et événementiels ou encore ses applications mobiles. Dans cette continuité, la Ville de Bordeaux met à votre disposition un portail open data spécifiquement dédié à l'ouverture de ses données publiques »⁸².

Ces quelques exemples nous amènent à un premier constat : dans une acception classique, l'engagement de l'État et les collectivités territoriales dans un projet open data s'inscrit dans leur volonté de moderniser et de créer de nouveaux services de façon à pouvoir en améliorer le fonctionnement. L'open data trouve écho dans des perspectives pour créer un État plateforme et des villes numériques cependant l'enjeu de la mise en place des projets open data ou d'autres services technologiques réside, comme l'a souligné Paillart (2002), dans l'organisation des acteurs, la réorganisation des structures, des usages et des modes de collaboration, que nous souhaitons présenter et questionner dans la troisième partie.

⁸¹ Voir : "Montpellier Territoire Numérique." *Montpellier.fr*. Accessed July 23, 2014. <http://www.montpellier.fr/3421-montpellier-territoire-numerique.htm>.

⁸² Voir : "Bordeaux, Cité Digitale : Votre Ville Résolument 'Open Data.'" *Open Data Ville de Bordeaux*. Accessed July 10, 2014. <http://opendata.bordeaux.fr/demarche>.

Conclusion du chapitre 3

Dans ce chapitre nous avons proposé une analyse de la notion d'open data dans le secteur public, et ce par une présentation des caractéristiques d'une donnée publique ouverte ainsi que par une analyse de l'intérêt de l'open data dans ce secteur.

Entre définitions opérationnelles et des bonnes pratiques, les acteurs s'intéressent de manière plus instrumentale à la façon dont une organisation peut ouvrir ces données. Le fait de produire une définition opérationnelle de la notion d'open data qui va, en principe, augmenter la validité de ce dernier, nous a mis en face d'un « open data hybride » associant plusieurs éléments et niveaux d'applications. Ceci provoque, selon nous, une confusion sur les principes de base de la notion et par conséquent à une confusion quant aux objectifs des politiques publiques d'open data.

L'open data apparaît comme une nécessité pour les organisations publiques. Empreints d'idéologies sur la transparence ou encore la possible participation des citoyens, les discours des acteurs se révèlent orientés par le développement accru d'un gouvernement ouvert où l'open data répond à un besoin d'efficacité et de transparence. Comme le nouveau modèle du gouvernement ouvert, le mouvement open data tente de renouveler la question de la transparence par l'élaboration et la communication en ligne des détails fins de leurs actions pour la formation de la confiance. Une autre conception, plus économique que politique, exige l'exploitation des données publiques en revendiquant leur valeur économique.

Dans le tableau 4 ci-après, nous proposons de restituer les principales approches vues dans ce chapitre en fonction des intérêts de l'open data dans le secteur public ainsi que les principaux acteurs et apports que nous avons dégagés de ces approches.

Tableau 4. L'intérêt de l'open data dans le secteur public

Approches	Nature	Bénéfices clés	Acteurs principaux
------------------	---------------	-----------------------	---------------------------

Politique	Le droit d'accès aux détails fins et des données quantifiables	Augmenter la transparence Rétablir la confiance	Les citoyens
Politique/Bureautique (Orientée administration)	Open innovation au gouvernement	Profiter de l'Intelligence collective Améliorer la qualité des services publics	Secteur public : les administrations
Politique/sociale (Orientée citoyen)	Collaboration et participation citoyenne	Utiliser des services publics efficaces Créer des services publics innovants	Les citoyens
Economique	Exploitation des données publiques	Prise de décision informée	Le secteur privé : les entreprises
Technologique	Etat plateforme Administration électronique Ville numérique Ville 2.0	Améliorer la qualité des services publics Créer des services publics numériques	Secteur public : les administrations

Chapitre 4. Les initiatives politiques et sociales de l'open data

Après avoir présenté l'open data comme le nouveau modèle du gouvernement ouvert et une extension des pratiques de communication et

de diffusion des informations exercées par les institutions publiques, notre objectif, dans ce chapitre, est de mettre en perspective les initiatives politiques et sociales prises d'une part par les gouvernements en vue de développer une politique open data mais aussi par des acteurs associatifs et des communautés internationales. La mise en perspective de ces démarches devrait souligner l'importance de l'articulation de la dynamique sociale militante pour l'open data et les politiques des gouvernements. Plusieurs initiatives mettant en œuvre des dynamiques en évolution feront l'objet d'une présentation afin d'identifier les acteurs aux niveaux international et national et fournir quelques éléments relatifs à l'expansion du mouvement open data et leurs influences directes ou indirectes sur l'implémentation de l'open data au niveau national.

Ainsi, nous présentons tout d'abord le mouvement open data comme un mouvement mondial adopté par les gouvernements de plusieurs pays et défendu par une communauté internationale. Nous nous intéressons également à la politique de la commission européenne en ce qui concerne l'ouverture des données publiques. Ensuite, nous découvrons la politique nationale de la France en signalant le rôle des acteurs publics en relation avec l'open data ainsi que les think tanks et les associations militantes de l'open data en France.

4.1 Expansion mondiale du mouvement Open data

Une référence immédiate souvent évoquée dans les discours relatifs aux initiatives open data sur le plan international, concerne le portail open data des États-Unis lancé en 2009. Cette première initiative, inscrite sous la démarche d'open government⁸³, est devenue un catalyseur pour d'autres pays qui se sont joints au mouvement de données ouvertes. Bien avant, Washington D.C. était la première ville au monde à ouvrir ces données sur data.dc.gov en 2007 (Tauberer, 2012). Vivek Kundra, le chef du gouvernement de la technologie dans la ville, qui a créé le catalogue de données dans une démarche novatrice, a ensuite occupé le poste de directeur fédéral de l'information du gouvernement (Chief Information Officer CIO) responsable de la création du portail national des États-Unis Data.gov en mars 2009. Le Royaume Uni a lancé en Janvier 2010 le portail Data.gov.uk, suivi par le portail data.gouv.fr ouvert en décembre 2011 en France. Open data Census, une initiative d'OKF, a recensé en octobre 2013, 70 pays impliqués dans le mouvement Open Data (OKF, 2013).

Lancé en septembre 2011 par 8 pays fondateurs, l'Open Government Partnership (OGP) ou Partenariat pour un Gouvernement ouvert est une initiative internationale qui regroupe, en octobre 2015, 66 pays membres (voir la carte ci-dessous) dont la France qui est devenu le 64e pays membre en 2014. Comme tous les pays participant à l'initiative, elle s'engage à respecter les principes définis par la Déclaration de Gouvernement ouvert⁸⁴ : transparence de l'action publique, participation citoyenne à l'élaboration des politiques publiques, intégrité de l'action publique, innovation et utilisation des nouvelles technologies pour moderniser l'action publique, notamment à travers l'ouverture des données publiques.

⁸³ Voir section 3.2

⁸⁴ Voir : Open Government Partnership. "What Is the Open Government Partnership?" *Open Government Partnership*. Accessed October 26, 2015. <http://www.opengovpartnership.org/>.



Figure 6. Les 66 pays membres du Partenariat pour un Gouvernement ouvert (Open Government Partnership, 2015).

Durant le 39^e sommet du G8 dans le comit  de Fermanagh, en Irlande du Nord, le 17–18 Juin 2013, les Chefs d'Etat du G8 r unis ont sign  la Charte du G8 pour l'Ouverture des Donn es Publiques. Cette Charte Open Data du G8 « marque l'ambition collective des Etats Membres de promouvoir une gouvernance plus ouverte et plus transparente : en  tablissant un principe d'ouverture par d faut des donn es publiques, en affirmant le principe de gratuit  de leur r utilisation et en privil giant les formats ouverts et non-propri taires, elle encourage l'acc s de tous   l'information et promeut l'innovation entrepreneuriale, citoyenne et sociale »⁸⁵.

Les initiatives d' tats se sont effectu es   un rythme soutenu par des organisations internationales comme la Banque Mondiale (2012) et les Nations Unis (2013). Nous signalons que les initiatives open data ne se limitent pas aux grands pays mais s' tend aux pays en d veloppement (Hogge, 2010). Cette expansion mondiale rapide des initiatives politiques d'open data des gouvernements est certainement le r sultat des promesses politiques et  conomiques de l'open data mais aussi, comme le signal Samuel Go ta (2012), sous l'effet de l'exemplarit  car les discours pr supposent l'id e d'une « course »   l'open data. Notre vision sur la question d'open data nous am ne   penser   une v ritable  mulation qui consiste   susciter et cr er une dynamique de progression au sein des pays.

⁸⁵ Voir : “Les Chefs d'Etat du G8 signent une Charte pour l'Ouverture des Donn es Publiques.” *etalab*. Accessed June 23, 2014. <http://www.etalab.gouv.fr/article-les-chefs-d-etat-reunis-a-loughe-erne-signent-une-charte-du-g8-pour-l-ouverture-des-donnees-publiques-118576420.html>.

Les gouvernements et villes « pionnières » qui ont réussi ou qui ont acquis une expérience doivent servir d'exemple. Ils sont décrits comme des modèles d'ouverture dont il faut suivre l'exemple. Ainsi les premières initiatives internationales sont considérées comme des catalyseurs qui accélèrent l'expansion mondiale du mouvement open data. A ce niveau, le groupement des acteurs internationaux dans le cadre du Partenariat pour un Gouvernement ouvert ou sous l'angle du G8 est devenu un moyen d'influencer les perceptions des politiques et un groupe de pression sur les pays membres pour se mobiliser rapidement en faveur de l'open data.

4.1.1 Associations internationales

En dépit de la jeunesse du mouvement, un effort et un progrès considérables sont accomplis par L'OKF⁸⁶, l'organisation de référence internationale dans le domaine. Créée en 2004 en Royaume-Uni, l'organisation joue un rôle central dans la promotion de la connaissance ouverte. Elle a donné une définition du terme « Open » pour former, comme nous avons vu, une base intelligible compte tenu de la complexité inhérente à ce concept. Les activités de l'OKF s'organisent autour plusieurs projets. Elle a créé un ensemble d'outils juridiques pour contrôler par des licences ouvertes l'usage des données. En plus, elle a publié Open data census et Open data index, deux outils de recensement et d'évaluation des pratiques open data des gouvernements dans le monde. L'association a offert également une solution technique pour libérer les données en développant le logiciel libre de gestion des données CKaN. En France, la communauté urbaine de Bordeaux et le département de Gironde l'utilisent comme plateforme pour ouvrir leurs données. D'autres projets internationaux ont contribué, selon nous, à la mobilisation et la création d'une communauté internationale autour l'open data comme : la journée internationale de l'open data, Open Government Data Camp et le catalogue modiales des données publiques ouvertes.

La « Journée internationale de l'open data » est un grand événement pour la communauté mondiale de l'open data à l'initiative de l'OKF. Initialement, David Eaves, activiste et membre du Groupe de travail de l'OKF sur l'Open Government Data et consultant sur le gouvernement ouvert au Canada, a lancé la première version de la journée internationale des données ouvertes

⁸⁶ 'Open Knowledge Foundation'. <https://okfn.org/>.

en Décembre 2010⁸⁷. Depuis, cet événement est organisé toutes les années simultanément dans plusieurs villes dans le monde. Le dernier International Open Data Day Hackaton⁸⁸ a eu lieu le 22 février 2014 dans 111 villes dans le monde⁸⁹ parmi elles 4 villes françaises (Lille, Montpellier, Nantes et Paris). Le nombre de villes participantes à l'International Open Data Day Hackathon 2014 est un indice de l'avancée et de l'expansion mondiale du mouvement Open data. Peu importe l'endroit, l'objectif était le même, selon l'OKF, qui a accueilli l'événement: « Dans de nombreuses villes à travers le monde, des citoyens se rassemblent pour développer des applications, libérer des données, créer des visualisations et publier des analyses utilisant des données ouvertes, tout cela pour supporter et encourager l'adoption de politiques de libération des données par les gouvernements locaux, régionaux et nationaux ». Tout le monde est invité à contribuer : développeurs, designers, bibliothécaires, statisticiens et citoyens. L'événement vise alors non seulement à créer des applications de données ouvertes, mais aussi de construire la communauté internationale d'utilisateurs de données ouvertes et étendre les applications existantes à d'autres endroits. Il s'agit de tirer parti des cas réussis, montrer les bonnes pratiques, et de créer des liens matériels et affectifs entre les communautés militantes et les communautés naissantes.

Open Government Data Camp⁹⁰ est un événement mondial annuel organisé par OKF depuis 2010. Un brainstorming collectif sur les données publiques en regroupant des fonctionnaires du service public, des développeurs, des organisations non gouvernementales (ONG) de 40 pays. L'objectif est de construire un consensus mondial autour des principes et des valeurs de données ouvertes et d'élargir et de renforcer la communauté internationale de données ouverte. Enfin, DataCatalogs.org⁹¹ est un autre projet lancé par

⁸⁷ Voir : Eaves, David. "Let's Do an International Open Data Hackathon." *Eaves.ca*, October 29, 2010. <http://eaves.ca/2010/10/29/lets-do-an-international-open-data-hackathon/>.

⁸⁸ Voir : Open Knowledge Foundation. "International Open Data Hackathon2014". Accessed 25 March 2014. <http://opendataday.org>.

⁸⁹ Voir : "2014/City Events", *Wiki International Open Data Hackathon* 2014. Accessed March 25, 2014. http://wiki.opendataday.org/2014/City_Events.

⁹⁰ Voir : Open Knowledge Foundation, 'Open Government Data Camp 2011', *Open Government Data Camp 2011*, 2011. <http://ogdcamp.org>.

⁹¹ "Datacatalogs.org". Accessed March 19, 2014. <http://datacatalogs.org/>.

l'OKF à Open Knowledge Conference organisé en 2011 à Berlin dans le but de recenser et créer un catalogue des catalogues de données ouvertes dans le monde. Il est organisé par un groupe d'éminents experts de données ouvertes du monde entier - y compris des représentants des gouvernements locaux, régionaux et nationaux, les organisations internationales telles que la Banque mondiale, et de nombreuses ONG.

L'association internationale OKF participe au mouvement open data par la promotion de la culture du libre. Pour synthétiser, nous pouvons dire que la stratégie d'OKF consiste à (1) préciser les fondements d'une donnée ouverte (proposer l'Open definition et des licences d'usage), (2) guider les administrations dans une stratégie open data (publier open data handbook), (3) proposer des outils de recherche de données ouvertes (Open data catalog), (4) proposer des outils de publication de données ouvertes (solutions techniques pour créer des plateformes de diffusion), (5) former à la manipulation des données (écoles des données), (6) créer un réseau et une communauté internationale d'open data (possibilité de créer un groupe de travail dans tous les pays), et enfin (5) animer l'open data (création d'événement, conférences, hackathons, expéditions et festivals) . Ces objectifs mettent en évidence la stratégie de l'OKF optant, selon nous, pour un positionnement de complémentarité avec les associations locales⁹². Nous soulignons également que malgré la qualité de ses travaux, son influence notamment auprès de l'État, est limitée à l'appui symbolique d'un projet open data.

4.1.2 La Commission européenne et l'Open Data.

A l'origine très orientées économiquement, les réflexions de l'Union européenne dans le domaine de la réutilisation des données publiques se sont développées depuis vingt ans. Actuellement, les institutions de l'Union européenne (UE) poursuivent une politique active en matière d'open data. En décembre 2011, Neelie Kroes, vice-présidente de la Commission européenne, a présenté la stratégie numérique en matière d'ouverture des données publiques⁹³, une stratégie de valorisation des données publiques européennes. Dans le cadre d'une approche économique, selon les

⁹² Nous présenterons dans la partie 4.2 les associations locales en France.

⁹³ Voir « Stratégie Numérique: Valoriser Les Données Des Administrations Publiques ». http://europa.eu/rapid/press-release_IP-11-1524_fr.htm?locale=en.

estimations, cette ouverture va soutenir l'économie de l'Union à hauteur de 40 milliards EUR par an. Les énormes volumes d'informations recueillies par les divers pouvoirs et services publics sont alors considérées comme une véritable mine d'or au potentiel économique inexploité. La stratégie de l'UE repose sur trois piliers à court terme : « premièrement, la Commission montrera l'exemple en donnant au public librement accès à ses informations par l'intermédiaire d'un nouveau portail; deuxièmement, il sera établi des conditions de concurrence équitables en matière de données ouvertes à travers l'UE; enfin, ces nouvelles mesures bénéficieront des 100 millions EUR qui, en 2011-2013, serviront à financer la recherche dans le domaine des technologies perfectionnées de gestion des données » (Commission Européenne, 2011). En Décembre 2012, le portail open data « open-data.europa.eu » de l'Union européenne contenant les données des services de la Commission et de l'Agence européenne pour l'environnement a été lancé. En Octobre 2013, le portail comptait 6138 jeux de données.

Bien avant, la commission européenne a publié « Le Livre vert sur l'information émanant du secteur public dans la société de l'information », un des premiers documents publiés en 1998 qui met en lumière l'importance de l'information du secteur public comme une ressource clef tout en soutenant la richesse économique et en montrant les opportunités qu'elle apporte pour la croissance économique. Ensuite plusieurs textes vont poser des principes sur l'accès aux informations publiques. La commission dans une communication du 23 octobre 2001 intitulée « e-Europe 2002 : créer un cadre communautaire pour l'exploitation de l'information émanant du secteur public »⁹⁴ relance le projet de la directive 2003/98/CE concernant la réutilisation des informations du secteur public tout en soutenant d'un point de vue économique les projets de valorisation économique des informations publiques. Ainsi la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 a posé le principe de la "réutilisation des informations du secteur public" au sein des États membres. La commission a également publié, en 2007, la directive européenne 2007/2/CE du 14 mars 2007 établissant une infrastructure d'information géographique dans la Communauté européenne (Inspire). Celle-ci vise à libérer les données géographiques et environnementales pour permettre l'échange, le partage,

⁹⁴ Commission Européenne. Direction générale de la société de l'information. "eEurope 2002 : Créer Un Cadre Communautaire Pour L'exploitation de L'information Émanant Du Secteur Public" Commission Européenne, 2002.

l'accès et l'utilisation de données interopérables et créer une banque de données géographique au niveau européen. En 2013, la commission européenne a publié la nouvelle version⁹⁵ de cette directive que nous présenterons de manière détaillée dans le chapitre suivant.

L'Union européenne reste très influente par rapport aux transpositions de lois qui ont introduit un nouveau système pour la réutilisation des données publiques. En France, comme dans les autres États membres, des centaines de textes législatifs et réglementaires ont été adoptés suite à la transposition en droit interne de la réglementation européenne. En effet, comme nous le verrons par la suite, la loi de 1978 sur l'accès aux documents administratifs a été modifiée en 2005 lors de la transposition de la directive européenne 2003/98/CE par l'ordonnance du 6 juin 2005⁹⁶ et mise en application par le décret n° 2005-1755 du 30 décembre 2005⁹⁷. Selon Cédric Favre⁹⁸, juriste et rédacteur du Guide pratique des données publiques, la transposition de la directive européenne dans le droit français a donné un nouvel élan en faveur des données publiques et de leur ouverture en France. La directive Inspire aussi transposée en 2010 en droit français sera, comme nous allons le voir, un des facteurs stimulant l'open data locale. L'influence de l'union européenne sur la réutilisation des données publiques montre d'être notable, surtout que la nouvelle version de la directive publiée en 2013 a été récemment (en décembre 2015) transposée en droit français en modifiant pour la deuxième fois la loi 78-753 du 17 juillet 1978 que nous présenterons de manière détaillée dans le chapitre suivant.

⁹⁵ Directive 2013/37/UE

⁹⁶ Ordonnance n°2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques.

⁹⁷ Les articles de cette loi qui régissent le droit d'accès et de réutilisation des documents publics seront examinés au chapitre 2, ce qui nous permet d'expliquer les pratiques de l'open data à la lumière de cette loi.

⁹⁸ Favre, C. (2011, juillet). Interview de spécialiste : Cédric Favre, juriste rédacteur du Guide pratique des données publiques. Retrieved from <http://www.data-publica.com/content/2011/07/interview-de-specialiste-cedric-favre-juriste-redacteur-du-guide-pratique-des-donnees-publiques/>

4.2 Open data en France : un modèle d'interaction entre les acteurs

Après avoir présenté les acteurs internationaux en mettant en perspective leurs initiatives politiques et sociales, nous souhaitons maintenant identifier les acteurs français impliqués dans l'open data et qui ont un impact, formel ou informel, direct ou indirect sur l'organisation et l'implémentation de l'ouverture des données publiques au niveau national⁹⁹. Si l'hétérogénéité des acteurs de l'open data ne peut se réduire à quelques catégories et illustre la complexité des dynamiques nationales à l'œuvre, quatre grands types peuvent néanmoins être dégagés :

- 1) les politiciens et législateurs
- 2) les agents publics
- 3) les think tanks
- 4) les associations

Cette typologie à visé évaluative nous aide à penser l'action publique en matière d'open data dans un contexte de redéfinition des politiques informationnelles de l'État et sous l'impulsion de repositionnement du rôle et de l'action des différents acteurs de la sphère publique ou sociale.

4.2.1 Politiciens et législateurs

Persuadé par la capacité de l'ouverture des données publiques à moderniser l'action publique, le Premier ministre, Jean-Marc Ayrault a réaffirmé, au festival du numérique 2013, « la détermination du Gouvernement à demander aux administrations et aux opérateurs publics de "jouer le jeu" »¹⁰⁰. Bien avant, cette priorité avait été inscrite dans la Charte de déontologie du 17 mai 2012 signée par tous les membres du gouvernement dès le premier Conseil des ministres de la mandature. Elle se traduit

⁹⁹ La partie 3 de ce manuscrit sera dédiée à une réflexion plus précise sur les acteurs au niveau local.

¹⁰⁰ Voir : "Le Premier ministre réaffirme son attachement à l'ouverture et à la mise à disposition des données publiques". *etalab*, 2013. <http://www.etalab.gouv.fr/article-le-premier-ministre-reaffirme-son-attachement-a-l-ouverture-et-a-la-mise-a-disposition-des-donnees-p-118492716.html>.

également par onze décisions prises lors des trois premiers Comités interministériels pour la modernisation de l'action publique (CIMAP), présidés par le Premier ministre le 18 décembre 2012, le 2 avril et le 17 juillet 2013. Enfin, une ambitieuse feuille de route stratégique a été adoptée lors du séminaire gouvernemental sur le numérique du 28 février 2013¹⁰¹. C'est également, comme nous l'avons vu, un engagement réclamé et souscrit par la France avec l'adoption, le 18 juin 2013, par les chefs d'Etat et de gouvernement du G8, de la Charte du G8 pour l'ouverture des données publiques.

Après l'an 2010, nous distinguons une accélération du rythme de publication de décrets et circulaires comme acte exécutoire et qui permet aux autorités administratives (ministre, recteur, préfet...) d'informer leurs services sur les applications en relation avec la réutilisation des données publiques :

- La circulaire du 14 février 1994 précisant les conditions de « diffusion des données publiques ».
- Le décret n° 2010-31 du 11 janvier 2010 relatif à la création de la Direction de l'information légale et administrative (DILA)¹⁰² qui assure des responsabilités de publication et de diffusion de documents administratifs et juridiques.
- Le décret n° 2010-32 du 11 janvier 2010 instituant un conseil d'orientation de l'édition publique et de l'information administrative (COEPIA) qui exerce une fonction d'évaluation, d'expertise et de conseil sur la mise à disposition des données publiques.
- Le décret n°2011-194 du 21 février 2011 qui a créé la mission « Etalab » chargée de la création d'un portail unique interministériel des données publiques.
- La circulaire du 26 mai 2011 a permis en pratique la création du portail unique des données publiques de l'Etat « data.gouv.fr » par la mission Etalab et l'application de nouvelles conditions de réutilisation provoquées par cette mission et, en particulier, la

¹⁰¹ “La feuille de route du Gouvernement en matière d’ouverture et de partage des données publiques.” *OverBlog*, February 28, 2013. <http://www.etalab.gouv.fr/article-la-feuille-de-route-du-gouvernement-en-matiere-d-ouverture-et-de-partage-des-donnees-publiques-115767801.html>.

¹⁰² Nous revenons sur la DILA dans le chapitre suivant en présentant les acteurs traditionnels de diffusion des données publiques.

généralisation de la « réutilisation libre, facile et gratuite des informations publiques » en tenant en compte le système de redevances existantes.

- Le décret n° 2012-1198 du 30 octobre 2012 portant création du secrétariat général pour la modernisation de l'action publique. Etalab est devenu un service directement rattaché au secrétaire général chargé de faciliter et de coordonner la mise à disposition des données publiques en vue de développer leur réutilisation.

Le législateur est intervenu à plusieurs reprises depuis la loi n° 78-753 du 17 juillet 1978, la loi initiale portant sur l'accès aux documents administratifs, sur des questions autour des informations publiques. Les résultats de l'étude de Balcou, Cottin et Iteanu (2011), ont montré l'importance de l'inflation législative et réglementaire autour du droit de l'information entre 1980 et 2010. Le nombre de lois publiées paraît dans les années 2000 assez stable par rapport au recours intensif aux ordonnances qui au contraire ne cesse de croître : 22 articles en moyenne en 1990, 33 en 2000, 41 en 2009. (Voir figure 8)

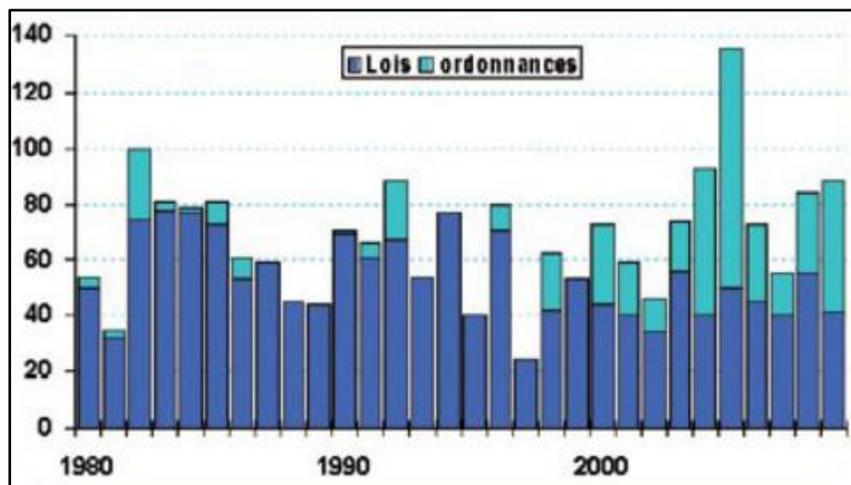


Figure 7. Évolution du nombre de textes de niveau législatif (Balcou, Cottin et Iteanu, 2011)

Nous citons :

- La loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal. Cette loi

a créé la Commission d'Accès aux Documents Administratifs (Cada)¹⁰³.

- L'ordonnance n°2005-650 du 6 juin 2005 modifiant la loi du 17 juillet 1978 et qui a transposé en droit français la directive européenne 2003/98/CE (modifiée en 2013) concernant la réutilisation des informations du secteur public
- L'ordonnance du 22 octobre 2010 qui a transposé en droit français la directive européenne 2007/2/CE du 14 mars 2007 « Inspire » sur la diffusion des données géographiques et environnementales.

Enfin, pour finir ce bref panorama des textes législatifs, nous nous devons de citer la loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public. Ainsi, le législateur français a procédé à la transposition de la directive 2013/37/UE du 26 juin 2013 concernant la réutilisation des informations du secteur public en modifiant pour la deuxième fois la loi 78-753 du 17 juillet 1978. Cette nouvelle loi érige en principe la gratuité de la réutilisation des données publiques qui a suscité un vaste débat depuis le début du mouvement open data et sur lequel nous reviendrons dans le chapitre suivant.

Force est de constater que cinq ans après la première initiative open data en France¹⁰⁴, une réelle dynamique politique et juridique s'est instaurée sur le sujet. A ce stade, les politiciens et législateurs jouissent d'un pouvoir de décision important concernant l'obligation des institutions publiques à adopter des politiques en faveur de l'open data. Malgré l'ambition du gouvernement (représenté par le secrétaire d'État au Numérique et ministre de la Réforme de l'État) de modifier la loi de 1978 au travers du grand projet de loi numérique¹⁰⁵, le retard législatif empêche d'avancer davantage encore en matière d'Open Data. Nous présentons dans le chapitre suivant les limites de cette loi par rapport aux applications de l'open data.

¹⁰³ Les agents publics créés par les textes législatifs (comme Cada, COEPIA, Etalab) seront présentés ci-après afin de mieux cerner leurs rôles et leur impact sur l'implémentation de l'open data en France.

¹⁰⁴ Ouverture des données de Rennes Métropole en 2010.

¹⁰⁵ Voir le texte intégral du projet de loi sur :

https://www.legifrance.gouv.fr/affichLoiPreparation.do;jsessionid=71AB31DC7004C58B6AF64347F7C13C11.tpdila23v_3?idDocument=JORFDOLE000031589829&type=contenu&id=2&typeLoi=proj&legislature=14

4.2.2 Agents publics

Nous présentons ici les divers agents publics concernés par l'open data mais que nous répartissons entre acteurs producteurs ou détenteurs de données et acteurs facilitateurs de l'ouverture mandatés par les politiciens et législateurs. La plateforme ouverte des données publiques françaises data.gouv.fr fournit une liste¹⁰⁶ pour une identification rapide des acteurs producteurs qui ont réagi positivement aux appels d'ouverture de leurs données. Les producteurs de données se constituent essentiellement d'acteurs publics de premier rang (ministère, établissement public interministériel ...) et des collectivités territoriales mais aussi d'associations et de sociétés privées, une stratégie adoptée par Etalab pour enrichir et diversifier leur offre de données¹⁰⁷. De statut différent et de niveau de compétences distinct, les acteurs facilitateurs offrent quant à eux un support technique ou occupent le rôle de régulateurs et de consultants chargés de l'organisation de l'accès et la réutilisation des données publiques. Dans cette catégorie nous identifions : Etalab, Cada, CNIL, COEPIA. Nous présenterons ici les caractéristiques de chacun, les formes de leur intervention et le degré de leur implication dans le processus d'ouverture des données publiques.

La mission « Etalab » est chargée de la création d'un portail unique interministériel des données publiques. Créée par le décret n°2011-194 du 21 février 2011, Etalab a coordonné et accompagné les administrations ministérielles dans leur démarche d'ouverture et le développement du portail national data.gouv.fr lancé le 5 décembre 2011, La mission Etalab, rattachée auprès du Premier ministre, est devenue en octobre 2012 un service directement rattachés au Secrétaire Général pour la Modernisation de l'Action Publique (SGMAP). Les services proposés par l'Etat aux organismes publics au niveau national et local afin de les aider à ouvrir leurs données se concentrent autour deux types de services :

- La création d'une plateforme web nationale destinée aux institutions publiques qui recherchent un outil technique centralisé pour diffuser leurs données et aux utilisateurs qui recherchent une base de données pour exploiter les données.

¹⁰⁶ Voir la liste complète sur <https://www.data.gouv.fr/fr/organizations/?page=2>

¹⁰⁷ Nous ne souhaitons pas dans cette recherche aller plus avant dans l'analyse des stratégies adoptées par Etalab en termes de l'hétérogénéité des organisations contributeurs aux données sur la plateforme nationale.

- L'organisation de concours open data « Dataconnexions » lancé chaque année depuis 2012 afin de dynamiser le secteur de l'économie numérique et recomposer des projets open data. Les concours permettent, trois fois par an, aux porteurs de projets innovants utilisant des données ouvertes par l'état ou les collectivités locales de concourir, de présenter leurs projets à un jury et de se faire connaître d'un réseau d'acteurs.

La Commission d'Accès aux Documents Administratifs (Cada) a été créée par la loi *De la liberté d'accès aux documents administratifs*, n°78-753 du 17 juillet 1978 pour assurer la bonne application du droit d'accès aux documents administratifs. Elle joue le rôle de consultant pour les réutilisateurs comme pour les administrations qui la sollicitent pour donner avis et conseils sur le caractère communicable ou non d'un document. En tant qu'autorité administrative indépendante, la Cada entend ainsi confirmer sa mission de « veille à la transparence de l'action administrative et fait connaître son interprétation des textes applicables ». Parmi ces compétences, « elle peut proposer au gouvernement les modifications nécessaires pour améliorer l'exercice du droit d'accès et, en matière de réutilisation des informations publiques, elle peut également prononcer des sanctions à l'encontre des personnes qui réutilisent des informations publiques en violation des prescriptions de la loi »¹⁰⁸. Pour assurer la communication entre les administrations et la Cada, la loi de 1978 (Art. 24), modifiée par l'ordonnance de 2005 oblige les administrations y compris les collectivités¹⁰⁹ à désigner une personne responsable de l'accès aux documents administratifs et des questions relatives à la réutilisation des informations publiques (PRADA ou PRADAQRIP). Nous verrons notamment l'importance du PRADA lorsque nous traitons la question de collecte de données publiques à ouvrir dans une collectivité.

¹⁰⁸ Voir : « Le Rôle de La Cada ». Accessed July 7, 2014. <http://www.cada.fr/le-role-de-la-cada,6077.html>.

¹⁰⁹ Selon l'article 42 du décret n° 2005-1755 du 30 décembre 2005, les administrations qui sont soumises à l'obligation sont : « les ministres et les préfets pour les services placés sous leur autorité ; les communes de dix mille habitants ou plus, les départements, les régions et la collectivité territoriale de Corse ; les établissements publics nationaux et locaux qui emploient au moins deux cents agents ; les établissements publics de coopération intercommunale regroupant une population de dix mille habitants ou plus ; les autres personnes de droit public et les personnes de droit privé chargées de la gestion d'un service public qui emploient au moins deux cents agents ».

La Commission nationale de l'informatique et des libertés (CNIL), quant à elle, veille à la protection des données personnelles. Créée par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, elle dispose d'un pouvoir de contrôle et de sanction à ce que « le développement des nouvelles technologies ne porte atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques »¹¹⁰. Quand il s'agit de l'ouverture et la diffusion des données publiques, la CNIL joue un rôle de contrôle et de conseil sur l'anonymisation des données ouvertes pour ne pas révéler des données personnelles. La désignation d'un Correspondant informatique et libertés (CIL) au sein d'une entreprise, d'une administration ou d'une collectivité locale assure la protection des données personnelles. La CNIL poursuit ses réflexions sur les contraintes de l'open data. Afin d'avoir une meilleure perception de ce mouvement et de ses développements à moyen terme, au regard de ses éventuelles incidences sur la protection des données personnelles, la CNIL a organisé le 9 juillet 2013 une journée d'échanges et de retours d'expérience¹¹¹. Le but était de sensibiliser les acteurs sur les enjeux de l'open data sur la protection des données personnelles. En outre elle a lancé une consultation en ligne auprès des acteurs de l'open data. Le but était de dresser un état des lieux des pratiques et des questions « informatique et libertés » qui se posent dans le cadre des projets d'ouverture de données publiques.

Dans une approche orientée vers les utilisateurs des données publiques, le conseil d'orientation de l'édition publique et de l'information administrative (COEPIA) joue le rôle de garant et veille à la bonne allocation des moyens et à la qualité du service rendu à l'utilisateur. Cet organisme créé par le décret n° 2010-32 du 11 janvier 2010, exerce une fonction d'évaluation, d'expertise et de conseil sur la mise à disposition des données publiques sur l'édition publique et la mise à disposition des données publiques. Le Coepia était chargée, à travers une formation spécialisée dédiée, de rendre un avis au Premier Ministre sur les informations publiques soumises à redevance. En

¹¹⁰ Commission nationale de l'informatique et des libertés. "Missions". *CNIL*. Accessed July 7, 2014. <http://www.cnil.fr/linstitution/>.

¹¹¹ Voir : "Open Data, Quels Enjeux Pour La Protection Des Données Personnelles?". 9 juillet 2013. <http://www.cnil.fr/les-themes/collectivites-locales/article/article/open-dataet-donnees-personnelles-lancement-dune-consultation-des-acteurs/>.

plus, Le conseil accueille plusieurs acteurs pour participer aux travaux du Conseil au sein de la formation spécialisée sur la "Mise à disposition et valorisation des données publiques".

L'Open data entre également dans le champ des activités de l'Agence du patrimoine de l'État (APIE) chargée de la valorisation du patrimoine immatériel de l'Etat. Le champ du patrimoine immatériel, dont il n'existe pas de définition juridique, est très large et recouvre selon l'APIE, différents actifs : données, images, marques, savoir-faire, logiciels et brevets, lieux prestigieux, etc¹¹². L'agence créée en 2007, propose une stratégie de gestion des actifs immatériels de l'Etat, en vue d'assurer une meilleure valorisation de ce patrimoine. En terme d'open data au niveau national, l'APIE a été chargée avec Étalab et COEPIA de travailler ensemble, sous la direction d'Étalab, sur la question d'une licence gratuite de ré-utilisation des données publiques durant l'été 2011. L'APIE a publié en juin 2011 une nouvelle version de sa licence de *réutilisation libre et gratuite*. Regard citoyens.org (2011) signale un problème de communication et de coordination des actions et entre ces acteurs : « L'APIE publie en Juin 2012 la version V2 de sa licence de réutilisation libre et gratuite sans prendre en compte les réalités et les tendances au niveau des collectivités des licences choisies¹¹³, et les licences déjà existantes ». La publication de la licence de l'APIE semble révéler un problème d'organisation de la communication autour de la future licence à adopter au niveau national. En octobre 2011, Etalab a finalement publié la Licence Ouverte adoptée aussi par plusieurs collectivités territoriales.

Ces divers agents publics jouent le rôle de consultants pour les administrations et les usagers, de garants de la bonne application de la loi et de la protection des données personnelles et enfin de promoteurs du patrimoine immatériel de l'Etat. A l'exception d'Étalab, leur engagement dans l'organisation et la mise en œuvre de l'open data au niveau national reste marginal. Les acteurs qui ne sont pas obligatoirement impliqués dans l'action d'ouverture, sont nombreux et ils ont gardé leurs places au schéma de l'open data vu l'interdépendance observable.

¹¹² Voir "Agence Du Patrimoine Immatériel de L'état - APIE," August 3, 2013. <http://www.economie.gouv.fr/apie>.

¹¹³ Alors que Nantes et la Gironde viennent de rejoindre l'ouverture en adoptant comme Paris une des licences libres existantes (l'ODbL),

De même des actions s'échappent de la politique nationale et aux agents publics comme la publication de guide officiel pour accompagner les administrations dans la mise en œuvre des projets open data et standardiser les pratiques et les formations pour les chefs open data.

4.2.3 Think tanks

Les think tanks se présentent comme des « groupes de réflexion et d'influence » (Huyghe, 2008). Au niveau national, pour ce qui concerne la thématique de l'open data, les deux think tanks les mieux actifs et connus parmi les acteurs sont le Groupement Français des Industries de l'Information (GFII) et la Fondation Internet Nouvelle Génération (Fing).

Le GFII¹¹⁴ est un groupement des acteurs du marché de l'information des secteurs public et privé. Son engagement pour l'open data est associé à la création, depuis 1997, d'un groupe de réflexion consacré à la diffusion des informations publiques. La présentation que se donne le groupement de leur think tank met l'accent sur leurs activités de réflexion, information, promotion, réseautage, représentation et formation¹¹⁵. Par la création des groupes de travail d'experts disposant de capacités d'analyse sur des sujets stratégiques pour le secteur, le GFII vise à créer un réseau d'acteurs informé et formé. Les différents événements ou réalisations permettent également de valoriser l'expertise de ses membres. Par la représentation des différents membres, le GFII est perçu comme un groupe d'influence pour faire valoir et défendre le point de vue des professionnels de l'information électronique auprès des organes de concertation, des instances françaises, européennes et internationale. La genèse du groupe de réflexion consacré à la diffusion des informations publiques correspondait aux premières réflexions de la commission européenne qui ont abouti à la publication de la directive de 2003 sur la réutilisation des informations du secteur public. En 2008, lors de la 11^{ème} réunion du groupe européen sur la réutilisation des informations publiques (PSI re-use), Denis Berthault¹¹⁶, vice président de GFII, affirme

¹¹⁴ Voir : "GFII - Les Acteurs Du Marché de L'information et de La Connaissance". Accessed July 4, 2014. <http://www.gfii.fr/fr/>.

¹¹⁵ D'après la présentation du GFII sur son site web <http://www.gfii.fr/fr/legfii/presentation-du-gfii>

¹¹⁶ Voir : Berthault, Denis. PSI re-use in France : position of the industry. 11th PSI Group meeting, Tuesday, 6 May 2008, Luxembourg.

www.epsiplatform.eu/sites/default/files/ezpublish_media/berthault.pps

que l'influence ou l'effet de lobbying du GFII d'engendrer une politique publique pertinente au niveau de la réutilisation des données publiques étaient perçue comme non pertinente malgré la mobilisation du groupement¹¹⁷. Avec l'émergence mondiale du mouvement open data, le GFII s'est mobilisé en suivant presque le même scénario¹¹⁸. Le GFII forme certes des groupes de réflexion et des réseaux d'acteurs sur leurs domaines d'expertises mais son influence dans le champ politique est pour le moment difficilement identifiable.

L'association Fing¹¹⁹, un think tank sur les technologies numériques, qui depuis 2010, engage une campagne sur l'ouverture des données publiques en partenariat avec plusieurs territoires. Ses objectifs sont de découvrir les initiatives open data, d'approfondir la compréhension des ses enjeux et d'accompagner l'ouverture et la réutilisation des données publiques locales¹²⁰ pour stimuler l'innovation de proximité. Fing a publié en 2011 le « Guide pratique de l'ouverture des données publiques territoriales »¹²¹ qui synthétise les enseignements tirés des expérimentations territoriales. La Fing a ainsi mené des études sur les enjeux de l'open data, établi des règles d'application et même offert des solutions techniques pour la mise en œuvre de l'ouverture. L'attitude de la Fing dont l'objectif est la stimulation de l'open data, se rapproche plus de la promotion que de l'influence à proprement parler. Les deux think tanks, le GFII et la Fing, essayent de jouer de leur influence mais ne réussissent pas en tout état de cause à se comporter comme un lobby.

4.2.4 Associations

¹¹⁷ Les actions menées comprennent des réunions de réflexions, des publications, des conférences, des lettres adressées au premier ministre, des propositions au ministère de l'industrie, des réunions avec la CADA et l'APIE.

¹¹⁸ Le GFII a publié en 2010 un communiqué de presse sur l'open data. Le groupe de travail intervient au COEPIA, plus particulièrement au sein de la Formation Spécialisée (FS2) consacrée à la réutilisation des données publiques (aspects économiques, fiscaux et juridiques) ; il entretient des échanges avec Etalab et suit également les actions du réseau européen ePSI platform.

¹¹⁹ "Fondation Internet Nouvelle Génération". Accessed July 4, 2014. <http://fing.org/?lang=fr>.

¹²⁰ Plusieurs exemples sont illustrés dans la dernière partie consacrée aux projets open data dans les collectivités locales.

¹²¹ Fondation Internet Nouvelle Génération, "Guide Pratique de L'ouverture Des Données Publiques Territoriales," January 1, 2011, <http://fing.org/?Guide-pratique-de-l-ouverture-deslang=fr>.

Deux nouveaux acteurs, LiberTic et Regards Citoyens, sont activement impliqués dans le débat et les discussions sur l'open data en France. L'association LiberTic est l'un des premiers militants Open Data en France. Créée en décembre 2009 suite à l'Opendata barcamp de la Cantine à Paris, elle est devenue un acteur de référence sur la thématique à l'échelle locale voire nationale. Elle est également un des membres fondateurs du Collectif Open Data France que nous allons présenter parmi les acteurs locaux de l'open data par ce qu'il s'adresse principalement aux collectivités territoriales. Pour comprendre le contexte de sa création, Claire Gallon, cofondatrice de LiberTic, explique : « À l'époque de la création de LiberTic, les cofondateurs faisaient de la veille numérique dans une association chargée d'évaluer les impacts des produits de consommation sur l'environnement. C'est là que nous avons pris conscience de l'importance de pouvoir accéder librement aux données et que nous avons découvert l'émergence du mouvement open data dans les pays anglo-saxons »¹²². De par sa mission de promouvoir l'ouverture des données publiques, le gouvernement 2.0 et la démocratie électronique, les membres du collectif ont participé aux forums sur l'open data et publiée sur le sujet sur leur blog. En suivant l'actualité de l'Open data en France, le blog est devenu une source pour faire la veille sur le sujet et suivre les initiatives des collectivités. Cependant, l'influence ou l'apport de LiberTic sur l'implémentation ou l'organisation de l'open data au niveau national est difficilement identifiable par rapport au niveau local plus visible et étudiés dans la dernière partie. A la différence de Libertic, la deuxième association Regards Citoyens milite afin de promouvoir la réutilisation des données publiques en se transformant elle-même en réutilisateur par le développement de projets comme Nosdéputés.fr. Ce projet permet au collectif développer des applications et des services informationnels sur les députés à partir des données publiques. Nous signalons que les communautés de « Civic hackers » aux Etats-Unis et en Grande Bretagne étaient des acteurs de pression car ils ont incité les administrations à l'ouverture. Becky Hogge (2010) dans son étude « Open Data Study » constate qu'aux États-Unis et au Royaume-Uni, la société civile, et en particulier un petit groupe motivé de « civic hackers » était essentiels à la réussite des projets d'ouverture des données. Les civic

¹²² Voir : « L'open Data Est Très Mal Estimé » Interview de Claire Gallon de L'association Libertic". *Data Publica*, January 2, 2014. <http://www.data-publica.com/content/2012/11/lopen-data-est-tres-mal-estime-interview-de-claire-gallon-de-lassociation-libertic/>.

hackers ont dans un premier temps participé à collecter des données non-ouvertes. Ils ont agi parfois hors la loi dans le but d'extraire et d'obtenir des données publiques, d'où l'appellation « Hacker ». Or la communauté de Civic hackers en France est presque invisible dans la phase de pré-ouverture.

Conclusion du chapitre 4

Dans ce chapitre nous avons mis en exergue, à la suite d'une description de plusieurs initiatives, l'importance de l'articulation de la dynamique politique et sociale en faveur de l'open data. Nous nous sommes interrogés sur les éléments relatifs à l'expansion du mouvement open data et l'influence des acteurs nationaux et internationaux sur l'implémentation de l'open data au niveau national. Nous avons présenté les enjeux de l'open data aux niveaux mondial et national. Dans cette optique, les premières initiatives ont eu un effet de catalyseur sur les politiques de plusieurs pays et la création d'une communauté internationale de l'open data. La Charte du G8 pour l'Ouverture des Données Publiques signée par les Chefs d'Etat du G8 est un bon exemple de la vision mondiale et l'ambition des États de promouvoir l'open data. La commission européenne a elle-même entrepris de stimuler l'ouverture des données publiques. La mise en place d'une politique d'ouverture repose sur une politique économique ciblée qui précise le cadre dans lequel les réutilisateurs sont invités à développer des richesses économiques en considérant l'information du secteur public comme une ressource clef. Avec la directive 2003/98/CE révisée en 2013, le principe de la "réutilisation des informations du secteur public" a été posé au sein des États membres

Dans le tableau 5, nous proposons de mettre en avant une typologie des principaux acteurs qui ont contribué d'une manière ou d'une autre à l'implémentation de l'open data en France ainsi que leurs différents statuts.

.

Tableau 5. Typologie des acteurs dans l'implémentation de l'open data

Type d'acteurs	Internationaux ou nationaux	Exemples	Apports
Catalyseur	Acteurs internationaux	Les États-Unis ; les pays du G8 ; Open Government Partnership (OGP)	1. Catalyseur open data à cause de l'effet d'exemplarité. 2. Moyen d'influencer les perceptions des politiques 3. Groupe de pression sur les pays membres
	Acteurs européens	Commission européenne	4. Catalyseur open data à cause de l'obligation de la transposition en droit interne la réglementation européenne
Politiciens et législateurs		Premier ministre ; Conseil des ministres ; Comités interministériels pour la modernisation de l'action publique (CIMAP) ; Les élus	5. Détenteurs du pouvoir législatif et exécutif. Ils jouissent d'un pouvoir de décision important concernant l'obligation des administrations publiques à adopter des politiques en faveur de l'open data.
Producteurs	Agents publics	Ministères ; Établissement public interministériel ; Collectivités territoriales	6. Producteurs et détenteurs des données publiques
Facilitateurs	Agents publics	Etalab, Cada, CNIL, COEPIA	7. Facilitateurs de l'ouverture mandatés par les politiciens et législateurs
Militants	Associations internationales	OKF	8. Promoteur de la connaissance ouverte

	Think tanks nationaux	GFII, Fing	9. Promoteur par des activités de réflexion, information, promotion, réseautage, représentation et formation
	Associations nationales	LiberTic,	10. Acteur de référence sur la thématique à l'échelle locale voire nationale et membre fondateurs du Collectif Open Data France
Militants et réutilisateurs	Associations nationales	Regards Citoyens	11. Promoteur et réutilisateur par le développement de projets comme Nosdéputés.fr.

La figure 8 vient répondre à la question de pouvoir et d'influence des différents acteurs nationaux et internationaux sur l'implémentation de l'open data au niveau national.

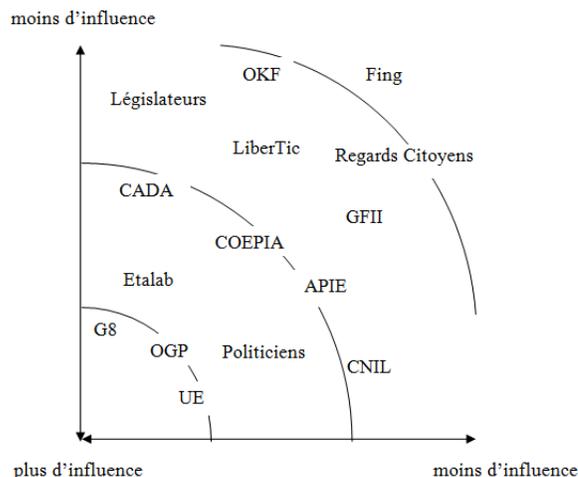


Figure 8. Influence des acteurs open data en France (Rahmé, 2015)

Nous pouvons tirer plusieurs conclusions du schéma. Premièrement, la politique nationale se traduit essentiellement par la création de la mission « Etalab » et une volonté politique fortement prédéfinie par le contexte international et européen. Les agents publiques de régulation joue un rôle plus marginal par rapport à l'implémentation de l'open data en France. Les législateurs ont moins d'influence vu les limites de la loi par rapport aux applications de l'open data. Enfin, nous signalons l'apport des think tanks dans le développement d'une dynamique sociale mais qui ne réussissent pas en tout état de cause de se comporter comme un lobby.

Chapitre 5. La communication et la diffusion des données publiques

Dans ce chapitre, nous nous intéressons aux pratiques concernant la communication et la diffusion des données publiques en France. Nous présenterons l'open data comme une extension des ce pratiques exercées par les institutions publiques et qui ne représentent pas un phénomène nouveau.

L'open data marque une évolution dans les pratiques, d'une part, des détenteurs des données et d'autre part, des usagers en face de nouveaux droits qui leur ont été accordés. Afin de comprendre ces transformations et les nouvelles pratiques qu'apporte l'Open data, nous étudions l'évolution des modes et des conditions de communication et de diffusion de l'information publique au sein des institutions publiques. Nous présentons l'open data qui compose ce que nous appelons le « nouveau système » de gestion et de diffusion des données publiques comme une extension du système et des pratiques traditionnelles des acteurs de la communication et de diffusion des informations publiques. Au-delà d'un simple accès aux documents publics, il s'agit désormais de s'interroger sur de nouveaux principes qui identifient ce qu'est une donnée ouverte diffusée dans des conditions différentes du système traditionnel de communication et de diffusion des informations publiques déjà existantes.

Nous plaçons les pratiques de l'open data en France dans un contexte historique et juridique qui date depuis 1978. Ce chapitre se base essentiellement sur une exploration des quelques sources disponibles sur le sujet comme les textes de loi de la base de donnée Legifrance et les différents rapports administratifs. Il convient tout d'abord de distinguer les notions documents administratifs, d'informations publiques et de données publiques considérées dans notre recherche. Ensuite, nous étudions les pratiques de communication et de diffusion des informations publiques à la lumière de la loi sur l'accès aux documents administratifs du 17 juillet 1978 ainsi que d'autres directives et circulaires qui interviennent directement pour régler ce système. Ainsi, nous étudions le système de communication et de diffusion des informations publiques en démontrant deux évolutions

essentielles : l'évolution des modes de communication qui met en évidence les distinctions entre la notion « d'accès » et la notion de « diffusion » en signalant les logiques d'offre et de demande d'informations publiques. Nous nous arrêtons également sur l'effet de la « société de l'information » sur l'évolution des modes de communication. Puis, l'évolution des conditions d'usage pour questionner la position de l'utilisateur, ses droits et ses pratiques d'usages dans un nouveau contexte d'utilisation.

5.1 Documents administratifs, informations publiques et données publiques

La loi du 17 juillet 1978 sur le droit d'accès à tout document administratif donne une définition du document administratif : « Sont considérés comme documents administratifs, quels que soient leur date, leur lieu de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission. Constituent de tels documents notamment les dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, directives, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions et décisions¹²³. Les documents qui ne sont pas considérés comme « documents administratifs » ne sont pas communicables. Par exemple, les documents législatifs relevant du pouvoir judiciaire sont exclus de son champ d'application¹²⁴. Il s'agit des « documents des juridictions, y compris les juridictions administratives et financières, qui sont liés à la fonction de juger ; les documents à caractère judiciaire ; les documents d'état civil ; les documents privés ; et les documents des assemblées parlementaires. Ce qui nous permet de souligner que la notion de « document administratif » désigne en même temps la nature et le contenu du document, son producteur et son statut communicable.

L'ordonnance de 2005 a modifié la loi de 1978, en incluant la notion d' « informations publiques ». Selon la loi, les informations publiques sont « les informations figurant dans des documents produits ou reçus par les administrations mentionnées à l'article 1^{er} ». Cette définition semble donc une reprise de la définition du document de Le Coadic

¹²³ Voir : “Loi N° 78-753 Du 17 Juillet 1978 Portant Diverses Mesures D'amélioration Des Relations Entre L'administration et Le Public et Diverses Dispositions D'ordre Administratif, Social et Fiscal | Legifrance” (legifrance.gouv.fr, 1978), <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006068643dateTexte=vig>.

¹²⁴ Voir : “La Notion de Document Administratif,” Cada, accessed May 24, 2014, <http://www.cada.fr/la-notion-de-document-administratif,56.html>.

(1994) : « Document » est le terme générique désignant les objets porteurs d'information » et de laquelle nous pouvons constater que pour favoriser sa communication, les informations doivent se matérialiser dans une médiation que nous qualifierons par « document administratif ». Ainsi, nous ne pouvons pas penser les informations publiques sans le document qui définit sa nature et son statut communicable.

La notion de « donnée publique » apparaît seulement dans la circulaire du 14 février 1994¹²⁵ relative à la diffusion des données publiques. Celle-ci donne deux définitions à la notion de donnée publique : au sens large, elle est équivalente à une information publique et produite sur n'importe quel support ; au sens étroit, elle désigne une information formatée pour être traitée par un système informatique. Nous pouvons donc dire que les données publiques, telle qu'elles sont décrites par les textes législatifs, ne se distinguent pas des informations publiques et elles représentent plus spécifiquement les données numériques. Cette définition souffre, selon nous, du peu de prise en compte des distinctions entre les notions de données et d'informations discutées dans la section 1.1, et de la propriété des documents numériques assez structurés destinés dans la logique d'open data et qui ne portent pas nécessairement de sens donc d'informations. Le développement des systèmes de gestion informatique dans les administrations, ainsi que la numérisation ont remis en cause l'utilisation de mot « document » et même de « l'information ». Le mode de gestion publique dans l'administration électronique défini par l'utilisation des nouvelles technologies d'information et de communication a produit des données numériques en dehors du cadre traditionnel des documents administratifs. D'où les exceptions pour considérer, par exemple, les courriels électroniques émis par les administrations, ou également les données contenues dans des bases de données qui peuvent en être extraites par un traitement automatisé d'usage courant, comme des documents administratifs communicables. En effet, les avis et les conseils émis par la Commission d'accès aux documents administratifs (Cada)

¹²⁵ Voir : *Circulaire Du 14 Février 1994 Relative À La Diffusion Des Données Publiques*, 1994 <<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000181007>>

constituent autant d'exemples de ce qui est qualifié en tant que « document administratif ». Par exemple, l'avis, référencé 20020741 du 14 mars 2002, avait précisé que «les courriers électroniques doivent être regardés comme des documents administratifs au sens de l'article 1er de la loi du 17 juillet 1978 modifiée»¹²⁶. Dans un avis référencé 20140831¹²⁷ du 27/03/2014, à propos des « informations relatives à la consommation de Dompéridone chez les femmes allaitant et chez les nourrissons depuis 2006 jusqu'à nos jours » produites par la caisse nationale d'assurance maladie des travailleurs salariés », la Cada avait indiqué que « sont regardés comme des documents administratifs existants, les informations qui sont contenues dans la base de données nationale SNIIRAM (système national d'information inter-régimes de l'assurance maladie).

En distinguant les particularités des données publiques par rapport aux « documents administratifs » et « informations publiques », nous pouvons constater une caractéristique conceptuelle et une autre technique de la donnée publique. La caractéristique conceptuelle représente la donnée publique comme identique à l'information publique et au document administratif, voir un élément attaché à ce dernier qui définit sa propriété et sa couverture (administratif ou pas) qui, par la suite, détermine sa communicabilité. Le caractère technique présente les données strictement dans leur format numérique pour désigner en d'autre terme les données numériques. Le flou autour de la notion de donnée publique explique pourquoi ces notions sont utilisées indifféremment dans les textes de loi et les rapports officiels. Comme le précise Salaün (2004), le numérique bouleverse profondément l'appréhension du document et nécessite alors de le réinterroger à l'aune des usages des données numériques. La distinction traditionnelle faite entre document, information et données tend à s'estomper à cause de la volatilité du support

¹²⁶ Voir : “Directeur Général de L'agence Française de Sécurité Sanitaire Des Aliments (AFSSA) : Conseil 20020741 Séance Du 14/03/2002,” CADA, March 14, 2002, <http://www.cada.fr/conseil-20020741,20020741.html>.

¹²⁷ Voir : “Caisse Nationale D'assurance Maladie Des Travailleurs Salariés (CNAMTS de Paris) : Avis 20140831 Séance Du 27/03/2014,” CADA, March 27, 2014, <http://www.cada.fr/avis-20140831,20140831.html>.

numérique. Une base de données, une extraction de ces données suite une requête ou la diffusion d'un flux de données dynamiques représentent-ils des documents ? Certainement, la dissociation des données du document affecte, dans le contexte de l'administration publique, la stabilité traditionnelle d'un document défini par son statut et sa fonction principale de preuve.

5.2 Système et acteurs traditionnels de la communication de données publiques

Les pratiques de communication et de réutilisation des informations publiques existaient avant l'open data. Le système traditionnel de communication et de diffusion de données et d'informations publiques est fondé sur deux modes d'accès :

- I. Principe d'accès sur demande (depuis 1978)
- II. Principe de diffusion par les « éditeurs publics » (depuis 1998)

Le premier est le principe d'accès sur demande. Il est régi essentiellement par la loi de juillet 1978 qui permet à tout citoyen de pouvoir consulter ou se faire communiquer un document administratif lorsqu'il en fait la demande. Ce principe est général par ce qu'il s'étend à tous les organismes publics et s'applique à toutes les informations publiques, sauf celles indiquées par cette loi.

Le deuxième mode d'accès se base sur le principe de diffusion d'informations publiques par les institutions publiques elles-mêmes. Il est assuré traditionnellement, par les « éditeurs publics institutionnels ». Ce principe est considéré selon nous comme partiel par ce qu'il concerne un nombre limité d'institutions publiques qui ont la vocation d'exercer une activité éditoriale et se réduit à leurs offres sélectives d'informations sous formes de publications. Ainsi, les établissements qui exercent des activités éditoriales ont créé un système dont les conditions de diffusion sont très réglementées. La diffusion peut-être directe ou indirecte, gratuite et/ou payante¹²⁸.

¹²⁸ Voir paragraphe 2.3.2

Les activités éditoriales de ces établissements étaient réglementées en 1998. En effet, la circulaire du 20 mars 1998¹²⁹ du premier ministre Lionel Jospin relative à l'activité éditoriale des administrations et des établissements publics de l'État « a donné un cadre normatif à l'activité d'édition publique et a désigné huit éditeurs dits institutionnels, dont l'activité d'édition participe des missions de l'établissement » (Levy-Rosenwald, 2012), parmi eux :

- Le Centre national de la recherche scientifique (CNRS), chargé de développer l'information scientifique et d'assurer l'élaboration et la diffusion de la documentation scientifique et la publication des travaux ;
- L'Institut géographique national (IGN), qui a pour mission d'assurer « la production, l'entretien et la diffusion de l'information géographique de référence en France » ; et
- La Direction des Journaux officiels, qui a pour vocation de publier et diffuser les textes législatifs et réglementaires et la Documentation française qui agit comme éditeur pour le compte d'administrations et d'organismes publics réunies par le décret n° 2010-31 du 11 janvier 2010 au sein de la Direction de l'information légale et administrative (DILA). La DILA assure plus particulièrement : l'édition et la diffusion des lois, des ordonnances, des décrets et d'autres actes ou documents administratifs qui doivent être publiés au Journal officiel de la République française ; la publicité des débats parlementaires ; la publication des informations contribuant à la transparence de la vie associative, économique et financière.

Il convient de signaler que l'activité éditoriale s'est développée au sein d'un grand nombre d'institutions publiques dont la vocation n'est pas de publier. Le Rapport du Médiateur de l'édition publique pour l'année 2011 (Levy-Rosenwald, 2012) a recensé 73 établissements qui exercent des activités éditoriales y compris L'Insee et Météo-France.

¹²⁹ Voir : Circulaire Du 20 Mars 1998 Relative À L'activité Éditoriale Des Administrations et Des Établissements Publics de l'Etat.

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000204563>.

Avant de présenter les évolutions des modes d'accès et des conditions d'usage des données publiques, nous revenons sur le contexte des années 90 qui a joué un rôle important dans l'organisation de l'édition publique et la construction des discours prospectifs sur l'accès aux informations dans les sociétés modernes. L'idée d'une « société de l'information » a connu un essor en Europe et en France. A l'initiative de l'État pour entrer dans la société de l'information, le « programme d'action gouvernemental pour la société de l'information » (PAGSI) a été lancé en août 1997 par le Premier ministre à l'époque Lionel Jospin et annoncé par Jean-Noël Tronc, conseiller pour la société de l'information, en Novembre 2000 à l'Unesco¹³⁰. Le gouvernement a élargi le principe d'accès aux documents publics en décidant la mise en ligne gratuite des « informations publiques essentielles ». Les informations essentielles au sens du projet de loi de la société de l'information¹³¹, présenté en 2001, constituent :

« 1° L'ensemble des actes et décisions, pris par l'Etat ou un de ses établissements publics administratifs, qui sont soumis à une obligation de publication en vertu de dispositions législatives ou réglementaires, ainsi que les documents qui leur sont annexés ;

2° Les informations sur l'organisation et le fonctionnement des services publics de nature à faciliter les démarches des usagers ;

3° Les rapports et les études sur les missions, l'organisation et le fonctionnement des services publics »¹³². Il s'agit que « les services et établissements publics administratifs de l'Etat mettent gratuitement à

¹³⁰ Tronc, Jean -Noël, 'Le Rôle Des Pouvoirs Publics Dans L'accès À L'information Faciliter L'accès Aux Réseaux et Services', *Unesco*, 2000.

http://www.google.fr/url?sa=trct=jq=esrc=ssource=webcd=3ved=0CD8QFjACurl=http%3A%2F%2Fwebworld.unesco.org%2Finfoethics2000%2Fdocuments%2Fpaper_tronc.rtfei=7bWJU-X4AcfJOfS3gPgKusg=AFQjCNFLMLdr58OllfPyHprXlx03ihMgsAsig2=otl12QzLrQ1J7A7VngOtZw.

¹³¹ Le projet de loi de la société de l'information a été abandonné avec l'adoption de la directive 98/CE du 17 novembre 2003 du Parlement européen et du Conseil concernant la réutilisation des informations du secteur public.

¹³² France. Assemblée Nationale. Projet de loi sur la société de l'information. 2001. <http://www.assemblee-nationale.fr/11/projets/pl3143.asp>

la disposition du public, sur des sites accessibles en ligne, les données essentielles qui les concernent »¹³³.

Ainsi, au niveau national, la mise en œuvre de la politique du gouvernement a donné lieu à la publication de plusieurs sites web pour « préparer l'entrée de la France dans la société de l'information ». « Il ne s'agit plus ici d'information spécialisée, professionnelle ou scientifique, mais de tout ce qui pourra contribuer à construire une " société de l'information " fondée sur une " culture de l'information " : ce programme d'action gouvernemental en faveur du développement d'Internet et du multimédia concerne l'éducation, la culture, les services publics, les entreprises, l'innovation et la régulation des réseaux »¹³⁴. Michel SAPIN, Ministre de la fonction publique et de la réforme de l'État dans un dossier sur « Le développement des sites internet des services de l'État » publié en 2001, affirme que l'internet public est aujourd'hui une réalité avec les sites des services de l'État¹³⁵. De cette époque, nous citons un ensemble d'organismes publics qui ont maintenu une famille de sites web publics qui offrent des informations gratuites sur le gouvernement et sur ses activités, par exemple, le portail de l'administration française lancé en 2000 (www.service-public.fr), Légifrance, le service public de la diffusion du droit lancé en 2002 (www.legifrance.gouv.fr), La Documentation Française (www.ladocumentationfrancaise.fr) et le portail de l'état sur les collectivités (www.collectivites-locales.gouv.fr). Moïse Mayo¹³⁶ explique que l'Insee partageait la même vision quand l'institut a commencé, en 2003, à publier sur le site web de l'Insee des tableaux de statistiques. Le but était à l'époque de minimiser les frais de production sur papier et de partager des données publiques gratuites.

¹³³ Ibid.

¹³⁴ Voir : « Le programme gouvernemental pour une société de l'information ». *Documentaliste* 35, no. 1 (January 1, 1998): 27–27. <http://www.adbs.fr/le-programme-gouvernemental-pour-une-societe-de-l-information-13481.htm?RH=REVUE>.

¹³⁵ Voir : Le Développement des sites internet des services de l'Etat : évaluation 2001. France : La documentation française, 2001. <http://www.ladocumentationfrancaise.fr/rapports-publics/014000796/index.shtml>

¹³⁶ Directeur de l'Insee à la région Bourgogne, entretien le 13 juin 2014.

Un premier retour sur les pratiques des institutions publiques concernant la diffusion de leurs données a donné des résultats décevants. En 2002, quatre ans après le lancement du programme d'action gouvernemental pour la société de l'information, Maurice Ronai¹³⁷ (2002), chargé à l'époque de mission du Commissariat Général du Plan¹³⁸, lors d'un Colloque sur la diffusion des données publiques, organisé par le Groupement Français des Industries de l'Information (GFII), signale que l'effort engagé par les administrations pour réaliser un inventaire préalable des gisements d'information disponibles et de définir une typologie des informations dont la diffusion est souhaitable s'avère un peu décevant car certaines administrations se sont contentées de recenser ce qu'ils diffusaient et non pas ce qu'ils produisaient comme systèmes d'information, bases de données ou gisements.

En revenant sur les origines historiques autour de la diffusion des données publiques, nous pouvons dire que l'open data s'inspire des dynamiques et discours en vogue autour de la société de l'information. Le mouvement open data, quant à lui, renouvelle l'engagement des institutions publiques à adopter une politique d'ouverture et poursuivre des efforts dans ce sens.

5.3 Evolution des modes d'accès : de l'accès sur demande à la diffusion proactive

Il nous semble pertinent de revenir aux concepts d'« accès » et de « diffusion » en soulignant la distinction entre eux. Le concept d'accès apparaît absolument fondamental à l'open data car il représente une des caractéristiques¹³⁹ d'une « donnée ouverte » accessible et disponible à tous. Le concept de diffusion représente cependant le

¹³⁷ Auteur en 1996 de Données publiques, accès, diffusion, commercialisation.

¹³⁸ Le Commissariat général du Plan (CGP) : une institution française qui existait entre 1946 et 2006 chargée de définir la planification économique du pays.

¹³⁹ Huit principes ont été énoncés par l'assemblée Open Government Data pour définir une donnée publique ouverte : complète, primaire, opportune, accessible, lisible par machine, non discriminatoire, non exclusive et non soumise à une licence. Voir chapitre 3.

processus d'ouverture qui révèle surtout l'évolution des modes d'accès et de diffusion des données publiques. En effet, le concept d'accès s'appuie sur l'ensemble des règles de la loi n° 78-753 du 17 juillet 1978 portant sur la liberté d'accès aux documents administratifs et de la réutilisation des informations publiques. L'accès, régi par cette loi, consiste pour tout citoyen à pouvoir consulter ou se faire communiquer un document administratif lorsqu'il en fait la demande. Pour accéder à un document le concernant, un individu doit en faire la demande, orale ou écrite, à l'autorité qui le détient. Avec le temps, ce principe d'accès sur demande s'est adapté aux évolutions techniques et aux pratiques numériques. Actuellement, le droit d'accès peut s'exercer de trois manières : a) consultation sur place ; b) délivrance d'une copie (aux frais du demandeur, au prix coutant le service de la copie) ou c) délivrance par courrier électronique.

Cependant, il existe une asymétrie entre la loi d'accès et le mouvement Open data. En effet, le principe d'accès à l'information publique est un instrument de la divulgation réactive, tandis que l'open data repose sur le concept de diffusion selon un système de divulgation proactive¹⁴⁰. La directive de l'association Sunlight Foundation¹⁴¹ sur la politique Open data souligne fortement ce point. De son point de vue, il faut libérer, de manière proactive, l'information gouvernementale en ligne. Les moyens de divulgation réactive supposent qu'une question ou une demande soit faite avant qu'une donnée ne soit divulguée. La diffusion est, à l'opposé, une action proactive, puisqu'il s'agit d'une libération de données avant que ne soit faite une demande. De même, la circulaire du 14 février 1994 relative à la diffusion des données publiques oppose la diffusion à l'accès. Elle précise que « *la diffusion des données publiques procède de la volonté de l'administration de faire parvenir au plus grand*

¹⁴⁰ Si nous parlons du concept d'« accès », il convient de le distinguer de celui défini par Rifkin. Comme le précise dans son ouvrage *l'âge de l'accès* (2000), « la notion d'accès se substitue à celle de propriété ». Ainsi, pour Rifkin, l'accès s'oppose à la possession, pourtant dans le cadre de l'Open data, l'accès est à l'opposé de la diffusion qui relève plus d'une diffusion proactive en ligne que d'un accès sur demande.

¹⁴¹ Sunlight Foundation, "Open Data Policy Guidelines," *Sunlight Foundation*, 2014, <http://sunlightfoundation.com/opendataguidelines/>.

nombre de personnes des informations collectées par elle ou élaborées en son sein »¹⁴².

Le terme « mise à disposition » est également accolé au mouvement Open data. Toutefois, ce terme ne comporte pas forcément une diffusion auprès du grand public. Mandelkern et Marais (1999) considèrent que l'administration crée un « système passif » de diffusion en mettant ses données à disposition sans réagir nécessairement par la mise en place d'une action de diffusion.

La distinction entre les deux concepts, accès et diffusion, montre un passage progressif d'une logique de demande à une logique d'offre. D'un point de vue matériel, selon Bruguière (2004), l'information accessible et l'information diffusée ne sauraient être confondues : la diffusion s'inscrit dans un processus unilatéral alors que l'accès suit un processus bilatéral. Il affirme que :

« L'information accessible est une information détenue à laquelle on peut, sous réserve de certaines conditions, avoir accès. Elle suppose donc une démarche volontaire de la part de celui ou celle qui y prête un intérêt. En ce sens, le processus est bilatéral. [...] L'information diffusée, disséminée [...] qui est par définition même non détenue, procède en revanche de la volonté de la personne qui a produit cette information de la faire parvenir au plus grand nombre ».

De fait, la diffusion proactive devient une opportunité permettant de faire évoluer le concept d'accès aux données publiques. Selon le modèle classique d'accès, l'utilisateur précise son besoin d'informations et l'institution publique réagit positivement ou négativement à sa demande. La diffusion telle qu'elle est envisagée dans le mouvement open data ne se réduit pas à un acte d'accès et les administrations publiques. D'ailleurs, le projet de loi¹⁴³ pour une République numérique introduit notamment l'ouverture par défaut des données publiques (*Open Data par défaut*) afin que la diffusion systématique

¹⁴² Voir : *Circulaire Du 14 Février 1994 Relative À La Diffusion Des Données Publiques*, 1994.
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000181007>.

¹⁴³https://www.legifrance.gouv.fr/affichLoiPreparation.do;jsessionid=71AB31DC7004C58B6AF64347F7C13C11.tpdila23v_3?idDocument=JORFDOLE000031589829&type=contenu&id=2&typeLoi=proj&legislature=14

en ligne devienne la règle. Or cela assume de nouvelles responsabilités de prise de décision et d'organisation d'un processus continu de diffusion. L'open data, comme nous allons le voir dans le chapitre 8, met en avant de nouveaux comportements et responsabilités au sein des administrations et celles-ci s'engagent alors, si elles adoptent une politique d'open data, à collecter et publier selon des modalités de gestion qui dépassent le système classique et traditionnel réactif de la requête d'accès aux données des administrations.

5.4 Évolution des conditions d'usage

Le concept d'open data est fortement lié à la question de « réutilisation ». Notre première définition de la réutilisation recouvre « *l'utilisation des données numériques structurées en vue d'un traitement informatique plus facile* » (Bruguière, 2004). Nous avons déjà montré que le potentiel de réutilisation des données est, en premier, lié à la structure même des données. Techniquement, plus les données sont structurées plus elles deviennent réutilisables. Juridiquement la notion de réutilisation recouvre premièrement « *l'utilisation des documents administratifs par des tiers à d'autres fins que celles de la mission de service public en vue de laquelle les documents ont été élaborés ou sont détenus* ». Dans ces conditions, toute réutilisation est une utilisation du contenu en dehors de son intention initiale.

En s'intéressant aux conditions d'usage des données publiques, nous avons repéré deux conditions assez essentielles et déterminantes qui affectent particulièrement la réutilisation des données publiques. Ainsi, la réutilisation dépend de la « finalité d'usage » d'une part, et des « conditions tarifaires applicables » de l'autre.

5.4.1 La réutilisation commerciale

Par rapport à la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 qui a posé le principe de la "réutilisation des informations du secteur public" au sein des États membres, la réutilisation est une exploitation des gisements de

données publiques. Comme nous l'avons déjà signalé, les pratiques de réutilisation existaient au sein du système traditionnel de communication et de diffusion informations publiques, pourtant le terme « réutilisation » est apparu dans les textes de loi français en 2005 lors de la transposition de la directive européenne 2003/98/CE par l'ordonnance du 6 juin 2005¹⁴⁴, modifiant la loi de 1978 sur l'accès aux documents administratifs. Le nouveau titre de la loi devient "*De la liberté d'accès aux documents administratifs et de la réutilisation des informations publiques*".

Comme nous l'avons déjà vu, pour l'Union européenne, la notion de réutilisation se retrouve dans une « approche économique orientée vers les entreprises ». La directive a donc posé le principe de non discrimination entre les utilisateurs commerciaux et non commerciaux et interdit les partenariats exclusifs. En effet, le but de cette directive transposée en France, était de créer un cadre communautaire pour l'exploitation de l'information émanant du secteur public. Rappelons le contexte de la création de la directive 2003/98/CE. La commission européenne dans une communication du 23 octobre 2001 intitulée « e-Europe 2002 : créer un cadre communautaire pour l'exploitation de l'information émanant du secteur public »¹⁴⁵ a relancé le projet de la directive 2003/98/CE concernant la réutilisation des informations du secteur public tout en considérant les données publiques comme ressource économique qui a des incidences positives sur le marché de l'information. D'où la valorisation des données publiques et l'incitation d'une part aux administrations d'ouvrir leurs données et d'autres part des acteurs économiques de les exploiter.

L'étude de Vickery, expert de l'OCDE (Organisation de coopération et de développement économiques), publiée par la commission européenne en 2011, montre en effet que le potentiel économique des données publiques est important. Une ouverture plus massive des données et une tarification marginale ou gratuite ont des impacts

¹⁴⁴ Ordonnance n°2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques.

¹⁴⁵ Voir : "eEurope 2002 : Créer Un Cadre Communautaire Pour L'exploitation de L'information Émanant Du Secteur Public" Commission Européenne, 2002.

économiques directs et indirects sur l'union européenne estimés de 140 milliards d'euros chaque année (Vickery , 2011). En France, la réutilisation commerciale des données publiques est en croissance lente depuis 2005. Les études du groupe SERDA en partenariat avec le Groupement Français de l'Industrie de l'Information (GFII) sur le marché des information électronique professionnelle montre une croissance lente du marché dérivant de l'information publique. En 2007, le marché est estimé à plus de 700 millions d'euros. En 2011, ce marché représente un chiffre d'affaire de 850 millions d'euros¹⁴⁶.

Dans les deux cas, la réutilisation constitue un nouveau pas en direction de l'usager des données publiques. C'est une « externalisation » qui consiste, pour une administration publique, à permettre aux usagers d'accéder plus facilement à des données publiques de valeurs afin de *pouvoir faire valoir leurs droits* et devenir un « réutilisateur », donc un « exploitateur » des données pour créer une plus-value. Les différents niveaux et modèles de réutilisation sont étudiés dans le chapitre 11.

5.4.2 De la redevance à la gratuité

Bien que la directive 2003/98/CE a posé le principe de non discrimination entre les utilisateurs commerciaux et non commerciaux et interdit les partenariats exclusifs, elle a gardé le principe de redevance. La notion de commercialisation de données publiques apparait comme centrale dans le système traditionnel de diffusion dans le secteur public. Dans sa thèse de doctorat en droit, Teresi (2007) a mis en avant les pratiques du commerce des données publiques en précisant que les institutions publiques procèdent depuis plusieurs années à la vente auprès du secteur privé de données produites par eux. « Des résultats d'examen aux données géo-spatiales, des informations sur l'état de l'atmosphère aux recensements de la population en passant par les décisions de justice, les annuaires et autres collections iconographiques, les données publiques font l'objet d'un commerce. La vente, entendue au sens économique, de données collectées ou produites à l'occasion d'une

¹⁴⁶ Voir Berthault, Denis, and Bernard Marx. "Tarification de la réutilisation des informations publiques en France." Les Cahiers du numérique Vol. 9, no. 1 (April 1, 2013): 39–53.

mission de service public est une manifestation des conséquences de la loi ainsi que, subséquemment, de la prévalence des valeurs des données publiques dans l'économie » (Teresi (2007)).

L'article 15 de la loi d'accès et de réutilisation des informations publiques précise les conditions pour l'établissement des redevances : « La réutilisation d'informations publiques peut donner lieu au versement de redevances ». La tarification est fondée sur une analyse des coûts assumés par le service public pour la « mise à disposition des informations, le traitement permettant de les rendre anonymes, la collecte et la production des informations. Selon cette loi, l'administration peut « inclure dans l'assiette de la redevance une *rémunération raisonnable de ses investissements* comprenant, le cas échéant, une part au titre des droits de propriété intellectuelle ». La transposition de la directive européenne par l'ordonnance du 6 juin 2005, a lancé, à l'époque, un débat entre divers acteurs publics et privés et la tarification était le principal sujet de contentieux vu que la directive prévoit une tarification qui va de la gratuité au recouvrement total des frais de production et diffusion (Battisti, 2004). Pour une mise à disposition, la « rémunération raisonnable » pose des difficultés comptables car « il semble impossible d'imposer un mode de tarification unique alors que les filières sont hétérogènes, car elles répondent à des missions de service public différentes » (Battisti, 2004, p.352). Même la version 2013 de la directive européenne n'a pas apporté des modifications majeures sur ce point. Les principes de tarification indiquent que « lorsque la réutilisation de documents est soumise à des redevances, les dites redevances sont limitées aux *coûts marginaux* de reproduction, de mise à disposition et de diffusion ». Cette règle ne s'applique pas dans les cas des organismes qui génèrent « des recettes destinées à couvrir une part substantielle des coûts liés à l'accomplissement de leurs missions de service public ou des coûts afférents à leur collecte, à leur production, à leur reproduction et à leur diffusion ». Les organismes concernés calculent le montant total des redevances en fonction de critères objectifs, transparents et vérifiables définis par les États membres. En plus les bibliothèques, les musées et les archives sont exclus du principe du « coûts marginaux » mais « le total des recettes provenant de la fourniture et des autorisations de réutilisation des documents pendant la période comptable appropriée ne dépasse pas le coût de collecte,

de production, de reproduction, de diffusion, de conservation et d'acquisition des droits, *tout en permettant un retour sur investissement raisonnable* » (Directive 2013/37/UE).

En application de l'article 16 de la directive, le paiement d'une redevance s'inscrit dans un cadre contractuel par la délivrance d'une licence de réutilisation : « Lorsqu'elle est soumise au paiement d'une redevance, la réutilisation d'informations publiques donne lieu à la délivrance d'une licence. Les administrations qui élaborent ou détiennent des documents contenant des informations publiques pouvant être réutilisées dans les conditions prévues au présent article sont tenues de mettre préalablement des licences types, le cas échéant par voie électronique, à la disposition des personnes intéressées par la réutilisation de ces informations » (Directive 2013/37/UE). La libéralisation des données par le secteur public se fait par le biais de contrats et de licences qui fixent les conditions dans lesquelles ces données pourront être réutilisées. Il s'agit d'un accès sur demande accordée à des cas précis et sous conditions. Pour les licences payantes, la difficulté est de déterminer l'exact coût d'une prestation. Dans ce cas, il faut connaître très précisément le financement et le coût du service en matière d'information. Selon Berthault et Marx (2013), deux méthodes s'appliquent. La méthode forfaitaire peut être modulée par certains éléments comme la périodicité des mises à jour, le volume d'information, la nature de l'entreprise, le nombre de filiales et l'objectif des réutilisateurs. Par contre, la tarification proportionnelle prend en compte le chiffre d'affaires généré par la réutilisation des informations publiques. Les prix de même que les modèles de tarification diffèrent considérablement d'un pays à l'autre, ainsi que d'un organisme du secteur public à l'autre dans un même Etat. Selon la commission européenne, la politique de prix devrait prendre en compte un certain nombre d'intérêts: un accès abordable pour tous ; un potentiel en termes d'exploitation ; une concurrence loyale.

Le mouvement de l'open data a renouvelé le débat sur la gratuité des données publiques et a mis en évidence la question de l'édition publique et le marché public de l'information et des données publiques. Dans le cadre de la politique nationale de la France en Open data, la circulaire du 26 mai 2011 qui a permis en pratique

l'application de nouvelles conditions de réutilisation, en particulier, la généralisation de la « réutilisation libre, facile et gratuite des informations publiques » en tenant en compte du système de redevances existantes. En Décembre 2011, une commission a été chargée de rendre un avis sur la liste des informations publiques dont la réutilisation est soumise à redevance. La liste publiée en juillet 2012 comporte 28 institutions publiques qui ont soumis leurs données à des redevances y compris la Direction de l'information légale et administrative (DILA), Centre national de documentation pédagogique (CNDP), Institut National de la Statistique et des Etudes Economiques (INSEE), Direction Générale des Collectivités Locales (DGCL), Institut National de l'Information Géographique et Forestière (IGN) et Commission d'accès aux documents administratifs (Cada)¹⁴⁷.

Dans le même sens, une étude¹⁴⁸ était ultérieurement menée, en 2013, par Mohammed Adnène Trojette, suite à la demande du premier ministre, pour évaluer et recenser des modèles économiques des redevances existantes portant sur l'utilisation des données publiques. Le rapport recense 20 établissements publics qui diffusent des données soumises à redevances. L'Insee et l'IGN sont des acteurs essentiels de production et de diffusion d'informations et de données publiques et ils perçoivent chacun près de 10 M€, équivalent ensemble à 70 % du total des redevances. Le rapport constate que celles-ci produisent des revenus décroissants (35 millions d'euros en 2012, -33% par rapport à 2010) et que 90% de ces revenus se concentrent sur dix jeux de données. La politique du gouvernement français sur la suppression de redevances et l'adoption du principe de gratuite sera en grande partie basée sur les propositions et les résultats de ce rapport¹⁴⁹. Les redevances fixées restent applicables

¹⁴⁷ La liste compète : <http://www.data-publica.com/data/14432--liste-des-donnees-publiques-payantes-juillet-2012>

¹⁴⁸ Trojette, Mohammed Adnène. Rapport Au Premier Ministre : Ouverture Des Données Publiques : Les Exceptions Au Principe de Gratuité Sont Elles Toutes Légitimes?, 2013. <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/134000739/0000.pdf>.

¹⁴⁹ France. Premier Ministre. Secrétariat général pour la modernisation de l'action publique. "Remise Du Rapport Trojette Sur L'ouverture Des Données Publiques" *Le Portail de La Modernisation de L'action*

en contradiction avec les appels à la gratuité. Cette étude démontre un écosystème des acteurs basé sur la redevance légalisée par la loi. Les institutions qui exercent des activités éditoriales ont créé un système dont les conditions de diffusion sont très réglementées. La diffusion peut-être directe ou indirecte, gratuite et/ou payante. Le choix de conserver le système de redevances semble non discutable car les redevances perçues représentent une activité lucrative pour les administratives dans le cadre de son auto-financement. Nous prenons comme exemple l'Insee et l'IGN, deux acteurs essentiels dans le marché des données publiques, qui vendent chacun près de 10 millions d'euros par an de données payantes. Nous signalons que l'Insee et l'IGN possèdent plusieurs modèles pour diffuser leurs données. Cela va de la gratuité au recouvrement total des frais de production et diffusion. Il existe plusieurs modèles et les prix dépendent du volume des données, l'identité du client et surtout des fins de la réutilisation des données de l'autre. Comme nous avons déjà signalé, l'Insee offre depuis 2003 des données gratuites sur son site web. Pourtant, l'Insee offre également des services informationnels dans le cadre du modèle payant progressif selon le volume des données. Dans la base de données des entreprises françaises « Sirene » produites par l'Insee, le tarif d'une liste d'établissements dépend du nombre d'établissements sélectionnés et du nombre de postes informatiques ayant accès à la liste. Si le prix d'une liste de 110 établissements coûte 99,35 €, la communication de la base Sirene complète est 70 374€. L'IGN a commencé, à partir de 2009, à publier gratuitement sur son site web des données géographiques moins détaillées que la version payantes. Il a ouvert ses données gratuitement aux organismes chargés d'une mission de service public en dehors des usages industriels et commerciaux. D'après le rapport Trojette (2013), cette initiative est justifiée par la montée en puissance des usages de données géographiques concurrentes, notamment celles issues d'OpenStreetMap. Par suite, l'IGN a développé des projets collaboratifs de «community sourcing» par la contribution de communautés d'utilisateurs impliquant, par exemple, avec les services de l'État et les collectivités territoriales.

Dans cet environnement concurrentiel, l'IGN produit les données géographiques de base qu'exige sa mission de service public. Ainsi le territoire français est intégralement couvert et la meilleure qualité est garantie. D'après Gaël Musquet, Président d'OpenStreetMap France la valeur ajoutée d'OSM est dans la micro géographie et la réactivité, cependant « il est plus facile d'être dans la donnée de l'immédiateté que dans le contrôle de l'information et de ses sources » rappelle Jean-Philippe Grelot, directeur général adjoint de l'IGN¹⁵⁰. L'institut planche sur la question du crowdsourcing et a commencé à développer des outils coopératifs. La nouvelle version du Géoportail propose par exemple aux internautes un outil de remontée d'informations et d'actualisation des cartes. Mais ces informations doivent être contrôlées avant d'être intégrées – quitte à perdre en réactivité – car la finalité reste la constitution d'une donnée de référence, fiable et qualifiée.

L'adoption du principe de gratuité a remis en question l'écosystème actuel de telles institutions. Selon Cédric Favre (2011), juriste rédacteur du Guide pratique des données publiques, le paiement d'une redevance est une question sensible et complexe. Il explique que « la constitution des données publiques est financée par l'impôt et elles ont pour but et finalité d'être directement utilisées par les administrations elles-mêmes. Leur réutilisation constitue une « utilisation anormale » dans le sens où elles ont été constituées pour être exploitées par les acteurs publics. En revanche, cette réutilisation offre une valorisation formidable de leur constitution. La redevance semble donc plus justifiée d'un point de vue économique que juridique. Mais faire payer une redevance peut créer un cercle vertueux, son versement pouvant dégager un autofinancement pour la constitution de nouvelles banques de données publiques » (Favre, 2011). Dans le même sens, François Perrussel-Morin¹⁵¹, directeur interrégional centre-est de l'IGN indique que l'IGN essaye de protéger son écosystème existant partiellement autofinancé par les redevances. Le budget est financé par 65% des subventions de l'Etat

¹⁵⁰ Interview dans «Dossier Open Data.» IGN Magazine, no. 68 (2012).

¹⁵¹ Discours tenus par François Perrussel-Morin lors du séminaire «L'Evolution de L'offre de Service Public IGN» dans *L'open Data : Nouvelle Richesse Du Web, Mais Pour Qui?* Lyon, 2014.

pour charges de service public et les redevances de réutilisation constituent un financement complémentaire et représentent 35 % du budget de l'IGN.

Ce débat open data entre redevance et gratuité des données publiques a débouché fin 2015 sur la gratuité de la réutilisation des informations du secteur public garantie par la loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public. Il s'agit d'une modification essentielle de la loi de 1978 mais cette nouvelle loi retient deux exceptions à la règle de gratuité : les administrations qui sont tenues de couvrir par leurs ressources propres une part substantielle des coûts engendrés par l'exécution de leur mission de service public (le cas de l'IGN) et les bibliothèques, musées et archives afin de compenser les coûts de numérisation de leurs fonds. Ainsi, dans un contexte de réduction budgétaire, la politique de tarification selon la règle d'exception demeure valable jusqu'au moment où l'application du principe de gratuité devient la règle générale.

Conclusion du chapitre 5

Dans ce chapitre, nous avons présenté le système traditionnel de diffusion de l'information publique en étudiant les évolutions qu'il a subies au niveau des modes d'accès ainsi qu'au niveau des conditions d'usage. Cette situation met en avant les lacunes du système traditionnel et nous aide à comprendre les impulsions au changement et à l'adoption de l'open data. Nous proposons de restituer le tableau 6 de synthèse de cette analyse comparative.

Tableau 6. Système traditionnel Vs Open data

	Système traditionnel	Open data
Objet de diffusion	Documents administratifs	Données publiques
Mode d'accès et de diffusion	Accès sur demande	Diffusion proactive
Acteurs de diffusion	Editeurs publics institutionnels	Tout établissement public
Conditions d'usage	Redevance sous des licences d'usage	Gratuité

En distinguant le principe d'accès sur demande du principe de diffusion proactive, nous avons montré le repositionnement, d'une part, des détenteurs des données publiques représentés par les institutions publiques, et des usagers et leurs droits de l'autre. Au delà des éditeurs publics institutionnels dont l'activité d'édition fait parti de leurs missions, les pratiques de diffusion des données publiques ont connu une véritable expansion pour atteindre d'autres établissements publics. Ce fait est remarqué suite à la montée de la société de l'information dans les années 1990. Considéré comme le premier pas vers l'ouverture, la politique de diffusion gratuite en ligne a donné lieu à la publication de plusieurs sites web pour « préparer l'entrée de la France dans la société de l'information ». Pourtant, cette diffusion était limitée et se réduit à l'offre sélective et distincte de données selon chaque institution. Afin d'assurer une diffusion plus large et complète des données publiques, le mouvement open data était un outil pour inciter les institutions publiques à renouveler leur

engagement et adopter la politique d'ouverture et poursuivre des efforts dans ce sens.

Concernant les usagers, nous avons remarqué une « approche économique orientée vers les usagers ». Une telle vision de la réutilisation mettra l'accent sur les bénéfices économiques de la réutilisation des données publiques en se basant sur le principe de non discrimination entre les utilisateurs commerciaux et non commerciaux et qui interdit les usages exclusifs. Enfin, de nombreux efforts ont été réalisés pour imposer l'ouverture gratuite mais sans trouver un substitut à l'écosystème basé sur la redevance toujours légalisée par la loi. Le développement des initiatives open data vont bien au-delà des obligations légales.

TROISIEME PARTIE : OPEN DATA LOCALE

Dans cette partie composée de six chapitres, nous souhaitons aborder l'open data par la présentation de la méthodologie et les résultats propres à notre terrain de recherche. Nous proposons dans cette partie analyser l'open data locale comme un processus suffisamment large qui recouvre l'ensemble de stratégies, d'actions ou de pratiques et des dispositifs qui sont mis en œuvre par les différents acteurs impliqués. Comme nous l'avons déjà souligné dans l'introduction, le processus comprend en premier les modes de production et de collecte de données, ensuite les méthodes de diffusion et les dispositifs utilisés, et enfin les modèles d'usage des données publiques numériques.

Pour cela, nous allons au chapitre 6 exposer notre terrain de recherche constitué de cinq collectivités territoriales et les méthodes de collecte de données nécessaires à l'avancée de notre recherche. Cette collecte de données nous amène par la suite à restituer les principales phases du processus d'ouverture (reproduites dans la figure 10) avant d'en présenter les résultats et de les discuter.

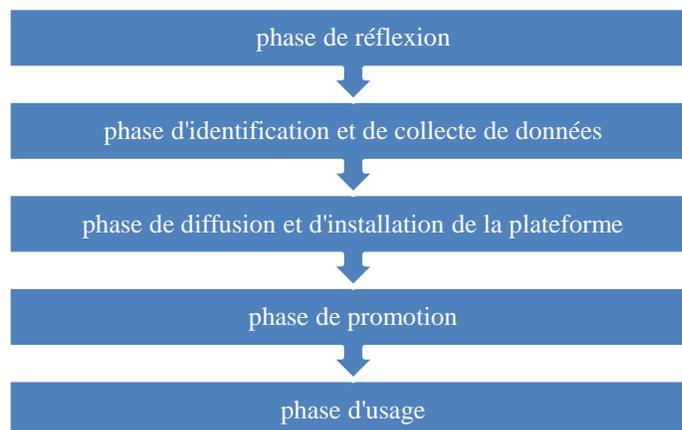


Figure 9. Les phases du processus d'ouverture

Ainsi la phase de réflexion, discutées dans le chapitre 7, montre la concertation des acteurs lors du montage du projet open data dans les collectivités. Le chapitre 8 aborde la phase de collecte de données qui concerne les facteurs qui influent sur le choix des données à ouvrir et le profil des acteurs producteurs de données publiques. Nous analysons dans le chapitre 9 les dispositifs de diffusion données

publique considérés comme des dispositifs de nature sociotechniques. Cette analyse nous permet également d'identifier le public d'utilisateurs ciblés. Le chapitre 10 exposera la phase de promotion du projet de l'open data au sein du territoire en intégrant les impacts et les usages résultants de la stratégie de promotion dans le contexte agité des usages sociaux des données. Ainsi nous proposons les caractéristiques du marketing de l'open data locale et les orientations de promotion de la donnée publique face aux logiques de réutilisations et les éléments fondamentaux de la réutilisation. Enfin, dans le chapitre 11, nous présentons bien entendu une analyse des profils des utilisateurs ainsi que les logiques et les modèles d'utilisation que nous avons pu repérés.

Chapitre 6. Méthodologie de recherche

Notre sixième chapitre présente notre terrain de recherche et les différentes méthodes de collecte et d'analyse nécessaires afin de répondre à notre problématique initiale : Quel est le positionnement des acteurs dans le processus d'ouverture des données publiques locales?¹⁵²

La réalisation de la recherche que nous présentons ici s'appuie fortement sur deux terrains de recherche : d'une part le terrain organisationnel et géographique et de l'autre le terrain technique. Le terrain organisationnel et géographique vise à affiner notre recherche au plan organisationnel aux cinq collectivités territoriales représentées par leurs conseils généraux et au plan géographique à cinq territoires dans lesquels les activités d'open data prennent place. Le terrain technique représente les dispositifs sociotechniques mises en place sur le web par ces collectivités dans le cadre de leurs projets open data.

Tout d'abord, nous présentons les résultats du recensement général que nous avons réalisés des projets open data des collectivités territoriales. Cela permettra de mieux situer notre recherche mais aussi de se référer à certains aspects propres aux collectivités sélectionnées lors de l'explicitation de choix opérés. Par la suite, nous reviendrons sur les critères de choix des projets open data des cinq collectivités territoriales que nous avons sélectionnées d'étudier en profondeur. Dans une troisième partie, nous proposons les méthodes de collectes de données ainsi que le corpus issus de nos terrains. D'un point de vue méthodologique, nous emploierons des mesures qualitatives dans nos analyses.

¹⁵² Voir l'introduction de cette thèse pour une mise en contexte de la problématique.

6.1 Recensement des projets open data locale

En première étape, nous avons dressé un panorama des expériences locales en matière d'ouverture des données qui consiste à rechercher non seulement les similitudes mais également les particularités de chaque projet.

Notre objectif est de recenser les projets Open data mis en œuvre au niveau des collectivités françaises quel que soit leur degré d'avancement. Le recensement systématique des activités d'Open data s'est fait via un travail de veille stratégique sollicitant plusieurs canaux et sources que nous détaillons ci-dessous. Nous avons pu créer un dossier sur l'état de l'open data dans les collectivités sans prétendre à l'exhaustivité¹⁵³. En effet, l'extension dynamique du mouvement Open data dans les collectivités françaises rend difficile le recensement total de toutes les initiatives. Les collectivités se trouvent dans des phases variées du processus d'Open data. Un nombre limité de collectivités a ouvert ses données et lancé des plateformes Open Data. Même à ce niveau, il existe encore des plateformes en « version beta » ou en « mode teste » avant le lancement de la version finale. D'autres collectivités sont « en cours » d'ouverture de leurs données. Des engagements publics des représentants de l'administration en faveur de la mise en place d'une « procédure » d'ouverture des données sont régulièrement annoncés. Des décisions politiques sont prises et les collectivités sont toujours en phase de « planification stratégique » avant de passer à la « phase d'action ». Nous trouvons également des collectivités qui envisagent d'ouvrir leurs données mais sont seulement au stade de la « réflexion ». Enfin, dans certains cas, la constitution de collectifs de citoyens sur un territoire pour promouvoir l'ouverture des données entre très souvent en ligne de compte comme un critère et une étape première vers l'ouverture des données dans une collectivité.

De plus, vu que le mouvement Open Data est récent, les critères de recensement sont régulièrement révisés pour rester en phase avec l'évolution du concept même de l'Open data. Même sur la carte

¹⁵³ Nous reproduisons dans l'annexe A le tableau de synthèse décrivant les projets open data en France.

participative de l'open data en France, mise à jour régulièrement sur le site de l'association LiberTic¹⁵⁴, les critères de sélection, pour figurer dans la liste des collectivités engagées dans l'open data, ne sont pas explicitement définis et déterminés. L'association invite régulièrement les internautes et les membres de la communauté open data à signaler les initiatives en matière d'ouverture de données publiques en France afin de suivre les avancées du mouvement. Cette carte identifie les projets de collectivités publiques, en cours ou déjà réalisés, de mise à disposition de données publiques. Elle montre les villes, les conseils généraux, les régions, les départements et les mouvements citoyens impliqués dans l'ouverture des données publiques en France. Mais, les différentes versions de la carte montrent des « entrants » et des « sortants » : de nouvelles collectivités sont ajoutées et d'autres sont retirées de la carte. C'est le cas de la commune « Le Havre », « Brest métropole » et le département d'Isère qui figuraient dans la version 3 de Juin 2011 de la carte Open data et qui ont été par la suite retirés de la version 4. Ces initiatives ne concernent qu'une « ouverture autonome » de données issues du système d'information géographique des collectivités. En effet, les conditions pour ajouter ou éliminer une collectivité de la carte ne sont pas définies. Parfois, un teaser sur Twitter qui annonce l'intention de la collectivité de s'investir dans l'open data est suffisant pour faire apparaître la collectivité sur la carte Open Data en France.

Dans le cadre de notre recensement, nous avons retenu seulement les collectivités qui satisfont aux critères suivants :

- A. Une ouverture effective (les collectivités qui sont en phase de réflexion avant l'étape d'ouverture ne sont pas prises en considération).
- B. Une ouverture dans des formats ouverts, structurés, et interopérables par machine (par exemple, fichiers Excel, CSV, XML. Notons qu'il existe des collectivités qui diffusent également sur le portail open data des formats propriétaires non structurés comme le format pdf).

¹⁵⁴<http://www.opendata-map.org/>

C. Une diffusion sous une licence libre (licence ouverte (LO), Open Database Licence, (Odbl), ou autre licence maison)

Nous présentons les résultats du recensement entre juin 2010 et Juin 2014 des projets open data au niveau des collectivités territoriales pour dresser un panorama des expériences locales. Ce recensement prend uniquement en compte les collectivités qui ont ouvert des données, y compris les données du Système d'Information Géographique (SIG), mais ne comprend pas les collectivités « en cours » d'ouvrir leurs données c'est-à-dire les collectivités qui sont toujours en phase de « réflexion » avant de passer à la « phase d'action ».

Le nombre des initiatives Open data locale évolue en France depuis la première initiative lancée par Rennes Métropole en 2010. D'après le site Keolis.Renne.com, Rennes Métropole a permis de déclencher un mouvement citoyen de grande ampleur sur le territoire Français¹⁵⁵. Les premières initiatives sont devenues un catalyseur pour les communautés et les villes à se joindre au mouvement de données ouvertes. En Juin 2014, nous recensons au total 69 collectivités territoriales qui ont ouvert des données en ligne soit en créant leurs propres plateformes Opendata ou en les hébergeant sur la plateforme nationale data.gouv.fr ou sur les plateformes mutualisées en partenariat avec d'autres collectivités souvent qui appartiennent à la même région. La courbe croissante (figure 5) établie en fonction de la date d'ouverture montre l'évolution du nombre des initiatives open data locale.

¹⁵⁵ Voir : Les Données Transport de Rennes Métropole En Accès Libre,” *Keolis Rennes.com*, accessed May 8, 2013, <http://data.keolis-rennes.com/fr/accueil.html>.

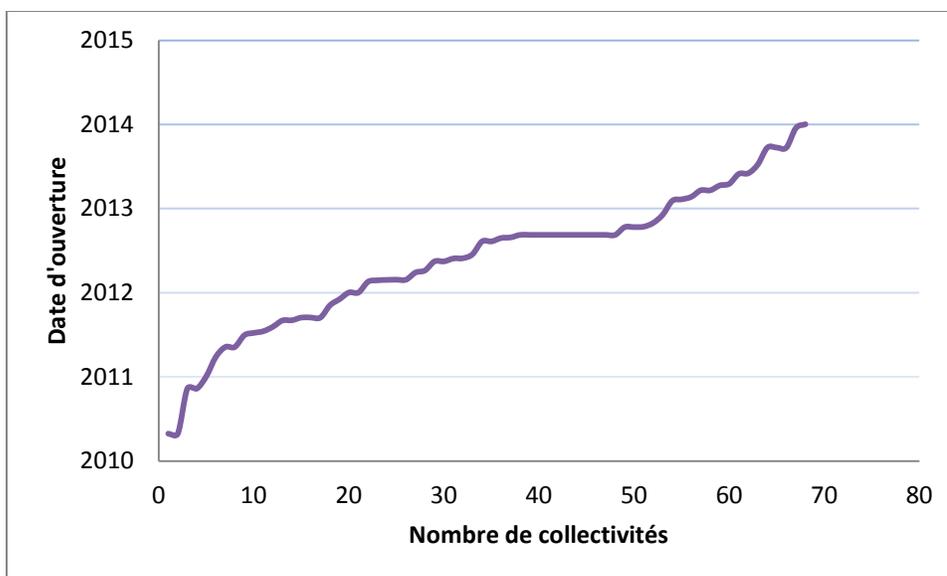


Figure 10. L'évolution des initiatives locales (Rahmé, 2015)

Le tableau suivant présente les 10 premières initiatives open data locale.

Tableau 7. Les 10 premières initiatives open data locale

	Collectivité	Date d'ouverture
#1	Rennes Métropole	Juin 2010
#2	Ville de Paris	Jan 2011
#3	Ville de Montpellier	Mar 2011
#4	Communauté urbaine de Bordeaux	Juin 2011
#5	Département de gironde en partenariat avec la région Aquitaine ()	Jui 2011
#6	Département de Saône-et-Loire	Sep 2011
#7	Département de Loir-Et-Cher	Sep 2011
#8	Communauté urbaine de Nice Côte d'Azur	Oct 2011
#9	Grand Toulouse	Oct 2011
#10	Nantes Métropole en partenariat avec la ville de Nantes, le département de Loire-Atlantique, et la région de Pays de la Loire	Nov 2011

Ce chiffre reste très restreint par rapport aux 36682 communes¹⁵⁶, 101 départements et 27 régions que la France compte. Bien que le mouvement open data ait été largement accepté au niveau national et dans les grandes villes, sa mise en pratique semble ne concerner que 18 % de l'ensemble des collectivités Françaises. De ce fait, nous constatons que l'intérêt pour l'open data reste inégalement représenté dans les stratégies et les politiques locales. Les collectivités ne sont pas au même stade de maturité dans leur réflexion et ne se représentent pas le problème de la même manière. Pour certaines, ouvrir les données sur des portails open data est une priorité à laquelle il faut répondre en adoptant une « politique volontariste » ; pour d'autres ce n'est qu'un service subalterne qui ne figure pas encore sur leurs agendas politiques. Il revient donc à chaque territoire, représenté par son conseil général, de fonder sa propre réflexion pour identifier le degré de priorité des projets open data dans leur stratégie d'action. Les collectivités qui ont franchi ce pas et ouvert leurs données sur des portails open data ont révélé des difficultés, des défis. Elles sont conscientes des avantages mais aussi des risques de l'open data et pourtant ils ont adopté une approche et une politique volontariste sans aucune obligation juridique.

6.2 Etude de 5 cas en profondeur

Le grand nombre de collectivités recensées (68 au total), nos moyens et le temps limité d'une recherche de doctorat, ne nous permet pas d'analyser tous les cas recensés. En plus, les particularités propres à chaque collectivité nous conduit à ne pas procéder seulement selon un critère de représentativité qui manque de profondeur, mais aussi selon un critère de pertinence pour montrer les différentes facettes et enjeux de l'implémentation du projet Open data. C'est pourquoi notre recherche et analyse s'appuient sur cinq projets open data dans cinq collectivités locales : Ville de Montpellier¹⁵⁷, Métropole de Lyon ou Grand Lyon, Métropole Grand Toulouse, Département du Saône et

¹⁵⁶ Au 1er janvier 2010, selon l'Insee, la France comptait 36 682 communes

¹⁵⁷ A ne pas confondre avec le projet open data de l'intercommunalité *Montpellier Méditerranée Métropole*

Loire, et la Région Provence-Alpes-Côte d'Azur (PACA) (voir tableau 8). Ces cas sont sélectionnés d'après des critères établis selon leur pertinence par rapport aux questions de recherche. Le critère principal de sélection est que la collectivité ait déjà ouvert ses données. La phase de post-ouverture est critique pour comprendre la réalisation des objectifs escomptés de l'open data, étudier le dispositif de diffusion et avoir un retour sur l'usage établi des données ouvertes.

Ainsi, nous privilégions les collectivités en phase avancée¹⁵⁸ de leurs projets open data pour pouvoir étudier tous les enjeux, les bonnes pratiques, les facteurs clés de succès et les expériences menées par les uns et les autres en matière d'ouverture de données. La sélection des cas doit répondre aux objectifs de notre recherche qui consistent à étudier la dynamique du mouvement de l'open data en contexte. Ainsi, nous retenons dans notre recherche les collectivités en phase de post-ouverture en tenant en compte des sous-critères de sélection suivants :

1. Date de lancement du projet : les cas choisis sont parmi les premières collectivités en France à ouvrir leurs portails open data. L'étude des enjeux de l'ouverture demande de prendre le temps du recul nécessaire pour tirer les enseignements positifs et négatifs du projet open data. Nous pouvons ainsi porter un regard sur la démarche développée, les pratiques et les méthodes employées, les productions réalisées, ainsi que le rôle et le niveau d'implication des acteurs concernés.
2. Volume de données ouvertes : c'est un autre critère à prendre en considération pour pouvoir étudier la typologie des données ouvertes et leurs caractéristiques techniques.
3. Dynamisme d'acteurs autour des données ouvertes : la présence d'une communauté open data locale, l'organisation des événements, et la réalisation de produits à partir des données ouvertes nous permettent de repérer les acteurs intéressés par l'open data et d'identifier leurs activités.
4. Type de collectivité : nous avons délibérément sélectionné des collectivités locales à différents niveaux (commune, communauté d'agglomération, département et région) pour

¹⁵⁸ La phase avancée correspond à l'ensemble de sous-critères expliqués dans le paragraphe qui suit.

montrer les enjeux de la taille de la collectivité sur le projet open data et vérifier les relations et partenariats intercommunales.

Tableau 8. Les projets open data étudiés

Collectivité	Type de collectivité	Portail open data	Date d'ouverture	Volume de jeux de données Aout 2013/ Novembre 2015
Montpellier	Commune	Montpellier territoire numérique – données publiques ouvertes http://opendata.montpellier.fr/	Mars 2011	95/247
Métropole de Lyon (Grand Lyon)	Intercommunalité	Data Grand Lyon http://data.grandlyon.com/	Août 2012	165/573
Toulouse Métropole	Intercommunalité	Data.toulouse-metropole https://data.toulouse-metropole.fr	Octobre 2011	49/212
Saône et Loire	Département	Open data 71 http://www.opendata71.fr/	Septembre 2011	49/280
Provence-Alpes-Côte d'Azur (PACA)	Région	Open PACA http://opendata.regionpaca.fr/	Juillet 2012	230/586

Chacune des collectivités sélectionnées représentent des particularités intéressantes à étudier.

- 1 La ville de Montpellier est très engagée dans l'ouverture des données. Suite au lancement du portail, divers événements organisés comme « Les info lab », les appels d'offres, les concours, la cartopartie d'OpenStreetMap (OSM) vulgarisent et favorisent l'utilisation et la réutilisation des données ouvertes par la ville.
- 2 Le projet Grand Lyon Smart Data s'oriente vers un modèle économique de l'open data regroupant le secteur public et privé. Le titre « Smart data », ainsi que la formulation de licences qui nécessitent une authentification des usagers et la possibilité de redevance expliquent cette déviation. Un modèle particulier des expériences de l'open data en France qui attirait notre attention doit être étudié davantage.
- 3 Toulouse Métropole était parmi les premières villes en France à s'engager dans une politique d'open data. Depuis octobre 2011, la métropole organise des concours pour dynamiser les acteurs locaux.
- 4 Le projet open data du département du Saône et Loire était le premier de ce genre au niveau des départements français. Le portail du département représente une particularité dans la représentation visuelle des données. Le portail a été le premier en France à intégrer, outre une interface présentant les données brutes dédiées aux développeurs, un outil de visualisation pour un grand public.
- 5 La Région Provence-Alpes-Côte d'Azur (PACA) était la première au niveau des régions françaises à s'engager dans un projet open data¹⁵⁹. La spécificité du projet open data de la région PACA réside dans la démarche partenariale par la mutualisation de la plateforme d'open data de la région avec divers partenaires producteurs de données publiques qui souhaitent publier leurs données : des communes, agence publiques, associations publiques et privées. Cette approche nous aide à s'interroger sur le type de partenariat entre les différents acteurs, les méthodes et les outils de publication, et la normalisation des données.

¹⁵⁹ Voir le tableau de synthèse des projets open data en France.

Ces cinq expériences étudiées sont loin d'inclure toutes les situations possibles, mais donnent cependant un aperçu global des enjeux de l'open data locale. Les expériences d'autres collectivités en France servent à enrichir notre étude en tirant des exemples de différents contextes. Ils viennent au besoin confirmer ou bien nuancer nos résultats et nos propositions.

C'est donc dans cet esprit que nous visons à identifier, d'une part, les stratégies des collectivités locales dans la mise en œuvre de leurs projets d'Open data, depuis la réflexion jusqu'à l'exécution en passant par la gestion du projet et la réorganisation des processus, et d'autre part, les pratiques des citoyens-usagers au regard des données ouvertes par les collectivités.

6.3 Méthodes de collecte des données

Si nous mettons méthodes au pluriel, c'est pour signifier qu'ici nous présentons plusieurs méthodes de collecte de données visant deux objectifs :

1. Etudier le terrain organisationnel et géographique par une analyse d'un corpus de discours.
2. Etudier le terrain technique par une analyse du corpus de dispositifs sociotechniques.

Notre analyse est alors fondée sur un continuum de deux corpus complémentaires mais qui ont des statuts et des objectifs différents. Notre préoccupation est de maintenir la cohérence entre notre approche sociotechnique présentée dans l'introduction de cette thèse et la méthode adaptée. L'avantage de croiser ces corpus est qu'il est particulièrement efficace en ce qu'il permet une compréhension et une analyse globale des enjeux de l'open data.

6.3.1 Corpus de discours

Nous nous intéresserons dans ce cadre aux discours tenus par les principaux acteurs de l'ouverture des données publiques en France et notamment aux discours des acteurs sur l'open data local. Nous procédons, ainsi, à une étude discursive de ce corpus afin d'identifier

les principaux traits du discours des acteurs l'open data. Nous nous mettons à distance égale de ces discours qui reflètent le point de vue de leurs énonciateurs pour avoir un recul et faire intervenir l'ensemble des acteurs impliqués dans la piste d'Open data.

Pour collecter les données et identifier les sources d'information, nous avons eu recours à trois techniques : la recherche et veille documentaire, les conférences, les entretiens semi-directifs.

I. La recherche et veille documentaire

L'open data est l'objet de nombreux débats sur les forums et les blogs des différents acteurs qui se caractérisent par leur forte présence en ligne. La communauté open data est active en ligne sur les blogs et les réseaux sociaux. Les développeurs blogueurs de la communauté open data partagent leurs expériences et réflexions sur l'open data. Les sites web, gérés par les chefs des projets open data dans les collectivités, forment aussi une ressource d'information riche et à jour et un milieu d'interaction entre les différents acteurs tels que les blogs et les forums des collectivités impliquées dans l'open data. Le blog¹⁶⁰ du projet open data à la ville de Montpellier, géré par Jérémie Valentin, le chef du projet, offre une veille exhaustive sur l'actualité de l'open data. Sandrine Mathon, responsable du projet open data dans la ville de Toulouse, partage sur scoop.it, l'outil de veille sur internet, les dernières publications sur l'open data.

Nous avons eu recours également aux discours portés par les medias. En effet, l'open data est régulièrement abordée dans les actualités française et internationale. C'est un thème vulgarisé par la presse et les magazines en ligne. Nous citons le dossier « Réutilisation des données publiques” de La Gazette des communes¹⁶¹, un magazine de référence sur l'actualité de la Fonction Publique Territoriale et des collectivités locales qui a créé un dossier de plusieurs articles traitant le sujet de l'Open data selon différentes approches.

¹⁶⁰ “Blog - Données Publiques Ouvertes - Montpellier Territoire Numérique”.
<http://opendata.montpelliernumerique.fr/Blog>.

¹⁶¹<http://www.lagazettedescommunes.com>

Nous nous intéresserons également aux discours des promoteurs de l'open data. Notre étude se base sur une exploration des différentes ressources disponibles – des rapports, études, guides, et articles de blogs - publiés par des associations militants pour l'open data. L'association FING, un Think tank de référence sur les transformations numériques a créé en 2012 dans son magazine « InternetActu » un dossier sur l'open data. En plus l'association a publié un « Guide pratique de l'ouverture des données publiques territoriales »¹⁶². Le « Manuel de l'open data »¹⁶³ de l'Open Knowledge Foundation (OKF), le blog de l'association Libertic¹⁶⁴, le site web de l'association Open Data France¹⁶⁵, le « Manuel d'ouverture des données publiques »¹⁶⁶ de l'association Sunlight Foundation, l'étude¹⁶⁷ de l'Open Society Foundation, et d'autres ressources, participent activement à la promotion de l'ouverture des données publiques en suivant l'actualité et les enjeux de l'open data.

Pour constituer ce corpus, une veille quotidienne était nécessaire pour suivre l'évolution rapide des pratiques de l'open data aux niveaux local. Le corpus a été constitué par la mise en place d'alerte email pour suivre les différentes ressources déjà cités : Nous nous sommes inscrits à des listes de diffusion comme Echodocs, portail destiné aux spécialistes en Information et en Documentation, et à « LaLIST » de l'Institut de l'information scientifique et technique (INIST). Nous sommes devenus « followers » de plusieurs acteurs et dossiers sur l'outil de veille Scoop.it. et sur le réseau social academia.edu. Les

¹⁶² Fondation Internet Nouvelle Génération, « Guide Pratique de L'ouverture Des Données Publiques Territoriales, » *Fing*, January 1, 2011, <http://fing.org/?Guide-pratique-de-l-ouverture-des&lang=fr>.

¹⁶³ Open Knowledge Foundation, *Le Manuel de L'opendata* (Open Knowledge Foundation, 2012), <http://opendatahandbook.org/fr/>.

¹⁶⁴ Association Libertic. « Blog de L'association Libertic, » accessed June 17, 2014, <http://libertic.wordpress.com/>.

¹⁶⁵ « Opendata France, » accessed June 17, 2014, <http://opendatafrance.net/>.

¹⁶⁶ Sunlight Foundation, « Open Data Policy Guidelines, » *Sunlight Foundation*, 2014, <http://sunlightfoundation.com/opendataguidelines/>.

¹⁶⁷ Becky Hogge and Open Society Foundations, *Open Data Study* (London, UK: Open Society Foundation, 2010), <http://www.opensocietyfoundations.org/reports/open-data-study>.

publications officielles françaises sont accessibles en ligne via la base de données de Legifrance.fr et le site ladocumenationfrancaise.fr

Dans la littérature, nous repérons seulement deux ressources qui traitent l'open data locale : un « guide pratique de l'ouverture des données territoriales » publié en 2011 par la Fondation Internet Nouvelle Génération et un autre « guide juridique » basé sur la loi Cada, publié en 2010, par la région Aquitaine, sur les enjeux et cadre juridique de l'ouverture des données publiques dans les collectivités.

II. Les conférences

En plus du corpus documentaire, nous avons collecté le discours des acteurs dans plusieurs manifestations sur l'open data. Les 15 conférences¹⁶⁸, séminaires et journées d'études auxquels nous avons assisté de Novembre 2011 à Juin 2014 étaient une occasion pour rencontrer les acteurs de l'open data en France. Des chercheurs, des chefs de projets open data au niveau national et local, des élus locaux, des développeurs, des chefs d'entreprises, des data journalistes, des graphistes, sont intervenus lors de ces manifestations sur leurs expériences et leurs positionnements en soulignant les différents enjeux de l'open data. Notre participation nous a permis également de repérer les acteurs/intervenants et de suivre les problématiques actuelles que pose l'open data.

III. Les entretiens semi-directifs

La technique d'entretiens semi-directifs apporte de nouveaux éléments au corpus de discours. Elle s'inscrit dans notre approche qualitative pour étudier les stratégies des acteurs à différents moments du processus de l'open data. L'objectif est de dépasser le discours « public », de rencontrer les acteurs en face à face et de découvrir les motivations réelles et leurs attitudes vis-à-vis de l'open data. Les entretiens nous aident à comprendre la stratégie des acteurs et leurs attitudes par rapport à l'open data : pourquoi ils approuvent ou non telle ou telle pratique par rapport au processus d'ouverture des données ?

¹⁶⁸ Voir la liste des manifestations auxquelles nous avons assisté en Annexe A.

L'essentiel ne réside pas dans le nombre de personnes interrogées mais bien dans le choix des individus interviewés, dans la manière de les interroger et dans la nature des questions à leur poser. La technique d'entretiens semi-directifs nous aide à approfondir nos réflexions et à décrire leurs expériences et leur savoir faire. Les personnes interviewées représentent les principaux acteurs impliqués dans le processus d'ouverture des données publiques locales. Ils partagent tous la même spécificité : ils sont concernés directement par l'ouverture des données publiques locales et leurs profils et statuts changent par rapport à leurs rôles dans le processus. Nous les avons regroupés en quatre catégories :

- A. les chefs de projets open data dans les collectivités
- B. les spécialistes de collecte, de traitement et de diffusion des données publiques
- C. les utilisateurs de données publiques
- D. les militants de l'open data locale

Pour construire notre grille de questions, nous sommes partis du constat que le processus d'ouverture des données publiques comprend à la fois les modes de production et de collecte de données, les méthodes de diffusion et les dispositifs utilisés, et enfin les modèles d'usage des données publiques numériques accompagnés par des stratégies de promotion. Dans le cas d'un échantillon segmenté en différentes catégories, la construction d'une grille de questions s'avère plus complexe. En effet, nous avons choisi d'élaborer plusieurs guides d'entretiens¹⁶⁹ que chacun correspond au profil et au statut de la personne interrogée. Voici donc la description des 12 personnes retenues pour analyse (tableau 9), ainsi que les motivations à les sélectionner.

¹⁶⁹ Les guides d'entretiens sont en annexes.

Tableau 9. Catégories des personnes interviewées et les structures auxquelles elles appartiennent

Catégorie	Structure
A. chefs de projets open data dans les collectivités	<ol style="list-style-type: none"> 1. Grand Lyon 2. Ville de Montpellier 3. Toulouse Métropole 4. Département du Saône et Loire 5. Région Provence-Alpes-Côte d'Azur
B. Spécialistes de collecte, de traitement et de diffusion des données publiques	<ol style="list-style-type: none"> 6. Institut national de la statistique et des études économiques (Insee) 7. Archives publiques de Saint Etienne.
C. Utilisateurs de données publiques	<ol style="list-style-type: none"> 8. Gazette des communes 9. Transpacity 10. Développeur indépendant
D. Militants de l'open data locale	<ol style="list-style-type: none"> 11. OpenStreetMap 12. OpenData69

Pour la catégorie A, cinq entretiens ont été réalisés avec des chefs et responsables de projets open data des cinq collectivités¹⁷⁰. Notre objectif étant de comprendre les facteurs qui contribuent et influent sur l'engagement des collectivités dans un projet Open data et les objectifs précisés. Quelle est la démarche d'ouverture, comment le projet a-t-il évolué et quelles sont les étapes clés dans un projet open data ? Qui sont les acteurs internes et externes de l'open data local ? En plus, nous cherchons à cerner les défis rencontrés dans l'implémentation, le type des usagers intéressés et les applications dérivées des données ouvertes. Nous avons alors rencontré les chefs des projets open data de la ville de Montpellier, de Grand Lyon, de

¹⁷⁰ Voir le guide d'entretien avec les chefs de projets open data en Annexe B

Toulouse Métropole, du département du Saône et Loire, et la Région Provence-Alpes-Côte d'Azur (PACA).

1. Le premier était avec Florence MARTINEZ , géographe et responsable de l'unité diffusion des données et géo-services au Service Information Géographique (SIG) de Grand Lyon.
2. Ensuite, nous avons interviewé Jérémie VALENTIN, géographe et chef de projet open data de la ville de Montpellier très engagé dans l'ouverture des données sur Montpellier et dans l'association open data France. Son travail de veille offre sur le blog du portail open data toutes les actualités de l'open data en France. Suite au lancement du portail, divers événements organisés comme Les info lab, les appels d'offres, les concours, la cartopartie d'OpenStreetMap (OSM) vulgarisent et favorisent l'utilisation et la réutilisation des données ouvertes par la ville.
3. Nous avons rencontré également Sandrine MATHON, juriste et chef de projet open data à Toulouse Métropole. Attachée à la Direction juridique de la CNIL (Commission Nationale Informatique et Libertés), elle est responsable du service Administration de la Direction des Systèmes d'Information de la Ville et de la Communauté Urbaine. La métropole s'est engagée dans une politique d'open data depuis octobre 2011. Sandrine MATHON est engagée au niveau national au sein de l'association Open data France.
4. Le quatrième entretien était avec Jean-Michel MARTIN, directeur de la Direction des Systèmes d'Information et de l'Information Géographique (DSIIG) chargée de l'ouverture des données dans le département du Saône et Loire en présence de Laurine DANTAN, Chef de projet informatique.
5. En fin, la dernière rencontre avec les responsables de projet open data dans les collectivités était avec Sophie TENDEIRO, Chef de projet « Services numériques innovants et mutualisation des données publiques », et responsable de l'ouverture des données de la Région Provence-Alpes-Côte d'Azur.

Pour la catégorie B, nous avons rencontré Moise MAYO¹⁷¹, directeur de l'Institut national de la statistique et des études économiques (Insee) à la région Bourgogne et Cyril LONGIN¹⁷², chef du projet Open data et directeur des archives publiques de Saint Etienne. Ces deux acteurs représentent les acteurs traditionnels et spécialistes de l'information publique. Leurs métiers et leurs implications dans des projets open data apportent un éclairage particulier qui nous a permis de comprendre le système actuel de communication et de diffusion des données et des informations publiques dans l'intérêt de découvrir ce qu'apporte le mouvement open data de nouvelles pratiques.

6. Moise MAYO fait parti, en tant qu'expert de production et de diffusion des données locales, du comité d'éthique du projet open data du département de Saône et Loire. Ce comité, composé de cinq membres, représente la société civile, un acteur passif dans d'autres collectivités.
7. Cyril LONGIN est chef du projet Open data et directeur des archives publiques de Saint Etienne. Il a participé également au groupe de travail sur la question de l'Open Data au sein de l'Association des Archivistes Français (AAF). Sa double position en tant qu'archiviste et chef de projet open data, nous interpelle sur la question de l'open data comme une opportunité pour les archivistes. Pourquoi les archivistes se montrent réservés à ce mouvement? Est-ce que les données publiques sont des biens communs?

Dans la catégorie C, nous avons rencontré également des utilisateurs de données publiques pour recueillir des informations sur leurs expériences d'usage de données publiques. La catégorie d'utilisateur comporte un journaliste, un développeur d'applications informatique et un chef d'entreprise.

8. Romain MAZON est journaliste et rédacteur en chef web de la Gazette des communes, une magazine de référence sur l'actualité de la Fonction Publique Territoriale et des collectivités locales qui a créé un dossier sur la réutilisation des données

¹⁷¹ Voir le guide d'entretien avec les chefs de projets open data en Annexe C

¹⁷² Voir le guide d'entretien avec les chefs de projets open data en Annexe D

publiques traitant le sujet de l'Open data selon différentes approches. Il nous a été recommandé par Fidelia Ibekwe-SanJuan car sa double position de Mazon en tant qu'acteur "observateur-évaluateur" des pratiques open data dans les collectivités et "chercheur-utilisateur" des données sur les collectivités est très intéressante. Son expérience montre les enjeux, les bonnes pratiques et les usages journalistiques de l'Open Data.

9. Cyril MOTTIER, développeur d'applications mobiles, a créé deux applications urbaines non officielles en utilisant des données sur les stations de VéloV à Lyon et à Paris. Ses pratiques de production nous éclairent sur le rôle du développeur en tant qu'intermédiaire entre le producteur des données et l'utilisateur final, ses pratiques de réutilisation de données ouvertes, la modalité de collecte d'extraction des données, ses sources et l'écosystème des applications gratuites.
10. Didier GORRON est président et fondateur de l'entreprise Transpacity (Transpacités avant 2013). Les activités principales de l'entreprise consistent à collecter, structurer et analyser les données sur les collectivités dans le but d'offrir des services informationnelles et de marketing aux opérateurs privés spécialisés dans quatre domaines publics ou « métiers » : les déchets, l'eau, le transport en commun, l'énergie. En plus, il a participé au concours « data connexion » organisé par Etalab. Il nous a été recommandé par Fidelia Ibekwe-SanJuan, et nous l'avons sollicité pour avoir son expérience avec l'usage de données publiques dans le cadre de son travail.

Enfin, dans la catégorie D, les deux derniers entretiens ont été menés avec des partisans du mouvement open data et mouvement du libre.

11. Florian FARGE est contributeur actif dans OpenStreetMap, un projet collaboratif visant à créer une base de données géographique réutilisable librement. Florian Farge y contribue depuis quatre ans. Il participe ainsi localement à la production de données et aux événements liés au monde de l'édition collaborative. OpenStreetMap, une pratique du crowdsourcing, est une base de données libre considérée comme une source

privilegiée de données géographiques ouvertes sur le web. Nous l'avons rencontré pour la première fois lors de notre visite au salon Primevère.

12. Sami RABIH, développeur en informatique, est membre du collectif Open data 69. Le collectif est formé d'un groupe informel d'habitants du Rhône dans le but d'être facilitateur et contributeur à l'application de l'open data. Sami Rabih montre la perception citoyenne de l'open data et le rôle des collectifs citoyens dans l'ouverture des données locales. Nous l'avons rencontré pour la première fois lors de notre participation à la réunion du collectif.

Les entretiens ont eu lieu entre 13 décembre 2012 et 21 juillet 2014. Le déroulement des entretiens a été globalement similaire pour chacun d'entre eux : ils ont tous eu lieu en face à face, chaque entretien a été enregistré et les temps d'enregistrement varient de 45mn pour le plus court à 2hrs pour le plus long.

6.3.2 Corpus de dispositifs sociotechniques

Si notre étude privilégie une approche compréhensive fondée sur une analyse d'un corpus de discours (corpus documentaire, conférences et entretiens) nous ne pouvons pas faire l'impasse sur le dispositif installé pour la diffusion des données. Comment observer la diffusion à l'aune du dispositif sociotechnique ? Ayant déjà formulé l'hypothèse que les plateformes open data se définissent comme étant des dispositifs sociotechniques, nous proposons de les analyser selon la « grille d'observation du dispositif sociotechnique » développée par Jouët et Le Caroff (2013). Cette grille d'observation est composée d'indicateurs qui attestent de l'imbrication étroite de la technique et du social (Voir tableau 10). Comme l'explique Jouët et Le Caroff (2013, p.149.): « les outils disponibles figurent du côté de la technique (colonne de gauche) mais, dès que l'utilisateur s'en empare, ils deviennent des outils sociaux car ils donnent lieu à des formes d'action (colonne de droite) ». Dans l'interface des plateformes open data, nous essayerons de repérer les signes qui révèlent l'identité du site qui se donne à voir dans son architecture et son contenu, ses

modes de participation et de lien social en ligne¹⁷³. Même si cette technique est compatible avec notre méthodologie et la dimension sociotechnique de notre étude, bien entendu, ce cadre méthodologique général demande, en raison de la diversité de la configuration des dispositifs, à être adapté aux spécificités des plateformes open data.

Notre corpus composé des dispositifs de diffusion des données installés par les cinq collectivités territoriales étudiées sera analysé par rapport à ces trois entités : la structure et le fonctionnement du site, les modalités de participation en ligne et enfin les configurations du lien social. La structure et le fonctionnement du site nous permet, en premier, d'observer la ligne éditoriale de la collectivité à partir de la mise en page, les rubriques de la plateforme, le contenu et les formats afin de comprendre comment sont représentées les données, de définir leurs natures, leurs sources et leurs formats, et d'identifier le public ciblé. La structure et le fonctionnement du dispositif disposent également d'une panoplie d'indicateurs pour connaître voire gérer leurs utilisateurs : nombre de téléchargement, de réaction déclenchées et de partages. Les modalités de participation en ligne permettent plusieurs degrés d'implication allant d'un simple clic sur « j'aime » à l'auto-publication. Enfin, les configurations du lien social réfèrent à la production d'échange entre les membres d'un site ou d'une plateforme. Dans notre recherche, nous souhaitons mettre en perspective une approche spécifique des dispositifs sociotechniques qui s'intéresse au design des technologies de diffusion et de participation en ligne.

Ce terrain nous permet également de discerner les usages et les produits dérivés de l'open data, ainsi que de démontrer le profil des usagers que nous présenterons et questionnerons dans les Chapitres 10 et 11. Dans ces espaces, les producteurs de données essayent de recenser les applications de l'open data. De même, les usagers sont invités à référencer leurs produits créés à partir des données ouvertes.

¹⁷³ Que nous présenterons et questionnerons dans le Chapitre 9.

Nous signalons que des bornes temporelles ont été fixées pour l'observation, opération indispensable en raison du caractère instable et évolutif des plateformes. Ici la période retenue s'est étendue sur les deux ans 2013 et 2014. Cette observation en ligne a requiert une immersion au cœur de notre terrain de dispositifs et qui a consisté une présence prolongée et régulière sur les espaces observés en ligne. Il s'agit par exemple d'une consultation fréquente des plateformes, une observation de l'architecture et du fonctionnement des sites, la recherche de jeux de données, un abonnement grâce au format RSS afin d'être informée immédiatement de l'ajout de nouveau contenu et l'adhésion au forum des plateformes.

Tableau 10. Grille d'observation du dispositif sociotechnique (Jouët et Le Caroff, 2013)

	TECHNIQUE	SOCIAL
Architecture et fonctionnement du site		
Ligne éditoriale	Morphologie du site	Captation de l'attention
	Mise en page, charte graphique	Attractivité
	Contenu	Public
	Rubriques, hiérarchie, format, ton	Préqualification
Gestion du site	Modération	Netiquette
	Modération, chartes de participation	Appréciation subjective
	Outils d'alerte	Régulation collaborative
	Audience	Tactiques de visibilité
	Indicateurs : Nombre de « lus », « réactions », de partages, de suivis (favoris, abonnements)	Autoévaluation, autopromotion des contributions
	Outils de visibilité	Stratégie personnelles d'audience
Participation		
Contributions	Outils de participation	Formes de contribution
	Posts	Auto-publications

	Comments	Réactions
	Like de contenus	Approbation
Personnalisation	Outils de personnalisation	Modalités d'affichage
	Signatures (pseudonymes, patronyme ...)	Choix de l'identité numérique
	Pages personnelles, photo, avatar, ...	Présentation de soi, arbitrage de visibilité de la vie privée
Lien social		
Discussion	Outils de discussion	Formes d'échange
	Réaction, chaînes de commentaires	Fils de discussion
	Messageries personnelles	Echanges privés internes du site
Partage	Outils de partage	Modalités de circulation
	Mails	Partage interpersonnel
	Facebook, Google +, Twitter	Affichage et réseaux interpersonnels
	Flux RSS	Diffusion

6.4 Démarche qualitative

Notre recherche nous offre une possible manière d'aborder les enjeux de l'open data locale par une interprétation des propos tenus par différents acteurs donc par une analyse de notre corpus de discours, ainsi qu'une observation des plateformes open data qui représentent notre corpus de dispositifs sociotechniques. Les enjeux particuliers de la collecte de données sur les acteurs impliqués dans l'ouverture des données publiques ainsi que sur la spécificité des dispositifs numériques révèlent plus de nature qualitative que quantitative. En effet, l'étude qualitative est mise à profit à différents moments d'analyse du corpus discours. Elle constitue la principale stratégie de recherche en vue de cerner le « pourquoi » des choses. Elle permet

de faire sortir les objectifs essentiels qui motivent les acteurs et d'expliquer leurs stratégies.

Cependant, nous nous appuyons de manière marginale sur l'approche quantitative afin d'exploiter les données chiffrées issues du recensement qui nous a servi de point de départ dans notre recherche. L'étude quantitative nous a permis de nous confronter à un grand échantillon (68 collectivités recensées) pour pouvoir établir une représentation approximative et une vision globale sur les tendances de l'open data. Nous avons créé des fiches de synthèses des collectivités qui ont ouvert leurs données sur le web en précisant : la collectivité, son échelle (région, département, commune), le collectif ou l'association open data (s'il existe), l'adresse de la plateforme, la date d'ouverture, le nombre de jeux de données ouvertes, les thématiques, les formats disponibles, la plateforme utilisée, les licences, etc. L'analyse de ces données nous permet de faire un état des lieux de l'open data en France, suivre son évolution, et découvrir les tendances par rapport aux thématiques des données ouvertes, le jeu de données le plus téléchargés, les formats et les licences choisies, etc.

Soulignons enfin que l'open data s'inscrit dans un environnement quantifiable qui est l'univers technique. Dans une perspective cybernétique, l'objectif de l'open data est la libération d'une plus grande quantité de données publiques. Cela se manifeste sur les sites open data qui survalorisent la quantité de données présentées comme si celle-ci est un facteur essentiel : « ce sont des changements quantitatifs (plus de numérique, plus de médias, plus de chaînes de télévision, plus d'informatique) qui vont provoquer un changement qualitatif (des relations plus épanouies, plus de libertés, une société meilleure) » (Breton , 2004, p.121). Plus les données seront nombreuses à être mises à disposition, plus la société pourra évoluer grâce à leur médiation. Tim Berners-Lee (2012) reprend, également, une vision quantitative des données en insistant sur la quantité des données existante pour les réutiliser et rendre le monde meilleur. En plus, le phénomène Big Data se caractérise en premier par la croissance exponentielle des volumes de données produites, stockées, traitées et échangées.

Cependant, dans notre étude, le volume des données ne constitue pas le seul indicateur d'analyse de l'ouverture des données. Pour cela

notre méthode s'oppose à la vision strictement quantitative de l'Open data, soit une approche qualitative plutôt que quantitative.

Conclusion du chapitre 6

Dans ce chapitre nous avons présenté les différentes méthodes de collecte et d'analyse propres à nos terrains de recherche. Nous tenons encore une fois à préciser que notre choix méthodologique est guidé par notre terrain organisationnel et géographique représenté par les cinq collectivités territoriales, ainsi que notre terrain technique lié à leurs plateformes Open data. La première étape a été consacrée au recensement des projets open data des collectivités territoriales en France et au classement des collectivités selon la phase d'implémentation du projet (réflexion, en cours, post-ouverture).

Si nous mettons méthodes au pluriel, c'est pour référer aux diverses méthodes employées dans cette étude :

1. la recherche et la veille documentaire en ligne sur les textes de lois, les blogs et les réseaux sociaux des acteurs ;
2. la participation aux conférences sur l'open data ou sur des sujets proches ;
3. la mise en place d'une enquête par entretiens semi-directifs menée auprès de 12 acteurs regroupés en quatre catégories : les chefs de projets open data dans les collectivités, les spécialistes de collecte, de traitement et de diffusion des données publiques, les utilisateurs de données publiques et les militants de l'open data locale ; et enfin
4. l'observation des plateformes open data des cinq collectivités.

Notre méthode d'analyse repose sur un « mix » des données collectées par ces différentes méthodes, mais il nous apparaît nécessaire de définir nos sources afin de guider aussi bien l'analyse que la progression de notre enquête de terrain. De plus, les différentes méthodes de recueil et les diverses sources doivent permettre l'articulation et la bonne cohérence de nos analyses. En effet, la troisième partie de cette recherche consacrée à l'étude de terrain se structure essentiellement autour des principales phases du processus d'ouverture que nous avons présentées dans l'introduction de cette partie. Notre grille d'analyse repose alors sur les cinq phases organisées autour de cinq grandes thématiques de recherche :

1. la réflexion ou la concertation des acteurs lors du montage du projet open data dans les collectivités,
2. la collecte de données,
3. la diffusion sur les plateformes,
4. la promotion de l'open data et enfin
5. l'usage.

Dans le tableau 11 ci-après, nous proposons de restituer les cinq phases, les méthodes de collecte de données, les sources, et les principaux objectifs. Cela nous permet de construire notre grille d'analyse et de définir les procédures de recherche essentiellement guidées par nos objectifs de recherche

Tableau 11. Notre grille d'analyse

Méthode(s)	Source(s)	Objectif(s)
I. Réflexion et concertation des acteurs		
<ul style="list-style-type: none"> - Entretiens semi-directifs - Veille et recherche documentaire 	<ul style="list-style-type: none"> - Corpus de discours des chefs de projets Open data dans les collectivités - Corpus de discours des spécialistes de collecte, de traitement et de diffusion des données publiques - Corpus de discours des militants de l'open data locale - Corpus documentaire en ligne 	<ul style="list-style-type: none"> - Etudier la dimension sociale qui englobe les motivations, les craintes et les interactions entre les acteurs pour le lancement du projet open data local
II. Identification et collecte de données		
<ul style="list-style-type: none"> - Entretiens semi-directifs - Veille et recherche documentaire 	<ul style="list-style-type: none"> - Corpus de discours des chefs de projets open data - Corpus de discours des réutilisateurs de données 	<ul style="list-style-type: none"> - Mettre en valeur et étudier les aspects originaux de la sélection et de la collecte de données afin de comprendre l'offre des collectivités en termes de données ouvertes.

	publiques - Corpus documentaire notamment les textes de la loi de 1978	
III. Diffusion et installation de la plateforme open data		
- Observation en ligne	- Corpus de dispositifs sociotechniques – les plateformes open data	- Questionner l’architecture et le contenu, les modes de participation et la configuration des liens sociaux sur les plateformes open data
IV. Promotion		
- Entretiens semi-directifs - Observation en ligne	- Corpus de discours des chefs de projets open data - Corpus de dispositifs sociotechniques – les plateformes open data	- Analyser la stratégie de communication et de marketing de la donnée publiques et du projet de l’open data au sein des collectivités étudiées.
V. Usage		
- Entretiens semi-directifs - Les conférences - Recherche et veille documentaire	- Corpus de discours des chefs de projets open data - Corpus de discours des réutilisateurs - Corpus documentaire	- Démontrer l’intérêt de l’open data pour les usagers - Définir le profil des réutilisateurs - Présenter les modèles de réutilisations

Chapitre 7. Lancement du projet : réflexion et concertation des acteurs

Cette première phase de réflexion et de concertation est essentielle afin de clarifier les attentes, les appréhensions et les craintes des parties concernées par le projet, ses opportunités, ses limites et inconvénients, ses objectifs à court et à long terme et le plan d'action. Il s'agit d'étudier les multiples facettes du projet avant de se soumettre au vote du conseil général de la collectivité. Les acteurs concernés ont discuté des multiples enjeux et aspects de l'ouverture. L'engagement dans un projet open data entraîne de grandes responsabilités. Une réflexion approfondie sur les avantages et les problèmes provoqués par l'ouverture serait d'une grande utilité. Dans ce chapitre, nous présentons la première étape de réflexion très cruciale car elle peut mettre en exergue certaines difficultés, comme le choix des données à ouvrir, le choix de licence, la constitution de l'équipe et les compétences qu'il est nécessaire de combler avant le démarrage du projet. La dimension sociale englobe les motivations, les objectifs plus larges et les interactions entre les acteurs afin de lancer le projet open data local dans leur environnement appropriés.

Afin de comprendre les raisons sous-jacentes à l'ouverture et étudier l'historique des initiatives dans les collectivités, donc en interne, nous revenons sur des éléments explicites qui émergent dans nos entretiens avec les chefs de projets open data des collectivités. Puis dans le but d'étudier la dimension sociale qui englobe les motivations et le rôle du mouvement associatif dans le lancement du projet open data local, nous reposons sur les propos des militants de l'open data locale mais également sur un corpus documentaire qui retrace les actions et les activités des associations dans plusieurs collectivités territoriales.

7.1 Des acteurs en mouvement

Dans le secteur public, la décentralisation des compétences des différentes collectivités territoriales définies par les lois de décentralisation de 1982 et 2002 a un double impact sur la politique locale. D'un côté, la question de l'ouverture des données était une initiative volontaire prise par les collectivités en l'absence d'une injonction juridique. L'attention de plus en plus soutenue à la valorisation des données publiques ouvertes s'est manifestée par le lancement volontaire de plusieurs portails open data local par les collectivités. Manifestement, cette volonté n'a pas pour origine une quelconque loi mais plutôt une motivation et une reconnaissance accrue des bienfaits de ce mouvement. Dans les faits, la législation actuelle repose sur une logique de demande d'accès aux données publiques sans obliger les collectivités à adopter la logique du mouvement Open data qui vise une diffusion proactive des données aux usagers, sans demande préalable. De l'autre, les collectivités ont acquis de nouvelles responsabilités et doivent posséder davantage l'esprit, le savoir faire, l'expertise et le financement pour gérer l'ouverture des données. Le choix apparaît comme une nouvelle épreuve pour les collectivités.

A cette étape nous distinguons deux catégories d'acteurs : des acteurs internes et des acteurs externes. Les acteurs internes sont les acteurs publics et les acteurs externes sont les associations et collectifs citoyens de la société civile. À cette étape préliminaire et incontournable du projet open data, quel sera le positionnement de chaque acteur impliqué ?

Nous rapportons le terme « mouvement » aux différents « acteurs en mouvement », qui militent pour que les données soient accessibles à tous, et ouvertes dans leur réutilisation. Les efforts délibérés des acteurs de l'open data sont à signaler.

7.1.1 Les acteurs internes : Élus et directeurs de services

L'acteur public est un acteur stratégique de l'open data locale. Il se positionne en tant qu'acteur dominant dans le système open data. En tant que source, producteur et diffuseur, la collectivité territoriale devient un porteur du projet open data définissant ses objectifs, sa

réalisation, ses règles, son calendrier et le budget consacré au projet. En interne, l'organigramme administratif distingue deux types d'acteurs : les élus locaux œuvrant au sommet stratégique de la hiérarchie administrative et les fonctionnaires des services à la base opérationnelle. Les acteurs internes partagent les rôles entre le niveau de décision (le pouvoir politique qui prend les décisions stratégiques et fixe les objectifs), le niveau d'exécution (le pouvoir de l'administration ou du gestionnaire qui prend les actions opérationnelles). Nous proposons la pyramide (figure 12) des acteurs internes où les élus occupent le sommet stratégique tandis que la base est formée largement des producteurs et détenteurs des données publiques locales, donc tous les services de la collectivité, et notamment des services responsables du projet open data dans la collectivité.

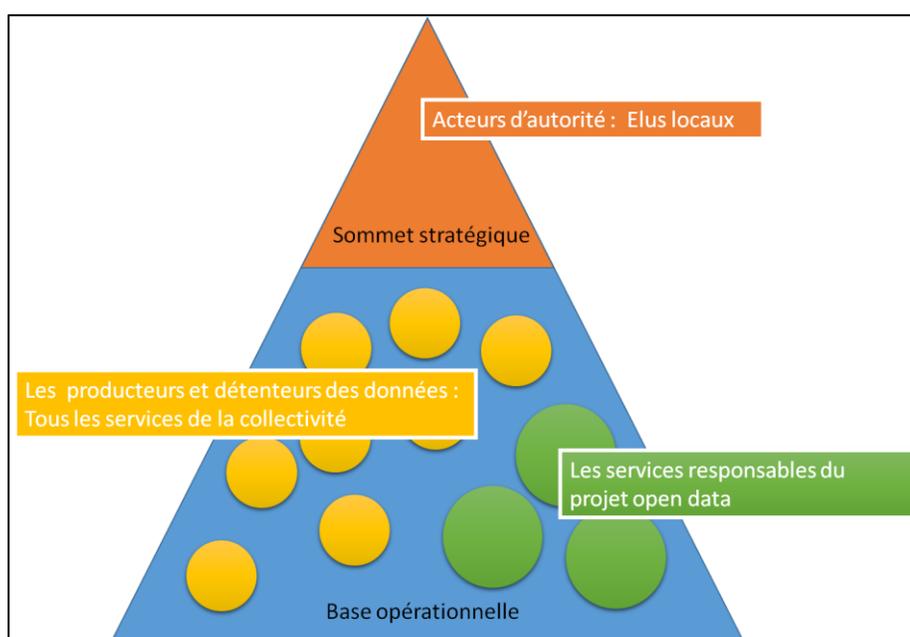


Figure 11. Les acteurs internes : Elus et fonctionnaires (Rahmé, 2015)

- Les acteurs d'autorité : les élus locaux représentent la haute sphère politique et les décideurs de la stratégie open data de la collectivité. Ils disposent du pouvoir d'adopter ou de rejeter un projet d'ouverture de données. Des décisions stratégiques relatives au fonctionnement d'ensemble du projet (la licence, la mutualisation, les partenariats) sont prises par le Conseil Général de la collectivité.

- Les acteurs opérationnels : ils regroupent tous les services de la collectivité que nous avons divisés en deux catégories : les services détenteurs des données et les services responsables du projet open data. Les premiers sont les producteurs et détenteurs des données et c'est au directeur de chaque service de sélectionner et faire le choix des données à ouvrir de son service, tandis que les acteurs de la deuxième catégorie de services sont chargés du pilotage et de la mise en œuvre du projet.

Si les entretiens ont révélé une perception commune de l'intérêt de l'open data pour les collectivités et qui correspond avec le discours général des avantages politiques, économiques et sociales de l'ouverture des données publiques présentées dans le chapitre 3, une série de facteurs apparaissent néanmoins dans leurs discours et qui influencent principalement la décision des élus pour adopter l'open data :

- Une volonté interne des acteurs de décision issue du climat international et national favorisant l'ouverture. C'est le cas du département du Saône et Loire. Initiée au printemps 2011, le projet open data en Saône et Loire est « né d'une volonté politique très forte du président du conseil général Arnaud Montebourg, actuellement ministre de l'Économie, du Redressement productif et du Numérique. Ce dernier était fortement inspiré par l'Open Government Initiative lancée par le président des États-Unis, Barack Obama », affirme Jean-Michel Martin¹⁷⁴, Directeur de la DSIIG qui pilote le projet Open data. La politique nationale de l'open data a naturellement un impact sur les politiques locales. Le comité d'éthique du projet open data du département de Saône-et-Loire a pris en considération la politique nationale en open data de manière à faire réagir sur une actualité riche avec notamment la feuille de route du gouvernement publiée fin février et qui réaffirme l'engagement de l'état dans l'open data.
- Une volonté interne des acteurs opérationnels influencée par le cadre européen de la directive INSPIRE. En parlant de la genèse

¹⁷⁴ Entretien le 30 Juin 2015

du projet open data au Grand Lyon, Laurence Martinez, indique que le projet a commencé dans un contexte européen marqué par la libération des données géographiques et environnementales pour permettre l'échange, le partage, l'accès et l'utilisation de données interopérables suite à la directive européenne INSPIRE du 14 mars 2007 établissant une infrastructure d'information géographique dans la Communauté européenne. En effet, la première version du portail open data lancée en Aout 2012, n'a diffusé que des données géographiques du service Information géographique du Grand Lyon.

- Une volonté interne des acteurs opérationnels en considérant que l'ouverture des données est compatible avec la stratégie et des projets prioritaires de la collectivité. Au Grand Toulouse, l'ouverture des données est cohérente par rapport à la stratégie de développement de la filière numérique. Open data Montpellier s'inscrit aussi dans une démarche plus globale de la Ville de Montpellier autour de l'innovation sociale et numérique dans le cadre du programme « Montpellier Territoire Numérique » lancé en 2010. Le conseil général de la ville de Bordeaux annonce que la démarche open data entre pleinement dans la dynamique du programme « Bordeaux Cité Digitales »¹⁷⁵. De même, le projet open data de la Région Provence-Alpes-Côte d'Azur (PACA) en 2011 répondait une volonté politique pour préparer Marseille-Provence à son futur statut de « Capitale européenne de la culture » annoncé pour 2013.

7.1.2 Les acteurs externes : le mouvement associatif

Pour mieux connaître le tissu associatif impliqué dans l'open data locale et mesurer son poids sur le territoire, nous avons identifié les acteurs locaux en essayant de définir leur rôle et le type d'activités qu'ils exercent.

¹⁷⁵ Voir : «Bordeaux, Cité Digitale : Votre Ville Résolument 'Open Data.'» *Open Data Ville de Bordeaux*. Accessed July 10, 2014. <http://opendata.bordeaux.fr/demarche>.

Au niveau local, le phénomène open data a conduit à la création de nombreux collectifs de citoyen militant à l'ouverture des données même si leur taille est plus modeste. Ils jouent un rôle équivalent mais avec un focus local, sauf l'association LiberTic qui est devenu un des membres fondateurs du de l'association Open Data France. La création de l'association Open data France, créée le 9 octobre 2013 à Toulouse, a pour but de regrouper et soutenir les collectivités engagées activement dans une démarche d'ouverture des données publiques et de favoriser toutes les démarches entreprises par ces collectivités dans le but de la promotion de l'open data »¹⁷⁶. Open data France rassemble un ensemble d'acteurs territoriaux afin de mutualiser leurs discours et coordonner leurs efforts. La Fing, l'association nantaise LiberTIC et Etalab sont des membres associés. Cet effort a pour but de regrouper et les collectivités engagées activement dans une démarche d'ouverture des données publiques et d'harmoniser et standardisés les démarches dans le but de la promotion de l'open data.

Nous avons repéré au total neuf associations et collectifs informels de citoyens : l'association nantaise Libertic, Open data 69 à la région Rhône Alpes¹⁷⁷, Open data Toulouse et Open data Bourgogne¹⁷⁸, Collectif Open Data Rennes, Collectif Elan pour un opendata à Poitiers¹⁷⁹, Ccollectif Saint-Brieuc Open Data¹⁸⁰, Collectif open data Strasbourg¹⁸¹, et collectif Open data Touraine #37. Ils partagent toutes la même mission : promouvoir l'open data locale.

Cette volonté se traduit par trois types d'activités :

a) Un moyen de pression

¹⁷⁶ "Opendata France". Accessed June 17, 2014. <http://opendatafrance.net/>.

¹⁷⁷ Open data 69. "Opendata69 : Collectif Pour L'ouverture Des Données Dans Le Rhône". Accessed July 4, 2014. <http://opendata69.org/>.

¹⁷⁸ "Open Data Bourgogne". Accessed July 4, 2014. <http://opendatabourgogne.org/>.

¹⁷⁹ "Open Data, Ouverture Des Données Publiques". Espace Mendès-France: culture médiation scientifiques, March 16, 2013. <http://emf.fr/14557/open-data-ouverture-des-donnees-publiques/>.

¹⁸⁰ "Opendata Saint-Brieuc". *Création Du Collectif Saint-Brieuc OpenData*. Accessed July 4, 2014. <http://opendata.la-matrice.org/>.

¹⁸¹ StrasWeb. "Open Data". Accessed July 4, 2014. <https://strasweb.fr/category/open-data/>.

- b) La vulgarisation et la promotion du concept d'Open Data
- c) La production d'applications à partir des données ouvertes

a) Un moyen de pression

Les activités de pression sur les conseils généraux des collectivités pour ouvrir leurs données se sont manifestées par la rédaction et la publication en ligne de lettres ouvertes adressées aux élus. En décembre 2010, l'association nantaise Libertic a lancé une lettre¹⁸² ouverte en faveur de l'ouverture des données publiques de Nantes. Le but de la lettre était, à l'époque, d'inciter la Communauté Urbaine de Nantes et la ville de Nantes à ouvrir explicitement leurs données publiques. Pour mesurer l'influence de la société civile sur les décideurs locaux, Claire Gallon affirme qu'au niveau local, « nous avons plus de visibilité sur l'impact de nos actions grâce aux échanges directs avec les décideurs »¹⁸³. Ainsi, la mobilisation d'acteurs et la lettre ouverte en faveur de l'ouverture des données de Nantes a pu jouer un rôle dans le développement d'une dynamique locale d'ouverture. L'interpellation de la Région Pays de Loire a également permis de lancer une réflexion qui s'est conclue par la validation du projet open data ». En effet, le portail open data mutualisé a été ouvert en Novembre 2011. Nous sommes dans une logique bottom-up, mais elle ajoute que la lettre n'a pu fonctionner que par ce que l'ouverture des données va également dans le sens du travail du Conseil Général de Loire-Atlantique sur la participation, le crowdsourcing, le développement de cartoparties.

Dans une démarche similaire, un collectif de citoyens, associations et entreprises de la région toulousaine a adressé, le 4 avril 2011, une

¹⁸² Voir : "Lettre Ouverte En Faveur de L'ouverture Des Données Publiques de Nantes," December 21, 2010. <http://libertic.wordpress.com/2010/12/21/petition-en-faveur-de-louverture-des-donnees-publiques-de-nantes/>.

¹⁸³ Interview de Claire Gallon de L'association Libertic". *Data Publica*, January 2, 2014. <http://www.data-publica.com/content/2012/11/lopen-data-est-tres-mal-estime-interview-de-claire-gallon-de-lassociation-libertic/>.

lettre¹⁸⁴ (inspirée de celle de LiberTIC), à M. Pierre Cohen, Président du Grand Toulouse pour s'engager dans l'open data. Le portail open data de Grand Toulouse et la ville de Toulouse était lancé le 22 Octobre 2012, lors du festival La Novela. David Chemouil, chercheur en informatique, et porteur de l'initiative Open data Toulouse¹⁸⁵ affirme que la lettre ouverte était un facteur stimulant à l'ouverture. Le 16 mai 2012, OpenData Touraine #37, le Collectif pour l'OpenData en Indre-et-Loire et Région Centre, a adressé la « Lettre ouverte aux élus d'Indre-et-Loire en faveur de la libération des données publiques »¹⁸⁶. Par contre, cette lettre n'a pas atteint son objectif prévu, à la différence de la lettre de collectif Open data Toulouse. Le dernier exemple est celui de la lettre ouverte et *la pétition en ligne* du Collectif open data Strasbourg (au sein de l'association StrasWeb) adressées en Juin 2012, à M. Roland Ries, maire de Strasbourg et M. Jacques Bigot, président de la Communauté Urbaine de Strasbourg pour les pousser à ouvrir leurs données. Pierre Rudloff, Président de l'association StrasWeb considère qu'ils ont réussi à convaincre les élus suite à l'annonce de Claude Graebing, vice-président en charge du numérique que la Communauté urbaine de Strasbourg allait commencer à ouvrir une partie de ses données. Par conséquence, la pétition en ligne était fermée à condition de « suivre de près ce projet afin d'assurer que la collectivité continue dans cette voie »¹⁸⁷. Le portail open data Strasbourg était lancé en Décembre 2012.

b) La vulgarisation et la promotion du concept d'Open Data

En plus du travail de lobbying, les associations ont assuré la pédagogie pour sensibiliser le public sur l'open data. Sami Rabih du

¹⁸⁴ Voir : "Lettre Au Grand Toulouse". *Opendata Toulouse*, October 22, 2011. <http://opendata-toulouse.numahel.net/doku.php/lettre>.

¹⁸⁵ Chemouil, David. "Opendata Toulouse". Accessed July 4, 2014. <http://opendata-toulouse.numahel.net/doku.php/>.

¹⁸⁶ OpenData Touraine #37. "Lettre Ouverte Aux Élus d'Indre-et-Loire En Faveur de La Libération Des Données Publiques". *Opendata37: Pour La Libération Des Données Publiques d'Indre-et-Loire (37)*. Accessed July 4, 2014. <http://opendata37.fr/>.

¹⁸⁷ Rudloff, Pierre. "Un Premier Pas Vers l'Open Data À Strasbourg". Accessed July 4, 2014. https://secure.avaaz.org/fr/petition/Ouverture_des_donnees_de_la_Communaute_Urbaine_de_Strasbourg/.

Collectif open data 69 indique que les membres du collectif créé en 2011 s'impliquent comme contributeurs et catalyseurs de l'open data¹⁸⁸. Ce collectif et d'autres de même nature proposent globalement les mêmes objectifs et visent pour les citoyens : comprendre les données publiques ouvertes, les manipuler et se les approprier, afin que tout un chacun connaisse mieux le territoire et ses enjeux et problématiques. Le collectif Open Data Rennes créé en Janvier 2012, un an et demi après l'ouverture du portail data.rennes-metropole.fr accompagne cette ouverture avec pour objectif de « fédérer la communauté de réutilisateurs de données et de tenter d'animer le territoire autour de l'open-data »¹⁸⁹. Le collectif rassemble des développeurs, journalistes, et aussi des habitants curieux de comprendre le fonctionnement de l'ouverture des données publiques. Les collectifs ont mis en œuvre différentes actions afin de sensibiliser les habitants à l'ouverture de données, aux travers de débats, conférences et ateliers. L'association libertic a créé et publié en ligne, en décembre 2011, un film très court sur l'open data. « Open data à la loupe » vulgarisant le concept de l'open data a été rapidement partagé sur les sites des collectivités et les autres associations. Le collectif « Élan pour un open data à Poitiers » a proposé un après-midi d'information et d'échanges sur l'intérêt citoyen de disposer des données publiques. Pour mettre en œuvre une dynamique autour de la visualisation, la manipulation et la production de ces données et de ce savoir, par tous et pour tous, des formations sont aussi offertes. Des membres du collectif Opendata Rennes ont animé un atelier sur la data-visualisation. Il avait pour objectif de faire découvrir le monde des données et les façons de représenter ces données qui nous entourent. LiberTic était co-organisateur de la première semaine européenne de l'Open Data. Par rapport au recensement des initiatives existantes et veille sur l'open data, LiberTic est l'association pionnier dans ce domaine. Le blog de l'association¹⁹⁰ et la carte open

¹⁸⁸ Entretien le 26 mars 2013.

¹⁸⁹ Collectif Open Data Rennes. "Nouveau Départ Pour l'Open Data À Rennes Métropole ?" *Collectif Open Data Rennes*, April 13, 2014. <http://blog.cod-rennes.fr/2014/04/13/nouveau-depart-pour-lopen-data-a-rennes-metropole/>.

¹⁹⁰ "Blog de L'association Libertic," accessed June 17, 2014, <http://libertic.wordpress.com/>.

data France sont deux outils de recensement et de veille sur les initiatives open data open France.

c) La production d'applications à partir des données ouvertes

Toutefois, outre les actions de pression, communication, sensibilisations, animations et formations autour des données ouvertes, nous identifions la réalisation d'applications web utilisant les données ouvertes, dans le but de donner des exemples d'applications. Même si la production de prototype de réutilisation semble être le but ultime d'un collectif selon Rabih, seulement le Collectif open data Strasbourg issu de l'association StrasWeb formée d'étudiants et de jeunes diplômés « ayant pour buts de favoriser l'insertion professionnelle des étudiants dans les métiers du Web et d'éduquer le grand public aux outils du Web » a développé deux applications web à partir des données ouvertes par la Compagnie des Transports Strasbourgeois : la première application montre la disponibilité de Vélhop (vélos partagés) à Strasbourg, et la deuxième application « Trouve ton campus » permet aux étudiants de l'Université de Strasbourg de visualiser les moyens de transports dont ils disposent pour se rendre sur leur campus.

Aux termes de ce parcours sur l'approche des acteurs à l'étape de lancement du projet, nous pouvons alors en retenir certains aspects. Les collectivités prévoient l'open data comme « opportunité » issue d'une conviction interne de ses impacts positifs sur la réalisation des objectifs stratégiques des collectivités. Par contre, les projets lancés pour répondre à une pression extérieure de la société civile s'apparentent selon nous comme une condition plus qu'une opportunité. Notons aussi l'absence des communautés de réutilisateurs dans la phase de pré-ouverture ce qui nous apparaît normal dans une logique d'offre que de demande.

7.2 Adaptation des acteurs et promotion de la culture de partage

Le projet open data a un caractère transversal. La transversalité est propre aux données, car les données sont présentes sur l'ensemble

des directions et des secteurs d'activités (économie, éducation, tourisme, santé, etc.) de la collectivité.

Le porteur du projet open data, élu ou fonctionnaire doit d'abord dégager ses motivations et les raisons de recourir à l'open data pour essayer de convaincre tous les acteurs de la nécessité de soutenir le projet. Selon Jérémie Valentin : « la première étape et la plus importante d'un projet open data consiste à convaincre les élus et les services municipaux de l'intérêt de la démarche »¹⁹¹. À ce stade, il est indispensable d'obtenir le soutien explicite des parties prenantes concernées par la bonne marche du projet, soit les élus, les directeurs de services et les fonctionnaires. L'objectif est de promouvoir la culture de partage, un facteur essentiel qui contribue à façonner l'ouverture des données. Tous les acteurs internes sont censés agir en développent des comportements et des pratiques de partage. Or l'établissement de cette culture de partage n'est pas simple car elle est confrontée à une longue tradition d'isolement et de fermeture qui n'encourage pas le partage et l'échange des données en dehors du cercle étroit de ses producteurs. Les données demeureront cachées limitant ainsi le potentiel d'ouverture. Comme la majorité des bureaucraties publiques, les données servent souvent à des tâches précises au sein de chaque service, plutôt qu'au développement d'une vision globale de l'organisation. Selon, Séverin Naudet (2011), ancien directeur de la mission Etalab¹⁹² « l'organisation classique de l'administration d'Etat se heurte en réalité trop souvent à celle de la société numérique, horizontale et décentralisée. Cette co-existence des modèles oblige l'Etat à repenser ses structures et ses missions, à faire évoluer les schémas précédents pour en inventer de nouveaux, afin de tendre vers une gouvernance mieux adaptée à notre époque ». Ainsi, même si une collectivité adopte l'open data, son implémentation est conditionnée par un processus d'adaptation car l'ouverture n'est pas composée d'un ensemble prédéterminé de règles et règlements statiques. Ce sont surtout des obstacles culturels qu'il

¹⁹¹ Entretien le 13 février 2013

¹⁹² Directeur de la mission Etalab depuis sa création en février 2011 jusqu'au Octobre 2012. Prédécesseur de Henri Verdier

convient de surmonter. Un changement culturel implique un processus de transformation lent et évolutif.

A ce stade, nous avons repéré trois facteurs de succès du projet :

- Plus la haute sphère politique est engagée et porte la parole sur l'open data, plus les chances de réussite du projet augmentent. Valentin souligne : « Le grand avantage est quand on se met dans une démarche open data en bénéficiant d'un support politique local fort. Nous avons de la chance que le projet soit attaché à un élu qui porte vraiment la thématique. Quand il n'y a pas d'élus sensibles à la thématique, il devient très difficile de faire avancer une démarche open data en interne ».
- Plus les porteurs du projet open data sont capables de défricher le sujet de l'open data, ses avantages et aussi ses difficultés, plus ils réussissent à convaincre les autres acteurs en proposant des pistes d'applications et des solutions aux problèmes en essayant d'apaiser les craintes et les incertitudes. Cela nécessite de créer un dossier sur le sujet suite à une étude approfondie et une veille sur les initiatives existantes en France et ailleurs.
- Plus le chef du projet open data est motivé et prend l'initiative de prendre contact avec tous les directeurs des autres services de la collectivité, plus la possibilité de la diversité des thématiques et le nombre des données à ouvrir augmentent. Un travail de pédagogie, sensibilisation, "évangélisation" en collaboration et approbation de ces collègues qui lui donnent leurs données. Cette tâche apparaît tout aussi difficile qu'importante surtout au sein de grandes collectivités qui regroupent des centaines de métiers et de services.

Cette adaptation peut exiger des investissements substantiels de la part de tous les acteurs et à tous les niveaux. Pour justifier l'ouverture des données et convaincre les autres acteurs, les arguments des porteurs du projet basés correspondent aux trois dimensions, politique, social et économique, que nous avons soulignées dans le chapitre 3. Les objectifs de l'ouverture des données : respecter ses engagements de transparence, créer des services sociaux, et valoriser le potentiel économique des données publiques.

En dehors de ces objectifs, le contexte national favorisant l'ouverture, la pression et la mobilisation de la société civile locale, ainsi que la présentation de multiples exemples de services et de d'applications produites à partir des données ouvertes aux États-Unis et en Grande Bretagne étaient des arguments et des facteurs puissants pour convaincre, en premier, les élus de voter à la libération des données et de pousser, en deuxième, les directeurs des services à collaborer et ouvrir leurs données.

7.3 Les risques de l'open data

Les collectivités locales engagées dans des projets open data sont conscients des avantages mais aussi des risques de l'open data et pourtant elles ont adopté une approche et une politique volontariste sans aucune obligation juridique. Xavier Crouan, Directeur de l'Information et de l'innovation numérique à la Ville de Rennes affirme lors d'une conférence du CapCom de Rennes en 2011 que : « Les nouveaux usages impliquent de nouveaux risques, mais c'est à l'institution de les prendre en charge. L'open data n'est pas un effet de mode. Il faut placer l'acteur public comme provocateur du risque, mais pas comme faiseur »¹⁹³.

Pendant la phase de réflexion, le discours standardisé et uniforme des chefs open data interviewés sur les avantages de l'open data est confronté par un discours plus « prudent » qui servait à démontrer les risques et les enjeux socio-politiques de l'ouverture. Moise MAYO¹⁹⁴, directeur de l'Insee à la région Bourgogne et membre du comité d'éthique du projet open data du département de Saône et Loire explique que l'objectif principal du comité d'éthique du projet open data est de réfléchir sur des problématiques notamment au regard du caractère «sensible» des données.

¹⁹³ Voir : Xavier, Crouan. "Libération Des Données Publiques : Faits, Enjeux et Perspectives." *Le Blog Du Modérateur*. Accessed February 20, 2014. <http://www.blogdumoderateur.com/liberation-des-donnees-publiques-faits-enjeux-et-perspectives-par-xavier-crouan/>.

¹⁹⁴ Entretien le 13 juin 2014.

Selon Sophie Tendeiro¹⁹⁵, le mouvement de l'open data est en construction et il est encore dans une phase expérimentale. Dans un contexte d'incertitude, les collectivités expriment leurs inquiétudes. En gestion du projet, l'équipe responsable de l'open data doit être audacieuse et créative dans ce genre de projet car il n'y a pas encore de modèles à suivre.

Des doutes subsistent quant aux effets secondaires et les conséquences possibles de l'ouverture. Les services ont peur de perdre le contrôle sur leurs données et ne peuvent pas prédire les possibilités de réutilisations. La crainte d'une mauvaise interprétation de données ou une mauvaise réutilisation est présente à l'esprit. En libérant les données, la puissance publique doit accepter la critique et prendre le risque. La qualité des données était une deuxième piste de réflexion qui méritait une auto-évaluation de la part des services. Si Bruguière (2004) fonde son engagement au droit des données publiques sur une vision politique, juridique plutôt idéaliste de la valeur des données produites dans le cadre de l'activité du service public¹⁹⁶, la préoccupation des chefs open data d'assurer l'ouverture de données publiques de qualité fait évoluer la conception des données publiques vers une attitude plus pragmatique et moins idéaliste. Les services sont en fait très sélectifs dans leurs choix de données à publier pour se rassurer que les données publiées ne contiennent pas d'erreurs. Selon Moise Mayo¹⁹⁷, la publication de fausses données ou des données non précises, de manière fréquente, pourraient mettre en doute l'efficacité de la collectivité et risque d'entraîner la perte de la confiance des citoyens. En plus, les chefs open data sont soucieux de ne pas ouvrir des données qui constituent une menace pour la vie privée des citoyens. La protection de la vie privée et l'anonymisation des jeux de données sont des conditions strictes à l'ouverture. Lorsqu'il s'agit de données personnelles, des observations sont faites sur la sensibilité de ce type de données en

¹⁹⁵ Entretien le 21 juillet 2014

¹⁹⁶ Pour l'auteur, la donnée publique est: « une donnée exhaustive, puisque recueillies pour tous, stable et complète, puisqu'il y a continuité de la collecte et du traitement par le service public et fiable en sens qu'il y a des sanctions attachées à la délivrance de fausses informations ».

¹⁹⁷ Entretien le 13 juin 2014

s'interrogeant sur le droit des citoyens et leur vie privé. Rappelons ici que la loi de 1978 protège la vie privée et interdit la communication de tout document qui porte atteinte à la protection de la vie privée. L'article 13 de la loi permet une réutilisation des données sauf si « la personne intéressée y a consenti, soit si l'autorité détentrice est en mesure de les rendre anonymes ou, à défaut d'anonymisation, si une disposition législative ou réglementaire le permet ». Malgré l'ouverture conforme à la loi et les mesures de sécurité mises en place par les chefs du projet open data afin que les données confidentielles des personnes restent protégées, la question de protection de la vie privée soulève des inquiétudes surtout par rapport au profilage de l'individu par le croisement de données¹⁹⁸.

Dans le tableau 12 nous synthétisons les avantages, les opportunités de l'open data ainsi que les solutions apportées face aux risques et craintes soulevés par les acteurs interviewés.

Tableau 12. Les avantages et les craintes de l'open data

Avantages / Solutions	Craintes
1. Montrer l'engagement de la collectivité dans la transparence. 2. Renforcer la confiance entre l'administration et les citoyens 3. Valoriser les données publiques 4. Profiter des commentaires des tiers, suite à l'ouverture pour corriger les données	5. Possibilité de trouver des erreurs 6. Questionner l'efficacité de la collectivité et risque de perdre la confiance des citoyens 7. Critique et mauvaise interprétation des données
8. La réutilisation pour créer des services sociaux pratiques. Ex. des applications web et mobiles	9. Ne pas prédire les possibilités de réutilisations Usage inadapté des données
10. Développement économique	11. Un mouvement en construction et les retours ne sont pas encore visibles.

¹⁹⁸ Voir à ce sujet le rapport de la CNIL: Quels Enjeux Pour La Protection Des Données Personnelles ?" In CNIL, 2013. <http://www.cnil.fr/les-themes/collectivites-locales/article/article/open-data-et-donnees-personnelles-lancement-dune-consultation-des-acteurs/>.

12. Anonymisation des données personnelles	13. La protection de la vie privée
--	------------------------------------

7.4 Le choix de licence

Les collectivités qui s'engagent dans d'ouverture des données sont confrontées au choix de la licence pour l'usage des données. L'adoption d'une « licence libre » est le choix logique pour accompagner une ouverture libre de données. Les données sont ouvertes par ce qu'ils sont sous licences ouvertes. Dans le fouillis¹⁹⁹ des licences²⁰⁰, la collectivité doit prendre une décision sur le droit et les formalités d'usage des données.

La licence (ODbL) donne la possibilité aux utilisateurs de partager, créer, et modifier et produire une création à partir d'une base de données sous réserve de mentionner la source de la base de données et la redistribuer sous les mêmes conditions imposées par la licence originale. Nous pouvons citer à titre d'exemple les données sous licence ODbL des collectivités locales comme Paris, Nantes, « Grand Toulouse Data, Communauté urbaine de Bordeaux. Le choix de cette licence est souvent justifié par les collectivités par le fait que les données publiques entrent dans le cadre des « biens communs ». Si la collectivité libère ses données, ce n'est pas pour que d'autres se les approprient.

La version 2 de la licence « Informations publiques gratuites » de l'APIE, qui impose de ne pas « altérer » les données, est aujourd'hui considérée comme obsolète²⁰¹. Le département du Saône et Loire et

¹⁹⁹ Voir : Blanc, Sabine. “Le Fouillis Des Licences Open Data S'éclaircit [Fiche Pratique].” *Lagazette.fr*, November 25, 2013. <http://www.lagazettedescommunes.com/208893/le-fouilli-des-licences-open-data-seclaircit-fiche-pratique/>.

²⁰⁰ Voir chapitre 2

²⁰¹ Voir à ce sujet : Ménégou, Karine. “Open Data : La CU de Bordeaux Change de Licence, Espérant Ainsi Générer Des Emplois - Lagazette.fr.” *Lagazette.fr*, January 25, 2013. <http://www.lagazettedescommunes.com/151375/open-data-la-cu-de-bordeaux-change-de-licence-esperant-ainsi-generer-des-emplois/#fn-151375-1>.

la communauté urbaine de bordeaux ont abandonné la licence APIE pour adopter respectivement la Licence Ouverte et la licence ODbL.

Durant l'été 2011, la mission Étalab, l'APIE, le COEPIA et les administrations des différents ministères dans la mesure ont été chargées par le Premier Ministre de travailler ensemble, sous la direction d'Étalab, sur la question d'une licence gratuite de réutilisation des données publiques. Cette période était consacrée au niveau national à la création d'une licence de type « licence Open data France ». La *Licence Ouverte* a vu le jour en Octobre 2011. Elle constitue un cadre pour la diffusion gratuite des données publiques sur la plateforme des données publiques de l'Etat « data.gouv.fr ». Cette licence assure une utilisation et exploitation gratuite, tout en veillant au strict respect de la paternité. Parmi les collectivités qui ont adopté la Licence Ouverte, nous citons le département du Saône et Loire, la ville de Montpellier, la ville de Strasbourg, et la ville de bordeaux. La Région Provence-Alpes-Côte d'Azur et ses partenaires ont fait le choix de plusieurs licences y compris la Licence Ouverte, Licence Creative Commons : CC-BY et la licence ODbL. Le Grand Lyon forme une exception en ce qui concerne le type de licences adoptées comme instruments de régulation de l'usage des données ouvertes par la collectivité. Nous trouvons trois licences différenciées :

- La Licence Ouverte qui couvre 95 % des données mises à disposition. Le but selon Laurence Martinez est « d'assurer la cohérence des services déployés avec les politiques publiques et garantir leur adéquation avec l'intérêt général » et « favoriser les réutilisations commerciales des données »

- La Licence engagée assure une mise à disposition gratuite des données après authentification des utilisateurs. La politique de Grand Lyon prend en compte que « les réutilisations de quelques données stratégiques soient compatibles avec les politiques publiques et l'intérêt général ».

- La Licence associée donne « possibilité de redevance à partir d'un seuil élevé du volume d'activité du produit ou du service créé afin d'éviter les situations de monopole ». L'objectif de cette licence est de

« favoriser le développement de nouveaux services et entreprises émergentes ainsi qu'un écosystème concurrentiel équitable ». Pourtant aucun frais est fixé afin d'éviter toute barrière à l'entrée pénalisante. Le Grand Lyon assure que cette licence ne s'applique pas pour les acteurs de la société civile et pour les PME innovantes.

Les deux dernières licences proposées par Grand Lyon n'entre pas dans le cadre juridique des licences libres attachées normalement à l'open data et qui défendent trois libertés :

- Accéder à la base de données gratuitement
- Modifier les données par le mixage avec des données d'autres sources
- Redistribuer la base de données dérivée.

Il faut signaler que si la base de données ou le jeu de données sont sous une licence libre, les produits dérivés de ces données ne sont pas nécessairement sous la même licence. Le même cas s'applique au logiciel libre qui permet l'accès gratuit au « source code » du logiciel, donc à la forme originale du logiciel. En fait, pour comprendre la différence, il consiste de séparer la couche de base qui comporte la base de données et le code source dans des formats réutilisables et modifiables de la couche superficielle d'usage plus commun (l'interface, l'application, le programme) d'un produit.

L'ensemble des licences libres offertes varient par rapport aux conditions attribuées à la réutilisation des données. Nous les classons à deux niveaux :

- Les licences les « plus libres » n'imposent aucune restriction pour la réutilisation des données. Il s'agit des licences CC0 et PDDL.
- Les autres licences sous certaines réserves :
 - l'attribution (citer le producteur des données). *Licence Ouverte* et ODC-BY tombent sous cette catégorie.
 - le partage à l'identique (la base de données dérivées suite à la modification sera ou pas sous la licence libre). La licence ODbL impose l'attribution et le partage à l'identique.

7.5 Constitution d'une équipe de projet

Dans les collectivités, le comité de pilotage du projet open data est constitué d'une équipe transversale dont les membres prendraient l'initiative de lancer le projet. Cette équipe de projet est souvent composée de chef du projet open data avec des responsables ou représentants de différents services comme le service des Systèmes d'Information, service de l'innovation et du numérique, service juridique, service de communication et, dans certains cas, d'autres services dont l'expertise était jugée indispensable.

Bien que le métier du chef de projet open data est encore neuf, des traits de caractère et des compétences spécifiques sont exigées de la personne qui occupe cette poste. Certaines compétences s'avèrent cruciales pour la réussite du projet.

Les chefs open data ont dégagé des traits de caractère ainsi que des compétences importantes quel que soit leurs parcours. Le responsable doit, par exemple, avoir des connaissances sur les problématiques et les pratiques de l'open data. Cela implique d'être au courant de toutes les questions qui agitent le secteur du numérique sur le territoire et d'assurer une veille stratégique, juridique et technique de l'Open data mais aussi sur les nouveaux usages numériques. Dans ce sens, plusieurs collectivités ont créé des sites web et des blogs de veille sur l'open data. @-brest est un site d'information édité par la ville de Brest sur « l'appropriation sociale des services, usages et technologies de l'information et de la communication »²⁰². Jérémie Valentin publie sur le blog du site opendata.montpelliernumerique.fr « La quinzaine des Données Ouvertes » en recensant les actualités de l'open data sur le web, Sandrin Mathon, chef Open data Toulouse utilise l'outil de veille scoopit pour diffuser et partager des informations sur l'open data. Selon Jérémie Valentin, il est important d'avoir un positionnement personnel et intellectuel sur le sujet de l'ouverture des données.

En plus, le caractère de transversalité pour l'implémentation du projet nécessite obligatoirement une bonne connaissance, par le chef open

²⁰² Brest. “@ Brest - Relier et Partager Autour Du Web.” Accessed June 7, 2014. <http://www.a-brest.net/>.

data, de l'organigramme général des services et des modes de fonctionnement des autres services de la collectivité. Un projet de ce type doit être suivi auprès de tous les services. Selon la description de poste du chef open data du département du Saône et Loire, parmi ses missions principales : « Animer des groupes d'acteurs internes et externes de la collectivité : directions du Conseil général, collectivités territoriales et intercommunalités, État, partenaires économiques et associatifs, organismes consulaires..., » et « Participer, en étroite liaison avec l'élu délégué aux instances et événements de l'Open data en France »²⁰³. De ce fait, des compétences communicationnelles et des capacités relationnelles et d'animation s'avèrent nécessaire pour créer des liens avec les autres acteurs internes mais aussi, avec les acteurs externes et les partenaires potentiels du projet.

²⁰³ Voir les détails de l'offre d'emploi du chef de projet Open data au Département de Saône-et-Loire. http://www.cg71.fr/le-departement/offres-demploi/detail-de-loffre-demploi/emploi/chef-de-projet-open-data-usa/listepointer/tx_eoffresdemploi_pi1%255BuidPage%255D%3D2%26cHash%3D28ec99955e9ed3ad8b94200f08d6e9d9/#.UzV4tYWF8jo.

Conclusion du chapitre 7

La question de l'ouverture des données reste une initiative volontaire prise par les collectivités en l'absence d'une injonction juridique. Une conviction interne des impacts positifs de l'ouverture sur le territoire devient une « opportunité » pour les collectivités porteuses du projet open data. Cependant, les projets lancés pour répondre à une pression extérieure de la société civile s'apparentent à une condition et non plus une opportunité.

En effet, la société civile représentée par les associations et collectifs citoyens locaux a contribué au mouvement open data en tant que moyen de pression, de vulgarisation et promotion du concept d'Open Data. Quelques associations ont également mené des activités de recensement des initiatives open data locale et de production d'applications à partir des données ouvertes. Dans tous les cas, un objectif commun pousse les acteurs internes et externes à coupler leurs expériences et expertises reliées à un besoin aux données publiques.

Même si une collectivité adopte l'open data, son implémentation est conditionnée par un processus d'adaptation des acteurs internes, élus et directeurs de service, à cause du caractère délicat et transversal du projet. Les collectivités locales engagées dans des projets open data sont conscients des avantages politiques, sociaux et économiques, mais aussi des risques de l'open data. Vu que le mouvement de l'open data est en construction et il est encore dans une phase expérimentale, un environnement d'incertitude entraîne la crainte de la part des conseils généraux de perdre le contrôle sur leurs données et ne peut pas prédire les possibilités de réutilisations ou de publier des données non exactes ce qui pourrait révéler l'inefficacité de la collectivité et entraîner une perte de confiance des citoyens. Enfin la protection de la vie privée et l'anonymisation des jeux des données sont des conditions strictes de l'ouverture.

Dans la phase de réflexion, les collectivités doivent prendre une décision sur le droit et les formalités d'usage des données. Dans le fouillis des licences, deux licences émergent : La licence Ouverte d'Etalab et la licence ODbL de l'OKF. La première assure utilisation et

exploitation gratuite, tout en assurant le strict respect de la paternité. La deuxième, moins libre, impose de mentionner la source de la base de données et la redistribuer sous les mêmes conditions imposées par la licence originale.

Bien que le métier du chef de projet open data est encore neuf, des traits de caractère et des compétences spécifiques sont exigés de la personne qui occupe cette poste. Certaines compétences s'avèrent cruciales pour la réussite du projet. Les chefs open data possèdent des traits de caractère ainsi que des compétences importantes quel que soit leurs parcours. Le responsable doit, par exemple, avoir des connaissances sur les problématiques et les pratiques de l'open data. Cela implique d'être au courant de toutes les questions qui agitent le secteur du numérique sur le territoire et d'assurer une veille stratégique, juridique et technique de l'Open data mais aussi sur les nouveaux usages numériques. Enfin, nous avons présenté les compétences spécifiques du chef du projet open data considérés comme cruciales pour la réussite du projet. Une bonne connaissance de l'organigramme général des services et des modes de fonctionnement des autres services de la collectivité ainsi que des compétences communicationnelles et des capacités relationnelles et d'animation s'avèrent nécessaire pour créer des liens avec les autres acteurs internes et externes. En plus, le chef doit avoir un certain degré de connaissances sur les technologies du web, applications mobiles et les bases de données.

Chapitre 8. Identification des données publiques : quelles données ouvrir ?

Dans ce chapitre, nous présentons le processus de collecte de données à publier en exposant les modalités de la production et de gestion des données publiques locales. Les collectivités territoriales proposent sur leurs portails une offre hétérogène de données. Le nombre, ainsi que les thématiques des jeux de données publiées varient d'une collectivité à l'autre. Ce chapitre relie ces variations de l'offre à une série de facteurs qui influencent le choix des données publiées et nous permet de mettre en valeur et d'étudier certains aspects originaux de la sélection et de la collecte de données. Les facteurs comprennent le choix stratégique des collectivités, la nature transversale des projets Open data, la sensibilité et la spécificité de la production et de la gestion des données publiques locales et enfin la doctrine d'ouverture. Le problème du choix des données fait également ressortir la doctrine fondamentale de l'open data qui est la diffusion proactive. Dans le cadre des projets étudiés, que deviennent la méthode de travail ? et quels obstacles ou facteurs affectent l'offre de données ?

De nombreux éléments développés lors de nos entretiens avec les chefs de projets open data et les réutilisateurs de données viennent en partie répondre et fournir des pistes de réflexions sur ces questions. De même, il était important de revenir au corpus documentaire notamment les textes de la loi de 1978 qui encadrent les fonctions et le travail des services.

8.1 Un choix stratégique

Quelques collectivités ont fait des choix stratégiques qui correspondent à leurs visions de l'Open data et à leurs agendas politiques en se focalisant sur une ouverture ciblée de leurs données. Plusieurs exemples méritent d'être rappelés car ils construisent un cadre de référence prégnant.

Au lancement du projet Open data de la Région Provence-Alpes-Côte d'Azur (PACA) en 2011, le choix de thématiques prioritaires répondait alors une volonté politique pour préparer Marseille-Provence à son futur statut de « Capitale européenne de la culture » annoncé pour 2013. L'Open data dans les thématiques - culture, transport et tourisme - constituait ainsi un élément important pour atteindre ce but, explique Sophie Tendeiro, Chef de projet Services numériques innovants et mutualisation des données publiques de la Région PACA²⁰⁴. La stratégie de promotion du territoire passe alors par celles de ses données.

Le département de Saône et Loire quant à lui, fait état d'un premier tri différent de celui de la région PACA. Il commence en effet par publier des jeux de données d'ordre financier relatives aux marchés publics, aux dépenses budgétaires, aux notes de frais des élus, ou encore aux subventions accordées aux associations dès l'ouverture du portail OpenData71 lancé en septembre 2011. Cette stratégie sert au département de Saône et Loire à mettre son action en cohérence avec son projet politique de transparence.

Enfin, dans sa version initiale lancée en août 2012, le projet SmartData de la Métropole de Lyon, était quant à lui, dédié aux données géographiques du service d'information géographique, Géomatique, porteur du projet au sein de la métropole. D'après Florence Martinez²⁰⁵, responsable de l'unité diffusion des données et géoservices de la métropole, les premières réflexions sur le projet Open data découlaient de la directive européenne INSPIRE²⁰⁶ du 14 mars 2007 visant à établir une infrastructure d'information

²⁰⁴ Intervention de Sophie Tendeiro sur la démarche partenariale de la région PACA lors du colloque « Open data : enjeux communicationnels et sociétaux ». Université Blaise Pascal. 7-8 novembre 2013.

²⁰⁵ Entretien le 20 septembre 2013.

²⁰⁶ Infrastructure for Spatial Information in the European Community

géographique dans l'Union Européenne pour favoriser la protection de l'environnement.

D'autres collectivités comme la Région Aquitaine ou le département de la Gironde ont également ouvert en premier lieu leurs données géographiques afin de se mettre en conformité avec les recommandations de cette directive.

De fait, l'Open data agit comme outil clé pour atteindre des objectifs plus larges : vérifier le statut « Capitale de la culture », faire preuve de la transparence financière ou protéger l'environnement. Dans ce cas, une ouverture thématique et synchronisée avec d'autres projets s'impose comme un choix stratégique majeur permettant aux collectivités de s'investir dans l'Open data et de gérer, à long terme, les opérations d'ouverture de leurs données publiques.

8.2 Un projet transversal

Au-delà des stratégies parfois orientées vers certains domaines économiques ou sectoriels, le choix des données à ouvrir se confronte à une réalité pratique de gestion de projet de nature transversale. Rappelons que ces données englobent toutes les données produites, collectées ou reçues par une autorité administrative dans le cadre de sa mission de service public. En effet, au sein de chaque collectivité, les services internes constituent les premiers producteurs de données publiques ; ainsi la répartition des tâches et la division du travail entre les directions et les services multiplient les producteurs et dispersent souvent les données. Le caractère de transversalité pour l'implémentation du projet nécessite donc obligatoirement une bonne connaissance, pour le responsable en charge de l'Open data, de l'organigramme général des services et des modes de fonctionnement de la collectivité. Il faut en premier lieu procéder à une analyse des activités et actions du service producteur ainsi que de la production de documents et données qui en découle. Ce point crucial est soulevé par Jérémie Valentin, lui-même chef de projet Open data pour la ville de Montpellier, qui décrit son travail comme « un vrai travail de

fourmi »²⁰⁷ quant à l'identification et collecte de données de différents services de la collectivité. Même dans les grandes collectivités, comme les communautés urbaines, les départements et les régions, le processus de collecte devient beaucoup plus compliqué.

La question d'identification et de traitement²⁰⁸ des données produites par les services est en fait un problème ancien qui a pris récemment des dimensions considérables lors de l'implémentation des projets Open data. En fait, en 2005, la transposition de la directive européenne, qui a modifié la loi du 17 juillet 1978 et a posé en France le principe de la réutilisation des informations du secteur public, a joué un rôle moteur en mettant en place l'obligation pour la collectivité territoriale, quelle que soit sa taille, de recenser les principaux documents publics en sa possession et soumis au droit d'accès et de réutilisation²⁰⁹. Par la suite, des mesures pratiques d'aide à l'élaboration de ce répertoire ont obligé les collectivités à désigner une « Personne Responsable de l'Accès aux Documents Administratifs », devenue une Prada. Selon le guide juridique et pratique sur les données publiques, la Prada doit « *identifier les données devant être mises à disposition en maintenant un dialogue constant avec les services et les élus, [...] coordonner la constitution du répertoire des informations publiques* »²¹⁰.

²⁰⁷ Intervention de Jérémie Valentin sur le projet Open Data de la ville de Montpellier lors du colloque « Open data : enjeux communicationnels et sociétaux ». Université Blaise Pascal. 7-8 novembre 2013.

²⁰⁸ Nous référons ici au traitement documentaire qui couvre des tâches allant de la réception du document à sa diffusion auprès des publics. Le processus de traitement de documents numériques se décompose souvent en traitement informatique (conversion de format), traitement technique (catalogage descriptif) et traitement intellectuel (indexation). Pour des études sur le traitement des documents numériques voir : Broudoux et Chartron, 2008. Voir également sur les questions de pérennisation des documents numériques : INRIA, 2006.

²⁰⁹ Le recensement ne prend pas en compte les documents exclus au droit d'accès et de réutilisation. Voir la liste des documents non communicables précisés par la loi de 1978 <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000339241>.

²¹⁰ Favre, Cédric, and Yann Mareschal. *Données Publiques : Guide Pratique et Juridique*. Bordeaux: Aquitaine Europe Communication, 2010. <http://www.aecom.org/Vous-informer/Juridique-TIC/Guides-juridiques/Les-donnees-publiques-decembre-2010>.

La constitution de ce répertoire de données devient alors une nécessité pour organiser le processus de diffusion car il ne s'agissait pas d'une simple liste de documents numériques, mais bien d'une description et d'une contextualisation de la production de ces documents pour faciliter son accessibilité et sa réutilisation. Les documents doivent donc être pourvus des métadonnées nécessaires à cette diffusion. D'après Dhérent²¹¹,

Les métadonnées concernent l'identification du document et de ses différentes versions, sa description et son contexte de création, son classement au sein d'une arborescence pérenne, la gestion de son cycle de vie, la gestion à long terme de son accessibilité. C'est également la mise en relation et en perspective des documents et données avec d'autres documents ou données qui donne ou ajoute à chacun d'entre eux du sens.

Selon l'article 36 du décret n° 2005-1755 du 30 décembre 2005, le répertoire publié en ligne précise pour chacun des documents recensés, la liste de métadonnées essentielles : « *son titre exact, son objet, la date de sa création, les conditions de sa réutilisation et, le cas échéant, la date et l'objet de ses mises à jour* ».

Si l'importance de l'identification et du repérage des données publiques paraît évidente depuis une dizaine d'années, son implémentation semble loin d'être complète. En pratique, plusieurs collectivités n'ont pas encore assigné de Prada et ne produisent pas un répertoire de leurs documents bien que la création d'un tel répertoire s'avère être une base solide et incontournable pour recenser les données produites au sein d'une collectivité. Son absence complique le processus d'ouverture et réduit l'offre de données sur les portails Open data qui ne repèrent pas tous les documents produits au sein de la collectivité, mais seulement ceux répertoriés au terme de la procédure de diffusion. Charles Népote, Chef de projet "Partage des données publiques" à la FING²¹², insiste sur le fait que l'ouverture des

²¹¹ Institut National de recherche en informatique et en automatique. *Pérenniser Le Document Numérique*. Paris: ADDBS, 2006. p.9

²¹² L'association Fondation Internet Nouvelle Génération.

données publiques sur les portails Open data reste partielle puisque les utilisateurs ne peuvent avoir connaissance de toutes les données publiques produites et communicables²¹³.

Tout cela concourt à faire de l'ouverture un processus difficile qui révèle la complexité de la production et de la gestion des données publiques locales.

8.3 Externalisation des données publiques

La nature transversale des projets Open data pose la question des liens et conflits d'intérêts entre le service responsable de la diffusion et les services détenteurs et producteurs de données publiques. En effet, les services des collectivités ne peuvent pas faire abstraction de la finalité de la production de données publiques. Malgré la volonté politique claire et la mise en œuvre d'une véritable externalisation, les données publiques apparaissent avant tout comme une matière première destinée à un usage interne. Nous l'avons vu, au sens de la loi du 17 juillet 1978, la finalité première des documents administratifs vise leur utilisation par les administrations elles-mêmes. Malgré l'interprétation relativement large de la notion de réutilisation, cette dernière recouvre en premier lieu « *l'utilisation des documents administratifs par des tiers à d'autres fins que celles de la mission de service public en vue de laquelle les documents ont été élaborés ou sont détenus* ». Dans ces conditions, toute réutilisation est une utilisation du contenu en dehors de son intention initiale.

Par conséquent, l'Open data marque une mutation dans les pratiques des détenteurs des données publiques suite à l'évolution des modes de communication de l'accès sur demande à la diffusion proactive²¹⁴, et ils seront particulièrement traités au quatrième chapitre. Le partage des données avec des tiers à des fins de réutilisation déroge aux pratiques habituelles des fonctionnaires dans le cadre de leurs activités professionnelles. Ainsi, des freins majeurs pèsent sur l'ouverture, d'une part, le système traditionnel administratif limite la

²¹³ Propos tenus par Charles Népote sur le partage des données publiques lors du séminaire Modra du 31 janvier 2014.

²¹⁴ Voir chapitre 5.

mission de diffusion aux seules institutions publiques qui ont la vocation d'exercer une activité éditoriale, et d'autre part la législation actuelle repose sur une logique de demande d'accès aux données publiques sans aucune obligation faite aux collectivités d'adopter une diffusion proactive. Ainsi, le concept d'ouverture n'est-il appréhendé que de manière très parcellaire dans les multiples services administratifs. Chaque entité applique ses propres régulations en termes de choix de données à ouvrir. Faute, jusqu'à présent, d'une culture d'ouverture orientée vers la satisfaction des utilisateurs en termes de données publiques, les chefs de services des collectivités territoriales accordent très souvent la primauté au fonctionnement interne de l'administration. Les collectivités engagées dans l'Open data n'ont pas pu empêcher les contournements de la loi de se perpétuer, preuve de la faiblesse des régulations au sujet de l'Open data. En effet, la loi du 17 juillet 1978 précise que « *la communication peut être refusée s'agissant de documents à confectionner, par le biais d'un traitement quelconque, de recherches, d'une synthèse, d'une analyse* ». Dans ce cas, le processus d'ouverture des données dans une collectivité ne doit, par conséquent, ni perturber ni surcharger le fonctionnement des services d'une collectivité quant aux documents nécessitant un traitement quelconque avant d'être publiés. En outre, la condition de *l'existence du document en l'état* décrite par la Commission d'accès aux documents administratifs est un facteur essentiel limitant l'ouverture. L'existence du document en l'état signifie que le document doit exister en son état actuel ou pouvoir être obtenu par un traitement automatisé d'usage courant. En d'autres mots, l'ouverture de certaines données peut-être refusée si elle nécessite un travail supplémentaire des services pour traiter les données en vue de l'ouverture.

Si l'Open data dépend fortement de l'externalisation des données publiques, il nous semble qu'une réforme de la loi sur l'accès et la réutilisation des documents publics devrait non-seulement obliger les collectivités à ouvrir leurs données²¹⁵, mais aussi à organiser les nouvelles tâches de diffusion afin de garantir une forme de

²¹⁵ Signalons que dans le projet de loi de décentralisation transmis au Conseil d'État en mars 2013, les collectivités locales [de plus de 3500 habitants et leurs EPCI] auront l'obligation d'ouvrir leurs données publiques.

coopération entre le service responsable de la diffusion et les services producteurs de données. Au-delà, l'Open data repose selon-nous sur la culture de partage, facteur essentiel qui contribuera à façonner l'ouverture des données. Ainsi, une sensibilisation apparaît nécessaire afin de convaincre les services des collectivités territoriales de l'intérêt de l'Open data. Tous les acteurs internes sont censés agir en développant de nouveaux comportements et des pratiques de partage. Ce sont surtout des obstacles liés à la culture professionnelle et administrative qu'il convient de surmonter et tout changement culturel représente plutôt un processus de restructuration lent et évolutif.

8.4 La diversité des producteurs de données sur un territoire

La loi de décentralisation a défini les compétences des différentes collectivités territoriales. Ainsi, les services des collectivités produisent des données seulement dans le cadre de leurs compétences. Le tableau de répartition des compétences entre les différents niveaux des collectivités et l'Etat confirme la complexité de la production des données publiques locales et montre que plusieurs producteurs interviennent à plusieurs niveaux dans le même secteur²¹⁶.

Face à ce défi, la collectivité territoriale ne peut plus être seule à produire les données sur son territoire. Dans une deuxième étape du processus de collecte de données, la collectivité s'oriente donc vers d'autres producteurs extérieurs de données en relation avec son territoire. Cette mise en réseau indispensable des producteurs de données, hétérogènes, portés par les projets Open data lancés par les collectivités favorise les partenariats, et agit comme un catalyseur de l'ouverture sur un territoire. Plusieurs éléments contribuent à créer une base commune de données ouvertes gérées par la collectivité: l'existence d'une plateforme préétablie²¹⁷ et officielle, l'échange de données entre les producteurs fondé sur la reconnaissance mutuelle

²¹⁶ Voir la répartition des compétences des collectivités locales sur le site de la Direction Générale des Collectivités Locales : http://www.collectivites-locales.gouv.fr/files/files/tableau_repartition_compétences_valideCILapublier%281%29.pdf.

²¹⁷ La collectivité offre à ses partenaires la possibilité de diffuser leurs données via son portail Open data

de la fiabilité de la source et de la qualité des données, l'identification de facteurs de progrès pour l'ensemble des acteurs en vue du renforcement des réutilisations possibles. Ainsi, la création d'un portail Open data assure par conséquent très souvent, un rôle fédérateur au sein d'un territoire, en rassemblant plusieurs producteurs autour d'un noyau de données à réutiliser. C'est le cas de Lyon Métropole qui cherche à faire venir et à associer d'autres acteurs dans le processus d'ouverture. Par exemple, elle s'associe au Sitra (Système d'Information Touristique de Rhône-Alpes) dans le but de fédérer des données touristiques sur les hébergements, les restaurants, le patrimoine culturel et naturel, les équipements, les commerces, les associations touristiques de la région Rhône-Alpes et l'Aéroport de Lyon, ouvrant la diffusion de données dynamiques en temps réel sur le trafic aérien, et cette question sera détaillée dans le chapitre 4. De même, dans une démarche partenariale, la région PACA a créé des partenariats avec les communes pour créer un portail partenarial régional.

Les chefs de projets open data interviewés mettent en avant leur rôle de fédérateur par le fait de réunir des données publiques produites par des acteurs aux profils différents - des établissements publics, d'autres collectivités locales, des entreprises, et même des citoyens - dans le cadre local et national.

8.4.1 La centralisation de la production et de la diffusion de données publiques au niveau de l'État

Nous pouvons mettre en exergue à la suite de Mazon²¹⁸ et Gorron²¹⁹ qu'une grande partie des données publiques locales est produite et gérée au niveau national par les ministères, les offices et instituts nationaux et autres administrations de premier rang. La dimension territoriale des données est omniprésente puisque la majorité des données publiques sont associées à des données géographiques qui les inscrivent à un territoire défini. A titre d'exemple, l'Institut National de la Statistique et des Études Économiques (Insee) dispose de données sur les collectivités. Pour le recensement national de la population, la loi

²¹⁸ Entretien le 19 février 2014.

²¹⁹ Entretien le 13 décembre 2012.

clarifie la répartition des tâches entre l'Insee et les communes et instaure de fait un partenariat plus étroit pour la mise en œuvre des enquêtes. Celles-ci sont préparées et réalisées par les communes, mais la collecte des informations est organisée et contrôlée au niveau national par l'Insee. Autre exemple, l'Office National de l'Eau et des Milieux Aquatiques (ONEMA), structure publique chargée de tenir à jour le Système d'Information sur les Services Publics d'Eau et d'Assainissement (SISPEA), permet également d'accéder gratuitement en ligne au descriptif des services d'eau d'une commune : mode de gestion, mission, éléments sur les indicateurs de performances, de prix et de qualité de l'eau. Les collectivités territoriales participent à la collecte et à la création de la base de données SISPEA par la saisie directe des données sur leur territoire. L'Institut National de l'Information Géographique et Forestière (IGN), quant à lui, est responsable de la production, de l'entretien et de la diffusion de l'information géographique de référence en France.

Ces trois exemples montrent la centralisation de la production et de la diffusion des données publiques en matière du territoire local opérées par les administrations nationales, même si dans certains cas les collectivités locales sont impliquées en tant que contributeurs dans la collecte de données concernant leurs territoires. En effet, le système de publication des données publiques se fonde principalement sur les activités éditoriales de quelques institutions publiques de premier rang, dont l'Insee et l'IGN, qui appartiennent à la catégorie des « éditeurs publics institutionnels » ayant pour vocation d'exercer cette activité éditoriale.

De même, il faut souligner que les stratégies et positions divergent au sein de ces éditeurs publics de grande ampleur car les conditions de diffusions sont très réglementées de sorte que la diffusion peut être directe ou indirecte, payante et/ou gratuite. Depuis 2003, l'Insee a commencé à publier sur son site web des données accessibles gratuitement alors que l'IGN montre toujours une résistance à l'Open data malgré sa nouvelle « politique d'adaptation » comme la décrit François Perrussel-Morin²²⁰, directeur interrégional centre-est de l'IGN. L'IGN a publié en ligne, sous licence ouverte, des données

²²⁰ Intervention de François Perrussel-Morin sur l'évolution de l'offre de service public IGN lors du séminaire Modra du 27 juin 2014.

géographiques moins détaillées depuis 2009 et a permis un accès gratuit aux organismes d'enseignement et de recherche fondamentale. Deux ans plus tard, il ouvrait gratuitement ses données aux organismes chargés d'une mission de service public en dehors des usages industriels et commerciaux.

Au vu des différents producteurs nationaux présentés ci-dessus, au-delà de leurs modes de publication des données publiques locales, nous mettons en exergue la dispersion et la fragmentation des données qui, selon nous, révèlent un problème d'accessibilité aux données publiques. Les initiatives nationales de mutualisation témoignent de l'engagement de l'État en vue de centraliser et unifier l'accès, la diffusion et la recherche de données publiques, par exemple data.gouv.fr permet aux collectivités de publier leurs données locales si elles le désirent. Une autre position plus intéressante à souligner est celle des collectivités qui collectent des données ouvertes par des services d'Etat pour les publier sur leurs propres portails Open data. Dans ce cas, le processus de collecte exige une extraction de fragments de données pour créer un nouveau jeu de données plus spécifique. Par exemple, la ville de Montpellier et le département de Saône et Loire font ainsi extraire des données publiées par l'Insee en relation avec leur territoire respectif afin de proposer une offre plus variée de données sur leurs portails Open data. Ainsi, afin de faire face à la dispersion des données locales produites et diffusées parfois par des institutions nationales, la stratégie Open data d'une collectivité repose en partie sur une stratégie de veille pour identifier et collecter les données externes qui la concerne.

8.4.2 Les entreprises chargées d'un service public

Dans le cadre du partenariat entre le secteur public et le secteur privé, les organismes privés chargés d'une mission de service public dans des domaines comme l'eau et l'assainissement, le transport public, le traitement des déchets et l'énergie, sont des producteurs et des détenteurs de données publiques, et ceci dans leurs champs d'activité respectifs. Ce sont des tiers avec qui l'administration a conclu des accords. Les règles du droit d'auteur excluent du champ de la loi sur la réutilisation des données publiques, les données sur

lesquelles des tiers ont des droits de propriété intellectuelle. Florence Martinez, responsable de l'unité de diffusion des données et géoservices de la Métropole de Lyon, souligne le fait que la collectivité territoriale ne peut pas céder des droits qu'elle ne possède pas en raison des dispositions particulières de ces accords²²¹. En effet, les données à caractère industriel et commercial, produites dans l'exercice d'une mission de Service Public à caractère Industriel et Commercial (SPIC), sont exclues du droit à réutilisation par la loi de 1978. Les contrats actuels entre les collectivités territoriales et le secteur privé ne prennent pas en compte la question de l'Open data. D'où la remarque du Conseil National du Numérique²²² qui a proposé de réformer la loi en élargissant le périmètre des informations publiques pour inclure le droit de réutilisation aux des données brutes des SPIC. Pour le moment, cette restriction représente un obstacle qu'il convient de surmonter en créant une réelle synergie entre collectivités et opérateurs privés qui prend en compte l'ouverture des données. La Métropole de Lyon offre une bonne illustration de ce type de démarche. L'équipe de l'Open data a réussi à créer deux partenariats différents avec deux entreprises du secteur des transports. Cette étape fait partie des objectifs majeurs de la métropole et qui vise à libérer le maximum de données sur le transport. Le Groupe Bolloré responsable du service des véhicules électriques, libère quant à lui, des données en temps réel sur la disponibilité des voitures, alors que la société Keolis libère des données sur le réseau de *Transports en commun lyonnais* (TCL) tels que les arrêts de bus, les lignes de bus, les parcs relai et les entrées/sorties des stations de métro.

8.4.3 Le crowdsourcing

L'Open data est souvent lié au modèle de crowdsourcing. Comme présenté dans le chapitre 2, ce dernier permet à des organisations de toute nature d'externaliser vers la foule des activités réalisées

²²¹ Entretien le 20 septembre 2013

²²² Conseil National du Numérique. "L'ouverture Des Données Publiques : Un Enjeu Majeur Qui Doit Etre Protégé Par Une Volonté Politique Déterminée," 2012. http://www.cnnumerique.fr/wp-content/uploads/2012/06/12-06-05_CP_OpenData.pdf.

traditionnellement en interne et représente une forme de production alternative aux autres formes de production en interne. En matière d'ouverture des données publiques, les projets Open data s'avèrent être une opportunité pour introduire l'intérêt du crowdsourcing dans le secteur public. Il s'agit d'une alternative de production qui repose sur la contribution des citoyens à la démarche d'ouverture et de mutualisation des données publiques d'une collectivité locale. La collaboration entre la ville de Montpellier et la communauté locale d'OpenStreetMap (OSM) illustre bien ce type d'expérience. En effet, le projet mondial communautaire de cartographie OSM représente une application concrète du crowdsourcing dans le secteur public, puisqu'il permet une production alternative, par les citoyens, aux autres formes de production assurées normalement par des services gouvernementaux en interne. Aujourd'hui, la base de données collaborative OSM est considérée comme une source privilégiée de données géographiques ouvertes sur le web et qui regroupe plus d'un million de contributeurs, comme l'indique Florian Farge, contributeur actif à OSM France²²³.

Les formes de collaboration entre la ville de Montpellier et la communauté locale d'OSM, ont porté, en premier lieu, sur la publication sur le portail Open data de la ville des données extraites de la base de données d'OSM. La géo-localisation des distributeurs de billets, des boîtes aux lettres, des stations-services, des pharmacies, des points de collecte des déchets, des téléphones publics, sont autant d'exemples de jeux de données extraits de la base de données d'OSM et publiés sur le portail Open data de Montpellier. Un autre témoignage de cette forme de collaboration apparaît dans le bilan de Jérémie Valentin, chef du projet Open data de la ville. Au-delà d'une simple extraction de données, la collectivité a créé une synergie avec la communauté d'OSM sur le territoire montpelliérain. Plantin et Valentin soulignent le rôle de la communauté locale d'OSM dans l'animation des ressources Open data :

« Dans le cas de Montpellier, il s'avère que la communauté d'OSM est très active dans la réutilisation des données publiques : les cartes qu'elle réalise constituent autant de

²²³ Entretien le 21 mars 2013.

« vitrines » et permettent dès lors de sensibiliser à la récolte citoyenne d'informations sur le territoire, qui peut trouver dans OSM un lieu de stockage structuré et réutilisable. Enfin, OSM peut servir de porte d'entrée dans l'univers du libre, autant d'un point de vue des logiciels que des données. OSM a alors un rôle de média qui sert à faire naître une culture de la donnée et du libre auprès des citoyens »²²⁴.

De plus, en partenariat avec la communauté locale des contributeurs d'OSM et des associations locales, la ville a organisé des "cartoparties"²²⁵ afin de pallier certains manques de données géographiques sur le territoire.

« Par exemple, sur la thématique du handicap, les collectivités ne renseignent que les lieux publics et la qualité du cheminement trottoir. Pour un réutilisateur désireux de créer un service destiné aux *Personnes à Mobilité Réduite* (PMR), ces informations sont certes importantes, mais ne suffisent pas à créer un service adapté à tout le territoire concerné. [...] Afin de compléter ces informations sur son territoire et de susciter la création d'un service destiné aux PMR, la ville de Montpellier, associée à des acteurs locaux du monde du handicap, de la communauté d'*OpenStreetMap* et du monde du libre, a tiré profit du fonctionnement contributif d'OSM à travers des collectes citoyennes d'informations, qui seront ensuite stockées dans la base de données afin qu'elles soient libres de droit »²²⁶.

²²⁴ Plantin, Jean-Christophe, and Jérémie Valentin. "Données ouvertes et cartographie libre." *Les Cahiers du numérique* Vol. 9, no. 1 (April 1, 2013): 85–107. http://www.cairn.info/article.php?ID_ARTICLE=LCN_091_0085.

²²⁵ D'après le site Open Data de la ville de Montpellier, la cartopartie, aussi appelée "mapping party" est « un événement permettant de cartographier en collaboration avec des volontaires, des éléments d'un quartier ou d'une ville. Le but d'une cartographie contributive (ou moissonnage de données) est d'arriver à produire une carte complète d'une zone plus ou moins grande, en récoltant le maximum d'informations sur le terrain. Pour se faire, on invite des pratiquants volontaires de tous niveaux à se réunir lors d'un après-midi afin d'arpenter à pied, à vélo, en fauteuil, en poussettes les rues d'un quartier ». <http://opendata.montpelliernumerique.fr/Cartographie-contributive,95>, consulté le 6 Mai 2015.

²²⁶ Plantin et Valentin. Op. cit.

Une telle expérience de production par les citoyens est considérée comme complémentaire au projet Open data mais elle constitue, de notre point de vue, un facteur de succès important qui montre le degré d'ouverture de la collectivité d'une part, et l'engagement de la population de l'autre. La participation des citoyens ne doit pas, pour autant, être surestimée quant à leur capacité à produire des données sur le territoire. D'après l'expérience de la ville de Montpellier, la participation des citoyens est orientée, voire supervisée, quant aux choix de données citoyennes à publier sur le portail Open data ou les thématiques à traiter.

Conclusion du chapitre 8

La question d'identification et de collecte des données démontrent la complexité de la production et de la gestion des données publiques locales. Dans ce chapitre, nous avons présenté les facteurs influant sur le processus de collecte et le choix des données à ouvrir sur les portails Open data des collectivités territoriales.

Au-delà des stratégies parfois orientées vers certaines thématiques, le projet Open data dans une collectivité est un processus complexe car il met en jeu l'ensemble des producteurs sur le plan interne, local et national. Les défis liés à l'ouverture sont donc de taille, puisque l'identification et la collecte des données dans un système multi-acteurs compliqué passe en effet par l'expérimentation et par une mobilisation accompagnée.

Outre les mesures visant à comprendre le mode de fonctionnement et de production internes des données, les cadres décisionnels des collectivités engagées dans l'Open data doivent produire des efforts pour promouvoir une culture d'ouverture et favoriser la libération des données initialement destinées à un usage interne. Il leur faut effectuer une campagne de sensibilisation au sein des services afin d'éliminer une culture de fermeture et substituer un état d'esprit ouvert à libérer leurs données.

L'offre hétérogène de données sur un portail Open data repose sur son propre rôle fédérateur rassemblant l'ensemble des producteurs autour d'un noyau de données. C'est de cette manière qu'un processus d'ouverture dépasse le stade de collecte interne et parvient à créer un écosystème d'acteurs autour des projets.

Malgré les efforts accomplis, les défis managérial et culturel de l'Administration limitent et retardent le processus d'ouverture. En effet, les militants en faveur de l'Open data exigent une mobilisation rapide et efficace de la part des institutions publiques. En fait, l'Open Knowledge Foundation (2012), dans son manuel sur l'Open data, incite les producteurs de données à les ouvrir le plus rapidement possible. D'ailleurs, les collectivités soucieuses de la compétitivité de leur territoire ont déjà envisagé une démarche rapide pour refléter leur volonté d'être en phase avec les dernières tendances. Il s'agit d'une reconnaissance dont ils espèrent bénéficier grâce au lancement d'un portail Open data.

Pour réduire le délai de l'ouverture, quelques jeux de données suffisaient à lancer une plateforme Open data le plus vite possible. Cependant, le nombre limité des jeux de données publiés sur les portails Open data ne résulte pas seulement d'une stratégie d'ouverture accélérée, mais dépend fortement des conditions de production des données et du processus de collecte présentés ci-dessus. Les collectivités publient d'abord les données les plus faciles à collecter car la collecte s'opère progressivement et la construction d'un répertoire de données ne se fait pas préalablement au lancement du portail, mais au fil de l'eau. Ainsi, il n'y a ni élaboration *a priori* de critères de sélection des données, ni processus formel et structuré de mise en forme d'une collecte stratégique même si les chefs de projet Open data à l'œuvre croient en l'émergence lente - mais progressive - sur un principe d'imitation qui suscite finalement la participation à l'ouverture. En outre, le principe du libre choix reste un élément important qui s'impose même si, selon nous, il est un frein à l'ouverture. Basé sur le fait que la communication des données ne doit pas perturber le fonctionnement de l'administration, la culture de partage reste une option, voire un choix éthique.

La mise en œuvre de l'identification et de la collecte des données, dans les conditions spécifiques que nous venons de décrire, est éclectique et instable, elle ne construit pas un processus solide de l'Open data. Malgré l'urgence suggérée par le contexte politique et sociétal, l'ouverture et le choix des données sont généralement déterminés par des considérations pragmatiques qui imposent une ouverture lente mais progressive sans prétendre à l'exhaustivité.

Chapitre 9. Dispositifs sociotechniques de l'open data

Suite aux étapes de réflexion et d'identification de données, nous proposons dans ce chapitre d'appréhender la phase d'implémentation technologique du projet open data en observant comment des principes et des stratégies peuvent être traduits en dispositifs numériques de diffusion. Selon Badouard (2014), l'approche de « mise en technologie » de projets politiques participatifs consiste à « observer comment ces technologies organisent des relations entre acteurs politiques (citoyens, décideurs, représentants de la société civile, etc.) et cadrent leurs pratiques de participation (mobilisations, consultations, délibérations, votes, etc.), contribuant ainsi à rendre opérationnelles certaines conceptions de la démocratie et de la citoyenneté » (Badouard, 2014, p.32). Le terme dispositif sociotechnique d'information et de communication (DISTIC) désigne à la fois « un lieu privilégié d'interaction entre communication et transmission [...] et un lieu de médiation, composé de multiples facteurs sémiotiques, esthétiques et techniques en interaction qui relie de manière médiate les acteurs sociaux. Il est caractérisé par une articulation spécifique et nouvelle entre un moyen permettant la transmission et la communication » (Peraya et Bonfils, 2012).

Dans notre recherche, l'analyse des plateformes open data ne repose plus seulement sur des activités de diffusion ou de transmission de données, mais aussi des technologies de médiation entre les acteurs. Ainsi que sur le mode d'organisation du contenu en mettant en avant que les plateformes open data sont considérées comme des portails²²⁷ définis par Stiller (2001) comme un « outil fédérateur d'information et de connaissances » mais surtout « un mode d'organisation, un système de classification dans lequel ces informations sont présentées selon des rubriques thématiques ». Nous signalons que notre approche pour analyser les plateformes open data s'intéresse non pas aux interactions établies entre les utilisateurs, mais au design des technologies de diffusion et de participation qui cadre ces interactions.

²²⁷ Par exemple les portails Opendata de la ville de Bordeaux et de Haute-Normandie.

De ces définitions, nous aborderons les plateformes²²⁸ open data au travers de la grille d'observation propre à notre corpus de dispositifs sociotechniques²²⁹ composé des plateformes open data installées par les cinq collectivités territoriales étudiées. Dans l'interface technique des plateformes, nous essayerons dès lors de repérer les signes qui révèlent l'identité du site qui se donne à voir dans son architecture et son contenu, ses modes de participation et de lien social en ligne.

Ainsi nous proposons d'abord d'analyser la structure et le fonctionnement du site définies par son architecture et sa ligne éditoriale. Dans ce cadre, nos observations nous permettent de mettre en exergue la morphologie du site accordée au contenu afin de comprendre comment sont représentées les données, de définir leurs natures, leurs sources et leurs formats. Nous cherchons plus précisément à identifier, en relation avec le contenu, le public ciblé et le type d'utilisateurs. Ensuite, il consiste de repérer les différentes formes d'engagement des internautes selon des modalités de participation incarnées par des applications techniques. Enfin, la troisième étape de la grille réfère à la production d'échanges entre les acteurs dans une optique de générer des liens sociaux entre eux. Nous questionnerons alors les outils de discussion et de partage comme forme de lien social.

²²⁸ Certaines collectivités ont publié leurs données sur une page de leur site web (le cas de la commune de Le Mans et de la région de Basse-Normandie qui publient sur la page web open data une liste et des hyperliens vers les jeux de données) et quelques autres ont eu recours aux services de l'État afin d'héberger leurs jeux de données sur data.gouv.fr. Nous signalons que la majorité des collectivités qui ont ouvert leurs données sur leurs plateformes locales ont référencé également leurs données sur le catalogue national dans un effort de centraliser les données publiques et d'avoir plus de visibilité pour leurs données.

²²⁹ Voir la partie 6.3.2 sur la méthode d'observation en ligne que nous avons adoptée pour l'analyse des dispositifs sociotechniques.

9.1 Structure et fonctionnement du dispositif

La ligne éditoriale des plateformes open data est définie par les collectivités des collectivités territoriales qui poursuivent des objectifs en apparence similaires. Si nous appliquons les principaux critères (voir tableau ci-dessous) à nos observations, nous pouvons constater que ces plates-formes sont à la fois mobilisées en tant qu'instruments de publication et d'organisation des données publiques collectées via un catalogue de jeux de données, et comme des outils de visibilité et de valorisation mettant en scène des objectifs plus stratégiques des plateformes open data.

Tableau 13. Architecture et fonctionnement des plateformes open data

	TECHNIQUE	SOCIAL
Ligne éditoriale	Mise en page, charte graphique	Attractivité
	Contenu	Public ciblé
	Rubriques	Pré qualification
Gestion du site	Nombre de jeux de données publiés	Tactiques de visibilité
	Indicateurs : le nombre de téléchargement de chaque jeu de données, le nombre de vus des pages, le nombre total de téléchargement sur le site	Promotion et valorisation des usages

En effet, nos observations des arborescences des sites et de la manière dont sont agencées les rubriques donnent à voir une standardisation de l'architecture des sites nécessaire, selon nous, à un déploiement de dispositifs plus ou moins stabilisés. Les plateformes open data déploient alors une approche en quatre segments :

1. les données (rechercher, présenter et télécharger les données dans un catalogue de jeux de données)
2. l'information (informer sur les actualités de l'open data, créer un blog, expliquer comment utiliser la plateforme, élaborer un glossaire, informer sur les partenaires)

3. l'usage (préciser les licences d'usage, répertorier les applications, s'adresser aux développeurs, annoncer les concours)
4. le projet (montrer la démarche et les objectifs du projet open data dans la collectivité)

Ces segments sont matérialisées à travers de plusieurs rubriques observées sur les plateformes : la démarche ou le projet, les données, la licence, les applications, les actualités, les concours, les appels à projet, les développeurs, le blog.



Figure 12. Page d'accueil de la plateforme open data de la ville de Montpellier

Nous signalons que les plateformes open data de Grand Lyon²³⁰ et le département de Saône-et-Loire²³¹ ont largement évolué depuis leurs premières publications. Les nouvelles versions ont apporté des changements concentrés sur une mise en page plus attractive et graphique, une nouvelle offre de données (des données de mobilité en temps réel sur la plateforme de Grand Lyon et des data visualisations

²³⁰ Grand Lyon a lancé la première version de son portail de données géographiques en version bêta expérimentale. La version Grand Lyon SmartData V1 lancé en Mai 2013 a apportée des évolutions majeures restructurant complètement l'application. Les deux versions sont conservées. La première diffuse uniquement les données géographiques du service Information géographique, tandis que la deuxième publie des données de tous les services de Grand Lyon. La version actuelle (V3) de la plateforme, lancée en Décembre 2013, (voir figure 13) a mis à disposition une version du site enrichie de données mobilité en temps réel.

²³¹ En novembre 2012, le Conseil général de Saône-et-Loire a lancé la version V2 de son portail open data. Dans cette version, la configuration du site a totalement changée de celle de la version V1 publiée en Septembre 2011

sur Open data 71), et sur une forme d'échange par l'installation d'un outil de discussion notamment sur la plateforme Open data 71 (figure 14). Ces changements donnent un aperçu du chemin de l'évolution des plateformes régulièrement révisés pour rester en phase avec l'avancement du projet et l'évolution des applications du concept même de l'Open data.



Figure 13. Grand Lyon SmartData V2 - Décembre 2013



Figure 14. V2 du portail open data 71 - Novembre 2012

Quelles sont précisément les éléments qui entrent en jeu dans la construction d'un modèle d'une plateforme open data? En rapport avec la ligne éditoriale et la gestion du site, certains éléments nous apparaissent comme incontournables.

9.1.1 Un catalogue de jeux de données

La rubrique « Les données », souvent la page d'accueil par défaut, représente le répertoire des données de la collectivité. Le catalogue

organise et regroupe les données sous des entrées thématiques pour permettre la recherche à facette par catégorie. Dans les cas étudiés, le choix des thématiques est basé sur les services et les métiers de la collectivité producteurs des données. Le moteur de recherche permet de trouver les jeux de données par l'interrogation des métadonnées. La description de chaque jeu de données comporte au moins le titre, la source, une fiche descriptive, une date de création, une date de mise à jour, des mots clés, les formats des fichiers disponibles en téléchargement, la licence, et un lien vers l'URL (Uniform Resource Locator) des données. Notons que les mots clés ne sont pas reliés à des listes d'autorités et des taxonomies publiques pré-contrôlées. La fonctionnalité de filtrage permet de raffiner la recherche par source, catégorie, format. Aussi une option de classement permet aux internautes de trier les résultats de recherche par date de publication, popularité ou par titre. Tous les portails offre l'option de télécharger tout le catalogue des métadonnées sous la forme d'un seul jeu de données.

Pour publier les données, deux critères d'accessibilité sont pris en compte : Les formats des données et leurs modes d'accès. Dans le but de respecter le principe de mise à disposition des données sous des formats réutilisables et non propriétaires, les collectivités ont publié une grande partie de leurs données dans des formats ouverts comme csv, json, kml, etc. Pourtant, il existe toujours des exceptions. A part Smart Data, tous les autres portails diffusent des fichiers dans des formats non-réutilisables et propriétaire comme xls, pdf ou doc. entre autre : le « Rapport Agenda 21 Direction de l'aménagement durable des territoires et de l'environnement » sur Open data 71, l'« Album plein de ville » sur Open data Toulouse, le « Compte rendu du conseil municipale » sur Open data Montpellier et l'« Annuaire des associations de Digne-les-Bains » sur le portail Open PACA. Signalons la multiplicité des formats de téléchargement disponibles pour chaque jeu de données.

Deux modes d'accès sont possibles sur les portails open data, en fonction de la nature des données publiées :

- les données statiques dans des fichiers disponibles par téléchargement et ;

- les données dynamiques accessibles via une Application Programming Interface (API).

Une API permet aux diffuseurs de sélectionner des parties spécifiques de leurs données, plutôt que de fournir leurs données en vrac comme des fichiers téléchargeables. API est généralement connectées à une base de données qui est mise à jour en temps réel. Cela signifie que la diffusion des données via une API peut s'assurer que les données sont mise à jour. Les APIs sont très utiles pour les données qui évoluent rapidement dans le temps. Pour les réutilisateurs, l'extraction des données, est facilitée par un processus automatique lors des mises à jour des données.

La nature des données et la fréquence de mise à jour des données publiées sont deux critères essentiels pour offrir les données à l'intermédiaire d'une API. En effet, les API sont encore peu proposées par les collectivités. Les données publiques sont prioritairement publiées sous un format statique de fichiers. Dans les cas étudiés, seulement deux APIs sont mises à disposition : une API sur les données de mobilité (données trafic et événements routiers) sur Smart Data de Grand Lyon et une deuxième sur les résultats des élections cantonales depuis 1994 sur Open Data de Montpellier. Les moyens et les compétences de la collectivité sont aussi des critères associés à la fourniture d'une API qui nécessite un cout supplémentaire pour son développement et sa maintenance. L'accès aux données en vrac veille à ce qu'il n'y a pas de dépendance à l'égard du fournisseur d'origine des données, ce qui signifie que si un cycle de restructuration ou le budget modifie la situation, les données sont encore disponibles.

Notons que le portail Smart Data de Grand Lyon a créé des niveaux d'accès privilégié. Contrairement au principe de « non discrimination » de l'open data, les usagers sont invités à créer un compte spécial pour accéder aux données dont l'accès est non autorisé au grand public. Vu que le projet de Grand Lyon donne accès aux données sous 3 types de licences différentes²³².

²³² Voir chapitre 4.5

9.1.2 Un outil de valorisation des données et des réutilisations

La plateforme open data est un outil de communication pour valoriser les données ouvertes et favoriser leur exploitation. La présence de la rubrique « Démarche du projet » ou « Le projet », dans tous les portails open data, explique les grands principes du concept de l'open data et présente ses avantages. Certains portails incluent un glossaire et une foire aux questions sur l'open data (*Frequently asked questions* FAQ). D'autres publient des posts d'actualités sur l'open data. Open data 71, Open data Toulouse Métropole, Open Data Montpellier et Open PACA, mettent en œuvre des stratégies de veille pour recenser et diffuser sur leurs portails des articles sur l'open data. Les portails affichent aussi les « dernières données libérées » pour montrer l'actualité sur les jeux de données. Ce qui nous fait dire que les plateformes open data révèle principalement la conception d'un dispositif de sensibilisation à l'open data.

Pour promouvoir l'usage des données, les collectivités organisent des événements, concours et lancent des appels à projet. Par la création d'une rubrique « Concours », les collectivités se servent du site pour faire les annonces sur ces événements. En plus, la rubrique « Applications » sur les portails open data est un espace de présentation des applications développées à partir des données ouvertes. Dans cet espace, les producteurs de données essaient de recenser les applications de l'open data ainsi que les usagers sont invités à référencer leurs produits créés à partir des données ouvertes. Le portail Open PACA a créé un catalogue des applications dérivées des données ouvertes. L'objectif de cette rubrique est de mettre en évidence l'intérêt des données et de susciter leur exploitation à travers des exemples de leur potentiel. Ainsi, les plateformes sont appréhendées comme des outils de valorisation des données publiques et de leurs réutilisations. Les plateformes offre souvent des statistiques et des informations sur l'usage tels que le nombre de téléchargement de chaque jeu de données, le nombre de vus des pages et le nombre total de téléchargement sur le site. L'intérêt d'afficher ces indicateurs d'usage ou de mettre en relief les « jeux de données les plus téléchargés », ou « most viewed » assure la valorisation des données et montre aussi l'utilité des données ouvertes. L'aspect informatif des plateformes comme par exemple sur

la page « licence » explicite le contexte légal d'utilisation et les responsabilités auxquelles s'engagent le producteur des données ainsi que le réutilisateur. La spécification de la licence garantit que les conditions d'utilisation des jeux de données sont claires et documentées. D'autres pages comme « Développeurs », « Documentation » et « Utiliser SmartData » donnent des indications sur l'usage, et les modes d'accès (téléchargement, API), etc.

9.1.3 Un outil de visibilité politique

Au premier contact avec le site, la mise en page, les graphiques, les couleurs sont des facteurs d'attractivité destinés à susciter l'attention. Certes, le portail est un outil de communication pour valoriser les données ouvertes et favoriser leur exploitation mais également un outil de visibilité à la collectivité qui cherche, par le projet open data, à offrir aux citoyens un nouveau service de valeur.

La rubrique « démarche du projet » explique, avec un ton optimiste et une volonté manifeste, le cadre du service proposé et le déroulement du projet. Cette approche de marketing utilise le portail open data en tant qu'outil de communication pour mettre en relief, au premier rang, le « produit » qui est « les données publiques » mais également, au deuxième rang, les efforts de la collectivité (élus et fonctionnaires) pour mettre en place un tel service public. Cette rubrique publie un discours politique qui confirme le positionnement de la collectivité dans la politique de l'open data, et sa volonté de tirer profit de tous les bénéfices qu'apporte l'open data : la transparence, le développement des nouveaux usages et services. Dans ce cas, le portail open data devient la vitrine de l'engagement de la collectivité dans ses missions.

Notre observation révèle d'autres éléments, devenus basiques dans toutes les plateformes open data, tel que l'affichage des indicateurs de comptage. Les compteurs de nombre de jeux de données publiés, le nombre de téléchargement de chaque jeu de données, le nombre de vues des pages, le nombre total de téléchargement sur le site, sont également considérés comme des signes de visibilité et d'autopromotion du projet de la collectivité. Ces indicateurs de mesure ne sont pas purement techniques car ils sont des signes de

distinction sociale hautement symboliques. Ils permettent aux collectivités de connaître l'intérêt suscité par la mise à disposition de leurs données en ligne.

9.1.4 Le public visé

Autant d'éléments observés démontrent que le contenu du site est construit sur une offre spécifique qui s'adresse à un public ciblé. Open data semble être le type de données adapté pour les experts puisque les formats standards techniques (csv, kml, ...), le volume des données, et les outils d'exploitation et de traitements de ces données requièrent un niveau minimum de compétence ou d'expertise pour les manipuler. Pourtant, la morphologie et la ligne éditoriale des portails reflètent une double approche expert/amateur. Loin d'être uniquement un outil et une source pour un public restreint d'experts de la donnée, les portails open data s'adressent également aux amateurs de la donnée donc au grand public. Ce résultat repose sur le postulat qu'il existe une interdépendance entre la construction stratégique du site et le public ciblé. En effet, plusieurs éléments observables démontrent que les portails s'adressent à la fois aux amateurs de la donnée et aux spécialistes.

D'un côté, le portail est, par excellence, un site de vulgarisation de l'open data. La rubrique « Démarche du projet » ou « Le projet », présente dans tous les portails open data, définit explicitement ce que c'est l'open data ? En plus, les outils de vulgarisation ludiques des données sont un autre indicateur de l'approche « open data pour tous ». Cette technique n'est pas généralisée car seulement, Open data 71 et Smart Data, offrent des outils de visualisation sur leurs portails. Même si le portail Open data Montpellier ne fournit pas ce type de service, Jérémie Valentin le considère comme une valeur ajoutée dans un projet open data. L'outil de Open data71 permet de créer des graphes (diagrammes circulaires, camemberts, courbes) et Smart data affichent les données sur des cartes quand des données de géo-localisation sont disponibles dans le jeu de données. Le développement d'outils de visualisation sur les portails permet d'envisager des réutilisations plus larges des données. Egalement, outre les formats de fichiers lisibles uniquement par machine (CSV,

WMS, XML, etc), la présence de formats comme pdf, doc, jpg et txt compréhensibles par tous, confirme cette double approche expert/amateur. Les données brutes dans des formats ouverts lisibles seulement par machine, les APIs, et la page dédiée aux développeurs s'adressent à un public expert et technique de la donnée. La page « développeur » a pour objectif de fédérer les développeurs en donnant des indications techniques et la documentation nécessaire pour accéder aux données.

Même si le portail Smart Data ne publie que des données au format ouvert, la rubrique « Utiliser Smart Data » sur le portail de Grand Lyon, propose deux niveaux de documentations pour l'accès et l'usage des données : pour « les débutants » et pour « les développeurs ». Le « coin des débutants » offre des informations basic pour s'appropriier des données tandis que le « coin des développeurs » contient une documentation très technique sur les formes d'accès aux données.

9.2 Modalités de participation en ligne

Dans notre optique de design, nous essayons d'identifier les applications présentes qui permettent d'agir sur le contenu des plateformes open data. Nous observons différentes formes de participation et de contribution qui implique plus ou moins l'utilisateur des plateformes open data et rendent possibles certaines actions (tableau 14).

Tableau 14. Modalités de participation sur les plateformes open data

	TECHNIQUE	SOCIAL
Contributions	Publication de nouvelles données	Coproduction
	Commentaire	Réactions
	Like de contenus	Approbation
	Noter les jeux de données de 1 à 5	Évaluation
Coopération	Outils de mutualisation	Choix de coopération

L'engagement des collectivités dans « un processus ouvert » vers différents partenaires se reflète par l'installation d'outils qui permettent aux réutilisateurs de contribuer à l'ouverture des données. Les internautes bénéficient d'une multitude d'applications pour produire, commenter et évaluer les contenus. Evidemment, il existe plusieurs degrés de contribution et d'implication des réutilisateurs sur le site, allant par ordre croissant, de l'approbation par simple clic sur « j'aime » ou sur les étoiles de popularité d'un jeu de donnée, à la réaction en publiant un commentaire, ou à la possibilité d'ajouter des jeux de données.

La majorité des plateformes open data²³³ observées permet à tout usager d'évaluer la qualité ou l'intérêt du jeu de données mis à disposition par la collectivité. Le module de « like » sur Open data 71, affecte le classement des jeux de données dans le catalogue. Le module des étoiles de popularité installé sur les sites de Smart Data, Open Data Toulouse et de Montpellier permet aux réutilisateurs, de noter sur une échelle de 1 à 5 l'intérêt du jeu de données. Cette technologie contribue selon nous à une possibilité d'action en attribuant aux utilisateurs le rôle d'évaluateur des données et conduisent à leur faire interagir à la qualité ou l'intérêt des jeux de données publiées. Elle soulève alors la question en terme de la politique open data des collectivités qui, comme nous l'avons déjà signalée dans l'étape de réflexion, s'orientent vers une vision politique plus pragmatique et moins idéaliste de la valeur des données produites dans le cadre de l'activité du service public. Cependant, le degré de réaction des utilisateurs sur les jeux de données reste très limité vu qu'aucune des cinq plateformes ne fournissent la possibilité d'ajouter un commentaire ou de laisser un avis ou suggestion sur chaque jeu de données. Cela empêche les utilisateurs des données et des plateformes open data de réagir librement et même de créer des chaînes de commentaires ou des fils de discussion, sorte de lien social entre eux.

La coproduction des contenus est une activité familière sur les plateformes participatives. L'idée que l'utilisateur n'est plus un simple consommateur de données mais un coproducteur de contenus se

²³³ Sauf celle de Open Paca.

matérialise sur les plateformes open data par la mise en œuvre d'une participation « contrôlée ». Le design des cinq plateformes ne permet pas à des personnes externes de référencer et d'ajouter librement des jeux de données au catalogue de données²³⁴ mais apporte une solution technique propre à la mutualisation. Ainsi la coopération entre les collectivités et d'autres acteurs producteurs de données se traduit aussi sur les plateformes open data mutualisées. En effet, pour la publication des données, la mutualisation est un choix à considérer. Dans un projet open data, la mutualisation locale (entre les collectivités locales) ou la mutualisation centralisé (entre une collectivité locale et le portail national d'Etalab) s'appuie sur une plateforme technique unique dotée d'un catalogue de données commun. La mutualisation centralisée, comme nous l'avons souligné, entre dans la catégorie des services que l'Etat offre pour aider les collectivités territoriales à ouvrir leurs données. Bien que les collectivités soient incitées à s'en servir, en étudiant les sources et les producteurs de données listés sur le site data.gouv.fr, nous avons repéré seulement dix collectivités (un département et neuf communes²³⁵) qui ont choisi un hébergement mutualisé sur la plateforme nationale. Par contre, une majorité des collectivités qui ont ouvert leurs propres portails open data ont référencé leurs données sur data.gouv.fr. D'un côté, les petites communes souhaitent publier leurs données sur le portail national pour réduire les contraintes financières et par manque d'expertise, de l'autre côté, les grandes collectivités apportent plus de visibilité à leurs données. Ce type de partenariat entre collectivités territoriales est encouragé voire même récompensé. En novembre 2013, la Ville de Nantes, Nantes métropole, le Département de Loire-Atlantique et la Région des Pays de la Loire ont été primés pour la mutualisation de leur plateforme de

²³⁴ Nous signalons que la plateforme open data du département de la Gironde « Ressourcerie datalocale », permet à des personnes autorisés et muni d'un identifiant de devenir des fournisseurs de données et de pouvoir référencer ou télécharger un jeu de données dans le catalogue. Selon le conseil général de la Gironde, l'objet d'une ressourcerie participative et citoyenne est d'organiser et de favoriser la rencontre entre les données descendantes (publiques) et les données remontantes (individuelles et collectives). <http://www.datalocale.fr>

²³⁵ Cantal, Coulommiers, Longjumeau, Louvier, Perreux-sur-Marne, Saint Quentin (Aisne), Saint-Maur des Fossés, Saint-Quentin, Sarlat-La Canéda, Villemomble

données publiques²³⁶. Les Prix Territoriaux créés par la Gazette des communes et l'entreprise d'assurance GMF (Garantie Mutuelle des Fonctionnaires) récompensent « les collectivités locales ayant réussi des projets locaux en impliquant plusieurs métiers et services au sein de leur structure ». La région PACA a proposé à d'autres collectivités dans la région de publier sur un site commun leurs données. La mutualisation portée par la région s'inscrit dans les politiques relatives à l'aménagement du territoire dans l'idée d'appuyer les plus petits territoires, moins bien équipés, dans l'usage d'outils techniques de diffusion. Elle donne la possibilité de profiter d'une architecture de publication, tout en gardant la propriété et la gestion de ses données. Pourtant, sur l'ensemble des 39 portails open data, seulement sept d'entre eux sont des portails mutualisés entre des collectivités du même territoire. Ce fait s'explique d'abord par des raisons politiques. Pour expliquer le choix d'un hébergement dédié, « la mutualisation dépend en grande partie de la volonté politique indispensable pour franchir cette étape », reconnaît Jean-Michel Martin²³⁷, responsable de la DSIIG du département du Saône et Loire. Dans un cas similaire, la communauté d'agglomération de Montpellier et la ville de Montpellier ont créé séparément deux plateformes open data pour publier leurs données. Cet exemple illustre la défiance des coopérations transfrontalières. En outre, la mutualisation montre aussi ses limites techniques. Les collectivités craignent que leurs actions ne soient pas assez visibles sur une plateforme collective. « Chaque territoire souhaite disposer de son propre espace sur son site, pour conduire le travail d'animation entre producteur et consommateur de données. C'est un point d'achoppement des projets de partenariats »²³⁸, reconnaît Gaëlle Rivoal en charge de l'innovation pour le département de la Loire-Atlantique. Pour résoudre ce problème technique, « la personnalisation » des portails donne la possibilité à chaque

²³⁶ Voir : « Prix Territoriaux : La Plateforme Open Data Mutualisée Récompensée ». *Nantes Ouverture Des Données*, November 20, 2013. <http://data.nantes.fr/actualites/detail-dune-actualite/prix-territoriaux-la-plateforme-open-data-mutualisee-recompensee-1/>.

²³⁷ Entretien le 30 juin 2014.

²³⁸ Cité dans Berdot, Vincent. « Pourquoi Les Collectivités Rechignent À Mutualiser Leurs Plates-Formes Open Data ». *01Business*, April 6, 2012. <http://pro.01net.com/editorial/567381/pourquoi-les-collectivites-rechignent-a-mutualiser-leurs-plates-formes-open-data/>.

collectivité de créer son propre site web avec ses rubriques et couleurs en séparant le stockage des données des frontaux web qui servent à les exploiter. En effet, dans le cadre de leur partenariat, la Ville de Nantes, Nantes Métropole, le Conseil Général de Loire-Atlantique et la Région des Pays de la Loire partagent la même plateforme dotée d'un catalogue de données commun mais les couleurs et le contenu des diverses rubriques (actualités, démarche du projet, concours, forum, etc.) changent par collectivité. La personnalisation n'est pas purement technique car elle est un signe de distinction sociale pour la visibilité. En plus, l'intérêt d'avoir son propre site Open Data devient un symbole de compétence, de modernisation et une marque d'efficacité permettant de se positionner dans l'activité générale au niveau local et national.

9.3 Configuration du lien social

En développant des applications qui permettent d'échanger sur les données, les dispositifs sociotechniques organisent des liens sociaux et cadrent la manière dont les utilisateurs font une communauté en ligne.

Tableau 15. La configuration du lien social sur les plateformes open data

	TECHNIQUE	SOCIAL
Discussion	Forum	Fils de discussion
	Messageries	Echanges privés
Partage	Mails, forme de contact	Partage interpersonnel
	Facebook, Google +, Twitter	Affichage et réseaux interpersonnels
	Flux RSS	Diffusion

En effet, les collectivités ont créé des espaces dédiés aux échanges autour du projet open data. Pour encourager l'interaction entre les producteurs et les réutilisateurs intéressés par les données ouvertes, plusieurs formes d'échanges via des outils de discussion sont possibles sur les plateformes. Ils découlent directement du projet politique porté par les collectivités et de leurs méthodes de travail

reposant sur la participation, l'ouverture et la collaboration. Tous les portails offre la possibilité d'entrer en contact avec le diffuseur des données ou le responsable du projet open data de la collectivité. Des formulaires de contact et des messageries privées, que nous avons personnellement utilisés pour prendre contact avec les chefs des projets open data, sont mis en place permettant aux réutilisateurs de poser des questions, donner des remarques, ou de demander ou suggérer la mise en ligne de nouveaux jeux de données. Smart Data de Grand Lyon précise le contact pour chaque jeu de données ce qui permet aux réutilisateurs de rejoindre le service producteur et d'échanger directement avec lui. Le design de ces applications produit des effets directs sur la dynamique des échanges entre les producteurs et les utilisateurs. En plus, l'espace dédié aux développeurs permet de référencer leurs applications. Ainsi ils peuvent partager leurs applications et présenter leurs utilisations des données sur la plateforme open data.

Sur le plan des possibilités d'interaction entre les utilisateurs sur les plateformes open data, seulement une application est proposée aux participants afin qu'ils puissent échanger entre eux. Il s'agit du forum sur la plateforme open data 71 qui forme un espace participatif entre la communauté des réutilisateurs et un espace d'échange sur les données. La mise en place de forum est un excellent moyen de montrer l'engagement de la collectivité dans le déploiement d'outils de discussion autour de l'open data. De ce fait, nous pouvons constater que la majorité des collectivités favorisent, sur leurs plateformes open data, une approche individuelle, directe et privée d'échange : la configuration du lien social est conçue alors comme l'agrégation des remarques et suggestions des utilisateurs qui prennent part au projet d'ouverture des données. Cependant, en termes d'interactions plus ouverte, les collectivités²³⁹ ont préférées par la création de plusieurs comptes sur les réseaux sociaux (comme twitter et facebook) de bénéficier de tels outils afin de construire leurs réseaux et de diffuser l'information au sein de la communauté en

²³⁹ Toutes les collectivités, sauf le département de Soane et Loire, possèdent des comptes sur les réseaux sociaux comme twitter et facebook.

ligne. Le flux RSS est aussi un outil de partage et de diffusion disponible sur la majorité des plateformes open data.

Comme nous l'avons dit au début de ce chapitre, dans cette recherche nous n'avons pas opéré d'analyse des formes de sociabilités qui s'y développent entre les acteurs sur les réseaux sociaux que nous avons pu détecter mais au design des technologies de diffusion et de participation qui cadre ces interactions.

Conclusion du chapitre 9

Pour conclure cette analyse des plateformes open data, nous soulignons l'osmose des dimensions techniques et sociales de l'implémentation technologique des projets open data par les collectivités. L'évolution des plateformes open data afin de rester en phase avec l'avancement du projet et l'évolution des applications du concept même de l'Open data a conduit à une standardisation de l'architecture des sites donc un dispositif plus ou moins stabilisé. Une standardisation de l'architecture des plateformes montre aussi une politisation du dispositif perçue comme un mode de régulation et de juxtaposition des regards et des pratiques de diffusion ce qui marque, en effet, une conciliation fonctionnelle entre les collectivités.

Les quatre éléments explicites que nous avons pu observés sur toutes les plateformes - les données, l'information, l'usage et le projet - montrent que celles-ci sont mobilisées en tant qu'instruments de publication et d'organisation des données publiques collectées via un catalogue de jeux de données, de valorisation des données publiées et de leurs usages et enfin un outil de visibilité politique mettant en scène des objectifs plus stratégiques des plateformes open data.

La structure et la ligne éditoriale des plateformes mettent en œuvre une participation « passive » visant avant tout à publier des données organisées dans des catalogues de jeux de données et à les mettre à disposition des internautes pour un usage possible. Les plateformes open data implémentent une participation « transformative » visant aussi à sensibiliser les internautes à l'open data. Elles valorisent les utilisateurs actifs (via le catalogue des applications produites), pour autant, le public cible comporte à la fois des experts et des amateurs. Loin d'être uniquement un outil et une source pour un public restreint d'experts de la donnée, les portails open data s'adressent également aux amateurs de la donnée donc au grand public.

Ensuite, les modèles de participation qui permettent d'agir sur le contenu des plateformes open data révèlent d'une architecture qui contribue à une participation limitée et une coopération contrôlée. Le degré de réaction des utilisateurs sur les jeux de données reste très limité. Dans nos cas, l'activité de participation attribut aux utilisateurs

le rôle d'évaluateur des données et conduisent à leur faire interagir dans un cadre prédéfini à la qualité ou l'intérêt des jeux de données publiées sans qu'ils ne réagissent librement ou laissent un commentaire. La coproduction de données se traduit par une collaboration de plusieurs producteurs de données et les plateformes open data sont alors conçues comme des outils de mutualisation de données publiques.

Enfin, et suite à l'analyse des outils de discussion et de partage, nous avons constaté que la majorité des collectivités favorisent un échange individuel, directe et privé via leurs plateformes open data et en termes d'interactions plus ouverte, elles préfèrent la création de plusieurs comptes sur les réseaux sociaux afin de construire leurs réseaux et de diffuser l'information au sein de la communauté en ligne.

Chapitre 10. Promotion et usages des données ouvertes

Charles Népote (2010), de la Fing, affirme que l'open data est « un mouvement accompagné voire stimulé ». Jérémie Valentin explique qu'une partie de son métier, en tant que chef de projet open data, s'agit d'un vrai travail de communicant. « L'Open Data ne peut fonctionner que si les gens utilisent ces données, explique le chef de projet. Pour cela, il faut nous faire connaître tant auprès de mes collègues que des citoyens ».

Les campagnes de promotion ont accompagné les premiers projets d'ouverture dans le monde. Comme nous l'avons déjà signalé, Washington D.C. était une des premières villes à ouvrir ses données sur data.dc.gov en 2007., DC, Vivek Kundra, Chief Technology Officer (CTO) de la ville (ensuite premier Chief Information Officer fédéral) et porteur du projet d'ouverture des données, a sollicité en Septembre 2008, iStrategyLabs, une agence de marketing dans le but de rendre le catalogue plus utile aux citoyens, les visiteurs du site, les entreprises et les organismes gouvernementaux de Washington. Selon iStrategy Labs, « la solution a été de tenir un concours d'innovation où nous avons mis les données dans les mains de nos citoyens talentueux, et leur avons donné des prix en argent et la reconnaissance pour leurs efforts dans le développement de technologies pour leurs voisins et le gouvernement de la ville. [...] La stratégie consistait à demander aux gens de créer des applications en utilisant les données du catalogue de données fraîchement lancé. Il comprenait une soumission en ligne pour les applications, de nombreux petits lots plutôt que quelques grandes, et plusieurs catégories différentes ainsi que d'un prix «Choix du public» ». Ainsi, « Applications pour la démocratie », était l'une des premières initiatives de marketing des données sous forme de concours de compétitions, lancée en Octobre 2008 par la ville de Washington D.C. aux États-Unis. Le concours était ouvert pendant 30 jours et a coûté au gouvernement DC 50,000 USD. En retour, un total de 47 iPhone, Facebook et les applications Web ont été développés avec une valeur estimée à plus de 2,6 millions pour l'économie locale.

Cet exemple montre qu'une stratégie de Marketing de la donnée pourrait apporter des solutions au défi de l'usage et l'utilité des données. En effet, dans la démarche d'accélération de l'ouverture que les collectivités adoptent, ils attendent des manifestations, des produits, et des réutilisations rapides à partir de leurs données, ce qui, rend nécessaire le développement de techniques spécifiques et efficaces pour promouvoir la réutilisation des données.

Ce chapitre s'arrête en premier sur la quatrième phase du processus d'ouverture des données publiques par une collectivité. Il s'agit d'analyser la stratégie de communication et de marketing de la donnée publiques et du projet de l'open data au sein des collectivités étudiées. Ainsi, nous proposons les caractéristiques du marketing de l'open data locale ainsi que les orientations de promotion de la donnée publique face à la communauté de réutilisateurs. Dans un deuxième temps, nous insérons les impacts et les réactions sur la stratégie de promotion dans le contexte agité des usages sociaux des données. Ainsi, nous transmettons tout d'abord le retour des chefs de projet open data sur les usages de leurs données et nous relient dans une approche marketing l'usage aux besoins des usagers. Alors en identifiant les besoins, nous pourrions comprendre la nature des usages et les motivations des réutilisateurs présentés dans le chapitre dernier.

Nos analyses sont basés sur les propos des chefs de projets open data mais également sur l'observation des plateformes open data des collectivités qui nous permet de discerner les usages et les produits dérivés de l'open data référencés sur les plateformes.

10.1 Les caractéristiques du Marketing de l'open data locale

Pour étudier et analyser les stratégies de communication des données publiques mises en œuvre par les collectivités, nous nous sommes appuyés sur l'approche de marketing standard de McCarthy (2005) basée sur les « quatre P » : produit, prix, place et promotion.

Le marketing mix de ces quatre éléments prend en considération :

- les caractéristiques du produit et comment il répond aux besoins de l'utilisateur final ;
- le coût de production y compris le temps, l'énergie pour la fixation de prix ;
- la région géographique de distribution du produit ;
- les méthodes de promotion tels que la publicité, la promotion des ventes, relations publiques, marketing direct, marketing numérique, etc.

Par ce qu'il s'agit de données publiques ouvertes gratuitement, nous avons retiré l'élément « Prix » sans toutefois déprécier son importance en tant que facteur d'attraction essentiel pour motiver les réutilisations. Nous présentons alors les trois éléments d'une stratégie de marketing open data locale.

10.1.2 Le produit : Les données publiques

En premier lieu, les collectivités font face à la complexité liée à la donnée publique elle-même. Les données publiées sur les portails, couvrent toutes les compétences, donc tous les domaines d'applications de la collectivité (environnement, éducation, finances publiques, vie associative, agriculture, arts et musique, tourisme, etc.). En plus, nous trouvons des données de différents degrés d'ouverture technique (des fichiers non structurés lisibles par tous et des fichiers structurés lisibles uniquement par machine). Cette diversité des données (plusieurs domaines et différentes propriétés techniques) élargit le public visé. La double approche amateur/expert des projets open data pose une autre complication dans la stratégie

de communication²⁴⁰. Les collectivités sont obligées d'utiliser une variété de techniques pour toucher les deux cibles : d'une part les amateurs et de l'autre les experts. Dans ce sens, Jérémie Valentin parle des méthodes qu'ils ont utilisées pour promouvoir l'open data : Organisation de conférences, de concours, de manifestations... Les initiatives sont variées car il faut toucher à la fois le grand public mais aussi les chercheurs, les développeurs, les industriels, etc.

En plus, dans une stratégie marketing des données publiques, il est difficile de repérer et d'identifier la nature des besoins que les collectivités visent à satisfaire par la publication de leurs données. Contrairement au cas de consultation sur demande, les usagers ne sont pas connus et leurs besoins ne sont pas exprimés. Ainsi, les usages issus des données ne sont pas prévisibles et à la collectivité d'anticiper les usages pour qu'elle puisse mettre en œuvre sa stratégie²⁴¹.

10.1.3 La place : le territoire

Nous nous référons à l'élément « territoire » pour désigner la région géographique attachée au processus d'ouverture. En effet, nous considérons que la dimension territoriale et l'open data locale sont inséparables. Même si, la « distribution » des données est via le web, les pratiques et usages des données sont loin d'être purement virtuels. La dimension territoriale est évidente par ce que toutes les données publiques publiées par les collectivités sont associées au territoire et au champ d'autorité de la collectivité. En outre, l'usage des données publiques locales est fondamentalement lié aux besoins sociaux spécifiques du territoire dans lequel vivent les usagers ou exercent leurs métiers, d'où la parole d'une « communauté open data locale ». Le fait d'étudier l'open data locale nous offre ici l'occasion d'analyser les formes de participation locale des usagers qui se transforment en projets afin de démontrer l'intérêt de l'usage des données publiques. L'idée du couplage de la vie sociale, politique et

²⁴⁰ Voir chapitre 6.4

²⁴¹ Voir la troisième partie sur les usages sociaux des données ouvertes

économique avec la communauté virtuelle des habitants du territoire en question semble être de plus en plus possible.

10.1.4 Les techniques de promotion

En étudiant les projets open data locaux, nous avons observé plusieurs méthodes et outils de promotion. Comme nous avons déjà signalé, le portail open data, avant d'être un catalogue de jeux données, est un outil de promotion et de vulgarisation du concept open data. D'autres outils en ligne comme facebook (Open data 71) et Twitter (#opendata71 et #openPACA) sont aussi utilisés pour atteindre la communauté open data en ligne. D'autres activités, face à face, menées dans le cadre des projets open data, comprennent typiquement des concours, des conférences, des ateliers de travail et des "hackathons".

Pour la promotion et l'animation autour de l'open data, ces activités sont souvent organisées par les collectivités en collaboration avec les associations au niveau national et local à la réalisation de cet objectif commun²⁴². En particulier, nous citons l'effort fait par la Fing qui accompagne les collectivités dans leur processus d'ouverture. Depuis 2010, elle s'engage dans une campagne sur l'ouverture des données publiques en partenariat avec plusieurs territoires. Un acteur actif sur le plan national, la Fing a co-organisé en partenariat avec les collectivités des événements autour de l'open data²⁴³. Elle a participé au lancement du concours "Rennes en accès libre" avec la ville et la Métropole de Rennes en Octobre 2010. La Fing a accompagné la Communauté Urbaine de Bordeaux et la région Aquitaine dans l'animation de la démarche en organisant des « Assises » de la réutilisation des données publiques. De même dans la région PACA, la Fing a participé au lancement de HACK DATA PACA.

²⁴² Voir chapitres 3.2.3 et 4.1.2

²⁴³ Voir : "Une Campagne Autour de l'OpenData". *Fing*. Accessed May 31, 2014. <http://fing.org/?Une-campagne-autour-de-l-OpenData&lang=fr>.

10.2 Les orientations de la promotion

Les techniques de promotion observées suivent trois orientations : la promotion du projet, la promotion de « la culture des données » et la promotion de la réutilisation.

10.2.1 La promotion du projet

L'objectif est de faire l'annonce du projet et présenter le portail des données. Des méthodes comme les communiqués de presse, des annonces sur le site web de la collectivité, ou des annonces pendant des événements publics locaux ont été adaptées pour faire de la publicité sur le projet et pour inciter les gens à aller sur le portail open data de la collectivité.

10.2.2 La promotion de « la culture des données »

Dans une stratégie qui vise les amateurs de la donnée, les collectivités doivent donc passer par un long « processus d'apprentissage », ce qui, par conséquence, prolongera le processus d'ouverture avant que les données publiques soient complètement exploitables. Cette approche est incontournable pour atteindre les objectifs définis par la collectivité et tirer pleinement parti des promesses de l'open data qui concerne tous les citoyens.

La sensibilisation du grand public à l'exploitation des données publiques est un effort commun réalisé par tous les acteurs des données publiques : collectivités, associations et collectifs open data et les éducateurs à la donnée. Les tentatives par lesquelles les différents acteurs s'efforcent de promouvoir une culture de la donnée publique se fédèrent autour d'une série d'actions, d'événement, et de formations autour de l'open data. Alors, dans le processus d'accompagnement par la formation, plusieurs programmes et événements se sont mis en place.

Le concept d' « InfoLab » de la Fing consiste à créer un espace d'enseignement, d'échange et de réflexion sur les données. Il vise à donner aux différents acteurs des compétences pour savoir repérer,

produire, comprendre, utiliser, échanger des données. Plusieurs collectivités et communautés locales se sont inspirées de l'initiative de la Fing pour créer des InfoLab locaux comme celui de Montpellier et de Rennes.

School of Data ou l'École des Données est un projet de l'OKF lancé en mai 2012 qui propose un ensemble de cours sur l'open data. L'école a pour but de donner aux organisations de la société civile, les journalistes et les citoyens les compétences nécessaires pour utiliser les données efficacement. « Le projet part du constat que la société civile (citoyens, ONG, journalistes, associations, etc.) pourrait bénéficier grandement des promesses de l'open data mais qu'elle ne dispose pas ou peu des compétences requises pour comprendre et analyser des données »²⁴⁴. Dans le même esprit, l'Open Data School imaginé par Aquitaine Europe Communication (AEC), l'agence aquitaine des initiatives numériques, en partenariat avec plusieurs acteurs locaux, « se veut "accélérateur d'innovations" grâce, notamment, à une mise en situation de production dans un espace-temps déterminé »²⁴⁵. Dans le cadre du programme open data qui associe la Communauté urbaine de Bordeaux et le Conseil général de la Gironde, un atelier a eu lieu en juin 2013 pour présenter des outils gratuits du web permettant de produire des visualisations de données sans connaissance technique spécifique²⁴⁶.

Il est évident que l'appropriation citoyenne des données publiques ouvertes se fait à plusieurs niveaux. La sensibilisation commence par les jeunes générations. En collaboration avec les institutions d'éducation, les éducateurs, et *l'Association d'Action et d'Education Numérique Citoyenne*, la sixième session d'info Lab de la ville de Montpellier a privilégié les lycéens pour leur faire découvrir le monde des données. Le but des interventions dans des classes de Terminale

²⁴⁴ "Ecole Des Données" Accessed July 16, 2014. <http://ecoledesdonnees.org/>.

²⁴⁵ Voir : Galy, Suzanne. "Open Data School : 130 Étudiants Pour Un Maximum de Créativité" *AEC Agence Des Initiatives Numérique*, February 3, 2012. <http://www.aecom.org/Vous-informer/Actualites2/Open-Data-School-130-etudiants-pour-un-maximum-de-creativite>.

²⁴⁶ Voir : Galy, Suzanne. "Les Outils de Data Visualisation" *AEC Agence Des Initiatives Numérique*, June 13, 2013. <http://www.aecom.org/Vous-informer/Actualites2/Les-outils-de-data-visualisation>.

est d'introduire aux étudiants comment les données publiques sont produites ou collectées, comment sont-elles utilisées et d'illustrer les manières de les approprier²⁴⁷.

Dans une démarche similaire, la région Rhône-Alpes, en partenariat avec l'association *Fréquence Ecoles*, a lancé en 2014 le projet #dataviz_rra sur la data-visualisation des données publiques par les lycéens²⁴⁸. L'objectif de ce projet académique expérimental est d'accompagner les lycées dans la réalisation d'infographies à partir des données ouvertes. Alors, les étudiants seront capable de comprendre la notion d'open data, de découvrir le métier de data-journaliste, d'apprendre à développer des compétences autour des données. En effet, au cœur de sa mission, l'association *Fréquence Ecoles* s'engage à développer chez les jeunes des compétences numériques solides leur permettant d'avoir une attitude critique face aux medias.

Le « Data Journalisme Lab » est un autre projet académique sur les données publiques²⁴⁹. Le projet lancé par l'Institut de Journalisme Bordeaux Aquitaine (IJBA) et l'agence des initiatives numériques AEC associe des étudiants journalistes en Master 1, des étudiants graphistes et des étudiants développeurs accompagnés par des professionnels de l'informatique et du graphisme, et entourés d'experts : cartographe, statisticien, démographe.

247 Voir : Valentin, Jérémie. "Infolab # 6 : Découverte Par Les Lycéens Du Monde Des Données." - Données Publiques Ouvertes - Montpellier Territoire Numérique, June 13, 2014. <http://opendata.montpelliernumerique.fr/Infolab>.

248 Voir : Fréquence Ecoles, and Violaine Lemaitre. "Découvrez Ici L'intégralité Du Dispositif Pédagogique #dataviz_RRA | Fréquence Écoles." *Fréquence Ecoles*, June 25, 2014. http://frequence-ecoles.org/2014/06/25/decouvrez-ici-lintegralite-du-dispositif-pedagogique-dataviz_rra/.

249 Voir : Galy, Suzanne. "Une Expérience de Data Journalisme Est Lancée À Bordeaux." *AEC Agence Des Initiatives Numérique*, January 23, 2012. <http://www.aecom.org/Vous-informer/Actualites2/Une-experience-de-data-journalisme-est-lancee-a-Bordeaux>.

10.2.3 La promotion de la réutilisation

En s'adressant aux experts, les collectivités visent à les motiver et à les faire participer au processus d'ouverture en profitant de leurs capacités. La promotion de la réutilisation par la production de prototype ou d'exemples entre dans la démarche accélérée de l'ouverture adoptée par les collectivités. La promotion de la réutilisation des données passe par une campagne de sensibilisation du rôle des experts en tant qu'intermédiaires et des acteurs essentiels dans la chaîne de co-production de services sociaux à tous les citoyens.

Dans ce cas, les collectivités doivent repérer les réutilisateurs potentiels, identifier leurs besoins de données, voire, créer une communauté locale et pousser l'innovation de nouveaux services ou produits. Pour attirer leur attention, des concours, des appels à projets, des hackathons et des ateliers de travail ont été organisés pour principalement la création d'applications web et mobiles à partir des données ouvertes.

Après la phase d'installation du portail open data la région Aquitaine, l'agence des initiatives numériques AEC, anime depuis septembre 2011 des ateliers de travail au sein desquels collaborent les professionnels du web et du numérique. Selon les organisateurs, ces ateliers ont contribué au repérage des acteurs du territoire intéressés par l'OpenData. L'OpenDataLab propose « de réunir des chercheurs, des collectivités, des startups et des grandes entreprises pour réfléchir ensemble au développement de services innovants basés sur la réutilisation des données publiques libérées par les collectivités locales. Cet espace de travail collaboratif permet de tester des concepts, valider des modèles économiques et juridiques, développer des maquettes et des prototypes, ou encore d'expérimenter des services et analyser des usages »²⁵⁰.

²⁵⁰ Voir : Chantry, Marion. "L'OpenDataLab : Un Lieu de Réflexion et de Tests En Mode Collaboratif." *Aquitaine Europe Communication. Agence Des Initiatives Numériques*, 2012. <http://www.aecom.org/Auberge-Numerique-incubateur-AEC/L-OpenDataLab>.

L'organisation des concours est un moyen répandu pour accélérer le développement d'outils ou de visualisations de données. Le Hackathon est un autre modèle de compétition pour les programmeurs qui consiste à développer dans un temps court (un jour ou plus) des applications en utilisant les données. Hack Data PACA définit un hackathon comme « un moment créatif intense qui réunit des développeurs informatique, des designers, des usagers, des "créatifs", etc., pour réaliser collaborativement des services/applications numériques : sous la forme de maquettes, de prototypes voire d'applications clé en main »²⁵¹. Des prix sont alors attribués aux projets gagnants. Des concours ont été organisés dans un grand nombre de collectivités dont le département de Saône et Loire (concours 2012), Toulouse Métropole (« Défi numérique » 2012, 2013 et 2014), ville de Montpellier, la région PACA, la région Aquitaine, la Communauté urbaine de Strasbourg, Rennes Métropole. Nous citons quelques exemples de ces activités :

Le département de Saône et Loire a organisé un concours en 2012. Les huit applications lauréates ont reçu des aides financières du Département entre 5,500 et 500 euros en favorisant les « jeunes créateurs ». Dans le cadre du programme Montpellier territoire numérique, « Montpellier Innovation » est un dispositif d'aide à l'innovation par le soutien des projets sous forme de subvention. Le troisième projet, "Mettre en scène ma ville" a été lancé en 2013 pour « soutenir les innovateurs et créateurs proposant de restituer, sous formes graphiques, affiches ou applications, les données informatiques du territoire »²⁵². Le premier appel en Mai 2012, aussi en relation avec l'open data, avait pour but la création de services innovants utilisant les données Opendata. Le projet a sélectionné neuf projets dont cinq qui ont bénéficié d'un soutien financier de la Ville et de ses partenaires pour un montant global de 20 000 Euro. Toulouse Métropole organise chaque année depuis 2012 les concours « Défi

²⁵¹ "Un Hackathon Késako ?" Accessed July 13, 2014. <http://hackdatapaca.net/post/25996058806/hackathon>.

²⁵² "Appel À Projet - Données Publiques Ouvertes - Montpellier Territoire Numérique." *Montpellier Territoire Numérique - Données Publiques Ouvertes*. Accessed July 19, 2014. <http://opendata.montpelliernumerique.fr/Appel-a-projet>.

Numérique » afin de favoriser la réutilisation des données mises en ligne sur son portail.

La Région Provence-Alpes-Côte d'Azur a organisé en 2013 un concours d'usages et de services numériques innovants liés à l'Open Data. L'objectif était de valoriser et soutenir les porteurs de projets qui inventent des projets d'applications, de services ou de visualisations de données interactives utilisant les données ouvertes du territoire. Les projets lauréats recevront une aide financière et/ou un accompagnement visant à accélérer le développement de leurs projets. Le concours a été doté d'une somme de 100 000 euros répartie sur dix projets.

Rennes Métropole avec ses partenaires a organisé un concours d'applications, qui s'est déroulé du 1er octobre 2010 au 30 mars 2011. Ce concours était doté de 50 000 € de prix et a récompensé des services et projets élaborés sur tous types de supports numériques (web, smart phone, téléphone mobile, autres supports) à partir des données publiques ouvertes. Le slogan du est « Fabriquer collectivement les nouveaux services numériques en PACA avec l'Open Data ! »

Benoit Deldicque, ingénieur spécialiste de la mobilité, et gagnant du prix « développement durable » au concours du département de Saône et Loire, partage son expérience suite aux 5 concours organisés par différentes collectivités auxquels il a participé. Dans une stratégie de relations publiques, il met en évidence les avantages de l'implication de l'équipe open data dans une relation avec les réutilisateurs : « La relation humaine, j'insiste, est un facteur important dans le développement de services numériques entre une collectivité et les personnes "travaillant" pour elle »²⁵³. la réactivité et la disponibilité du chef de projet open data ainsi que faire un retour aux réutilisateurs stimule la coopération entre les parties. « Le point intéressant avec la région PACA, c'est la réactivité et la disponibilité de Sophie Tendeiro qui a su répondre aux mails de façon humaine et

²⁵³ Voir le blog de Benoit Deldicque : "Concours Opendata: Les Collectivités S'engagent (ou Pas)." *Benoit Deldicque*, December 12, 2013. <http://www.benoit-deldicque.com/blog/concours-opendata-les-collectivites-sengagent-ou-pas/>.

non robotique ou automatisée. Dernière remarque, c'est le seul concours auquel j'ai participé qui se permet de me contacter pour m'annoncer que les jeux de données inexistantes précisés dans mon dossier sont disponibles et qui se permet de me relancer sur l'avancement des évolutions de l'application ».

Ces activités de communication et de promotion de la donnée nécessitent selon nous de les placer dans le contexte agité des usages sociaux des données. Comment qualifier les réutilisations issues de la logique de promotion ? Comment les insérer dans une logique plus pérenne d'usage ? Et plus globalement, quelles sont les éléments fondamentaux de la réutilisation ? Ensemble d'interrogations que nous proposons maintenant de traiter dans la deuxième partie de ce chapitre.

10.3 Les usages sociaux des données ouvertes

La question de l'usage et le degré de réutilisation des données publiques ouvertes sur les portails open data est au cœur du mouvement open data. Elle reflète l'approche centrée sur la réutilisation des données qui met en question l'objectif de l'ouverture pour les collectivités et confirme l'utilité des données ouvertes au service des usagers. Généralement, « les recherches sur l'usage sont nées d'interrogations sur les opérations des usagers afin d'atteindre une meilleure connaissance de leurs pratiques, manière de faire, comportement, habitudes, modes d'opération ou schémas d'action » (Le Coadic, 1997). Josiane Jouet (1993, p.371) fait d'ailleurs la distinction entre l'usage, la simple utilisation d'une technique, et la pratique qui « recouvre non seulement l'emploi des techniques mais également les comportements, attitudes et les représentations des individus qui se rapportent directement ou indirectement à l'outil ». Dans le domaine de la sociologie des usages, la notion d'usage, selon les contextes d'analyse et les cadres théoriques mobilisés, renvoie également à un continuum de définitions allant de l'appropriation à l'invention de nouveaux usages en passant par l'interactivité des usagers (Vidal, 2012). L'approche de l'appropriation, souligne Millerand (1998), consiste à analyser comment se constituent des usages différenciés selon les groupes sociaux, notamment à travers la signification de ces usages par les usagers eux-mêmes.

L'appropriation fait référence à des oppositions entre usages prescrits et usages réels et donne une signification des usages, en s'intéressant aux représentations et aux valeurs qui s'investissent dans l'usage d'une technique ou d'un objet.

Le contexte général maintenant posé, notre approche d'analyser les usages de données publiques consiste à mesurer l'usage des données publiques ouverte, montrer les logiques et les éléments fondamentaux de l'usage de données.

10.3.1 Mesurer l'usage des données publiques ouvertes

Une des principales limites à l'étude de l'usage des données ouvertes est qu'il est quasiment impossible de déterminer et d'identifier tous les usagers des données ouvertes sur le web. Nous sommes face à un principal avantage et défaut de l'open data. Le principe de la « non-discrimination » assure un accès non discriminatoire aux données, ce qui signifie que toute personne peut accéder aux données à tout moment sans s'identifier ou fournir des justifications pour le faire. Au sens large, le profil du réutilisateur, l'objectif de la réutilisation et le produit ou service dérivé ne sont pas identifiés. Les statistiques issues du système de gestion des données permettent de quantifier le nombre de téléchargement d'un fichier, le nombre de visites, le nombre de pages vues, mais n'enregistre pas des données autour des pratiques de réutilisation. Jérémie Valentin considère que « le principal avantage de l'open data (ne rien demander au téléchargement, pas d'inscription obligatoire) est pour nous les gestionnaires son principal défaut. Nous ne savons pas exactement qui et pourquoi le fichier a été téléchargé et dans quel but »²⁵⁴. Ainsi, le premier retour capté par les chefs des projets open data sur l'usage des données est basé sur le nombre de téléchargements des données. La « mesure de l'usage », en comptant le nombre de fois un fichier est téléchargé, semble le seul indicateur pertinent pour mesurer et évaluer l'usage. Cet indicateur est bien suivi et calculé par les chefs des projets. Après la mise à disposition en open data et en temps réel des résultats des élections municipales 2014, Valentin ajoute que :

²⁵⁴ Entretien le 13 février 2014

« d'un point de vue comptable, cette opération a plutôt bien fonctionné, car le site open data a enregistré 7.362 visites dimanche soir contre généralement une soixantaine ». La deuxième limite est liée au mode de diffusion proactif des données. Contrairement au mode réactif par accès sur demande, les usagers ne sont pas connus et leurs besoins ne sont pas exprimés. En plus, la réutilisation des données ouvertes est structurée à deux niveaux : directe et indirecte. Le réutilisateur directe est l'utilisateur qui utilise les données brutes en tant que source primaire tandis que le réutilisateur indirecte utilise les produits dérivés des données brutes, donc les produits secondaires, d'où la difficulté de repérer les différents usages et usagers des données publiques.

Dans une initiative de détecter et recenser les projets ou produits dérivés des données ouvertes, les collectivités ont encouragé les réutilisateurs à raconter ce qu'ils font avec les données. Les portails open data, sous la rubrique « Applications » répertoire les applications mobiles et web créées à partir des données, la catégorie de réutilisations la plus visible de l'open data selon Simon Chignard²⁵⁵.

Le nombre de téléchargement et le recensement des applications, ne sont pas suffisant pour apaiser les inquiétudes perçues par les chefs de projet. Selon Sandrine Mathon²⁵⁶ : « l'absence d'un standard clair des usages ralentira encore plus le processus d'ouverture en attendant que les usages se manifestent ».

10.3.2 Les logiques d'usage

L'expérience des collectivités avec les usagers est guidé par deux logiques : 1) la logique de la promotion et 2) la logique de la négociation

La logique de la promotion de la réutilisation est une démarche de provocation à l'usage de données. Dans ce cas, nous sommes dans une situation d'offre de données par les collectivités qui cherchent à

²⁵⁵ Chignard, *Open data*. Op.cit

²⁵⁶ Entretien le 4 avril 2014

dynamiser leurs usages. Pourtant les collectivités reconnaissent deux défis :

1. Une offre de donnée restreinte car elle dépend de l'existence des données et l'absence de la culture d'ouverture dans les administrations²⁵⁷
2. La difficulté de repérer les réutilisateurs et d'identifier la nature de leurs besoins.

Ainsi, les collectivités ont dû anticiper les usages en fonction de leur offre limitée de données. La stratégie de promotion des collectivités, étudiée en phase 5, consiste essentiellement en la création d'une communauté de développeurs et graphistes et à les pousser à développer d'applications web et mobile et la création de visualisations. Ce modèle d'usage a fait ses preuves selon les chefs de projets open data – la création de plusieurs applications mobiles et des visualisations - mais le focus sur un seul modèle d'usage accéléré par des concours nous conduit à questionner la pertinence du modèle, à s'interroger sur sa finalité, sur sa viabilité et son utilité. En effet, ce modèle est basé en grande partie sur une seule source de données qui est les données ouvertes par les collectivités ce qui rend le champ d'application relativement limité. La structure et la nature des concours posent aussi des questions sur la viabilité et la pertinence des applications produites. Le règlement du concours *Défi Numérique* de Toulouse Métropole, précise les critères de sélection : « la pertinence, efficacité du service proposé utilisant les données publiques, l'originalité, caractère innovant du projet ou de la réalisation, et la viabilité de la proposition »²⁵⁸. Les applications sont sur plusieurs domaines au sein de plusieurs petits projets plutôt de quelques grandes. L'aspect financier et pérenne du projet se révèle un point très important surtout après le concours. Est-ce que les porteurs de projets garantissent la mise à jour des applications ? En observant les projets issus des concours, nous avons remarqué que certains des lauréats n'assurent plus le bon fonctionnement de l'application. Nous citons l'application web Striim qui a remporté le grand prix du jury au

²⁵⁷ Voir chapitre 5

²⁵⁸ Voir : "Concours Open Data Défi Numérique." *Data.toulouse-Metropole.fr*. Accessed July 19, 2014. <https://data.toulouse-metropole.fr/web/guest/concours-opendata>.

concours 2012 du département de Saône et Loire et le projet *Où vont nos impôts?* gagnant du prix Visualisation de données au concours de Toulouse Metropole #2 qui ne sont plus accessibles sur le web. L'application Calendev Concert Théâtre a remporté le Grand Prix de Toulouse Métropole, doté de 8.000 euros, pourtant elle n'est pas mise à jour après le concours. En plus, Cyril Mottier²⁵⁹, développeur d'applications mobiles, a questionné la qualité des applications créées durant des journées de hackathon.

La logique de négociation impose que les collectivités ajustent leurs offres (nombre, qualité, domaine) dans une démarche de concertation avec les réutilisateurs pour répondre à leurs besoins de données. La satisfaction des attentes des usagers se révèle de plus en plus importante pour préparer une offre adéquate avec les besoins des réutilisateurs. Il ne s'agit pas nécessairement d'élargir l'offre dans le sens de diffuser plus de données mais de constituer une offre pertinente en fonction des demandes directes : « des données sur mesure ». Les réutilisateurs sont dans la meilleure position pour déterminer quelles sont les données utiles et pertinentes et les mieux adaptées à leurs besoins. Pourtant, des limites persistent toujours comme l'existence des données et la volonté de l'ouvrir par ses détenteurs qui peuvent affecter l'usage. La prise en compte du public se fait par les échanges informels entre les réutilisateurs et les administrations au sein des manifestations, mais aussi par contact directe via les outils de communication installés par les collectivités via le portail open data, facebook ou twitter. L'open data comme toute autre nouvelle tendance et une nouvelle technologie au sein des administrations publiques n'atteint pas immédiatement ses potentialités. Elle nécessite des ajustements au milieu où elle sera intégrée, ce qui amplifie le problème de l'usage.

Les éléments de bilan de l'expérience des collectivités territoriales avec l'ouverture de leurs données reposent, d'une part, sur le nombre de téléchargement de jeux de données sur le site, et le dynamisme de la communauté open data autour des manifestations organisées par les collectivités, de l'autre. La « mesure de l'usage », en comptant le nombre de fois un fichier est téléchargé, semble le seul indicateur

²⁵⁹ Entretien le 28 novembre 2013.

pertinent pour mesurer et évaluer l'usage, vu que dans la logique d'open data, le profile du réutilisateur, l'objectif de la réutilisation et le produit ou service dérivé ne sont pas identifiés. Dans une phase de post-publication, nous classons les réutilisations observées et marquées par les chefs des projets soit dans le cadre des initiatives de la promotion, ou de la logique de négociation. La logique de promotion met les collectivités en situation d'offre de données mais se heurte à la difficulté de repérer les réutilisateurs et d'identifier la nature de leurs besoins. Les réutilisations sont contrôlées et surveillées par les collectivités dans le cadre des concours ou des projets d'appels organisés, pourtant l'offre de données reste restreinte et ne répond pas nécessairement aux besoins des usagers. En contre partie, la logique de négociation montre des réutilisations sur mesure. Ainsi les collectivités ajustent leurs offres (nombre, qualité, domaine) dans une démarche de concertation avec les réutilisateurs pour répondre à leurs besoins de données. La prise en compte du public se fait par les échanges informels entre les réutilisateurs et les administrations au sein des manifestations ou par contact directe entre les deux acteurs.

10.3.3 Éléments fondamentaux d'usage

La stratégie de marketing basée sur la promotion du projet, la culture des données et la promotion de la réutilisation, désigne comment les collectivités cherchent à accompagner les utilisateurs dans le passage nécessaire d'une prise en compte passive de la donnée publique disponible, à une prise en compte dynamique pour exploiter la donnée publique.

Pourtant, malgré les efforts des collectivités, les usages des données publiques dépendent d'un environnement et de connaissances considérés comme des variables sur lesquelles une collectivité ne peut pas bâtir pour contrôler la réutilisation de ses données. Le comportement d'un réutilisateur ne peut pas être prédit. Dans ce sens, la réutilisation des données fait ressortir quatre éléments :

- L'intérêt
- L'offre de données
- L'appropriation

- L'action

La notion d'intérêt est strictement liée au besoin du chercheur de données sur un sujet ou objet précis en relation avec un territoire. Certains usages montrent l'inconstance ou l'irrégularité des intérêts des réutilisateurs d'une collectivité à une autre. Par exemple, le fichier « Fontaine d'eau potable du Grand Lyon » est consultée 1859 fois par contre le fichier « Fontaines à boire » sur le territoire de Toulouse métropole est téléchargé seulement 17 fois. Les fichiers les plus téléchargés sur les portails des collectivités témoignent la diversification des intérêts (voir tableau 16). En s'interrogeant sur les raisons de l'usage de certaines données, Tendeiro se demande, par exemple, sur les besoins derrière le téléchargement de la liste des orgues recensés en région Paca, qui est parmi les fichiers les plus téléchargés sur le portail open Paca (2,768 téléchargement).

Tableau16. Les fichiers les plus téléchargés sur les plateformes des collectivités

Portail	Fichier	No. de téléchargement
Open data 71	Photos des points d'arrêts	37,329 (vus)
Montpellier Territoire numérique – données publique ouverte	Liste des Associations	10, 149
Open paca	Mobilité étudiante et jeunesse : année 2010-2011	6,469
Toulouse Metropole Data	Réseau cyclable et vert	7,731
Smart Data	<u>Orthophotographie 2012 du Grand Lyon</u>	5,106 (consultations)

La question de l'offre met en question le volume, la certitude, la qualité technique et la mise à jour des données. Par conséquent, l'usage des données dépend de ces éléments qui qualifient les données. En expliquant le processus de réalisation du mashup à partir de données ouvertes sur le web, Gauthier Poupeau, data architect chez Antidot, indique que la construction d'un nouveau produit dépend, d'une part, de la nature des usages possibles de l'objet

construit et, d'autre part, des données disponibles. Ainsi, la façon dont les utilisateurs se saisissent des données ouvertes pour définir les spécifications du produit fini est liée aux intérêts des usagers mais également à l'offre des données donc des données existantes. Ce rapport module la stratégie dont les utilisateurs perçoivent leur produit créé à partir des données ouvertes. La logique de négociation²⁶⁰ pourrait être une solution pour rapprocher l'offre des administrations de l'intérêt des usagers afin de parvenir à une solution réaliste.

Dans un premier retour d'expérience sur la réutilisation des données ouvertes, Romain Mazon, journaliste et rédacteur en chef web de la Gazette des communes, parle des limites qui persistent dans le processus d'ouverture des données publiques locale surtout au niveau de la quantité et la variété de l'offre de données. « Beaucoup, beaucoup de données ne sont pas encore accessibles dans des domaine tels que la santé et la culture ». Dans ce sens, Maxime Vaudano, data journaliste au Monde.fr, parle de la variété de sources pour collecter les données. Le journaliste interroge les gens et cherche des ressources de « données non ouvertes » via des fournisseurs internes dans l'administration. Selon Poupeau, les sources de données ne correspondent pas toujours à ces critères. Le critère de certitude des données est étroitement lié à la *confiance* dans la source qui publie les données. Alors seules les collectivités sont en position de garantir l'exactitude des données. Ce critère oppose les institutions publiques aux sources de données citoyennes. Malgré la richesse des réseaux sociaux, la qualité des données n'est pas toujours garantie. La qualité technique de la donnée correspond au degré de structuration des données. Les formats CSV, XSL, XML, et RDF sont les plus réutilisés.

La notion d'appropriation employée dans le contexte d'analyse des usages de données porte sur les connaissances et compétences propres à un utilisateur de données nécessaires pour trouver, manipuler et interpréter les données. La question des compétences requises pour faire un usage des données ouvertes, publiques ou non, se pose dans tous les environnements et à plusieurs niveaux. Les

²⁶⁰ Voir section 8.2.2

revendications de l'« open data pour tous », supposent que tout citoyen ait les capacités à exploiter les données publiques. En plus, le secteur économique et les entreprises cherchent, comme nous l'avons déjà évoqué au chapitre 1, le profil du « data scientist » qui puissent tirer avantage des données internes ou externes.

Dans monde de l'open data locale, l'enjeu de l'appropriation des données publiques est double : offrir des données de haute qualité technique, mais aussi qui touche la plus grande population possible. Pour préparer les citoyens à faire un usage des données publiques ouvertes, une formation à l'appropriation citoyenne des données publiques ouvertes ou « Data literacy »²⁶¹ permet d'acquérir des compétences dans la collecte, l'évaluation, la manipulation, l'analyse et l'interprétation des données. Nous avons déjà relevé la mobilisation des collectivités, associations et des éducateurs pour promouvoir la culture de la donnée et former le grand public. Contrairement aux produits/services traditionnels basés sur l'interprétation des informations publiques publiées dans des documents, rapport, etc. le grand public n'est pas familier ni avec les techniques d'exploitation des données ni avec les attributs des données publiques par rapport à ses besoins. L'appropriation des données est tout à fait différente de l'appropriation des informations (Information literacy). L'origine de cette différence réside principalement dans la grande complexité de la manipulation de données que tout autre type d'informations (Thompson et Edelstein, 2004).

Un questionnement se pose sur les capacités des citoyens à appréhender les données : Quelles sont les compétences minimales requises par un citoyen? En effet, la Fing au cours de deux ateliers Infolab : «*Peut-on démocratiser une culture de la donnée*» pendant le Connecteur Recherche à Futur en Seine, et «*Les compétences de la "Data Culture"*» pendant l'Open Data Week 2013 à Marseille a identifié quatre niveaux de pratiques dans l'expérience quotidienne des données : « la réception, la (ré)utilisation, la production et la diffusion », ainsi que sept champs de compétences associés :

²⁶¹ Voir : Calzada Prado, Javier, and Miguel Ángel Marzal. "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents". *Libri: International Journal of Libraries Information Services* 63, no. 2 (June 2013): 123–34. doi:10.1515/libri-2013-0010.

« Lecture-écriture (sens critique et culture visuelle), Culture numérique (maîtrise TIC et machine informatique), Information-documentation-communication (sens critique et techniques documentaires, stratégies), Droit (identification du statut juridique des données, des licences liées à la consultation, aux réutilisations), Statistique (calcul et traitement des données), Informatique (programmation), Design, graphisme (mise en forme et lisibilité, ergonomie des interfaces techniques) »²⁶². Le bilan du projet #dataviz_rra sur la data-visualisation des données publiques par les lycéens²⁶³ a montré les compétences acquises par les lycéens réparties sur cinq domaines d'application : 1) Travailler dans un environnement numérique évolutif, 2) Être responsable, 3) Produire, traiter, exploiter et diffuser des documents numériques, et 5) Communiquer, travailler en réseau et collaborer (voire figure 15).

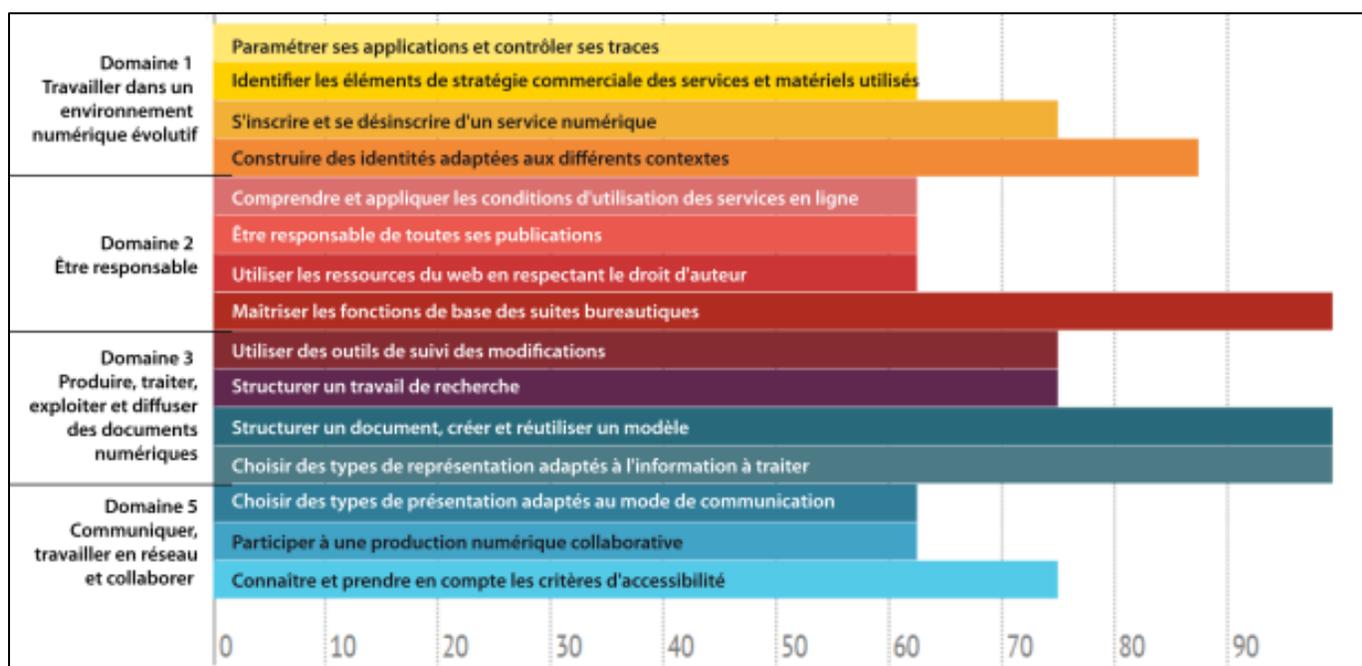


Figure 15. Les compétences acquises par les lycéens (Projet #dataviz_rra, 2014)

²⁶²Voir : “Les Compétences ‘Data.’” *FING*, September 2, 2013. <http://www.fing.org/?Les-competences-Datalang=fr>.

²⁶³ Voir : “Découvrez Ici L’intégralité Du Dispositif Pédagogique #dataviz_RRA | Fréquence Écoles.” *Fréquence Ecoles*, June 25, 2014. http://frequence-ecoles.org/2014/06/25/decouvrez-ici-lintegralite-du-dispositif-pedagogique-dataviz_rra/.

En complément des compétences déjà écrites, les connaissances ou l'expertise d'une part du domaine ou thématique des données ouvertes et d'autre part du fonctionnement des établissements publics sont également indispensables. Selon Tendeiro²⁶⁴, la spécificité et la granularité des données publiées sur les portails open data n'intéresse qu'un public très restreint de spécialistes. En donnant l'exemple du budget de la collectivité de la région Paca, Tendeiro, indique que seulement les experts comptables peuvent s'intéresser à ce type de données.

Malgré son importance, la démarche d'acculturation à la donnée s'avère difficile et longue. Une démarche alternative repositionne le citoyen/amateur de sa position directe avec les données ouvertes à une position de deuxième degré. Dans ce cas, le citoyen amateur a un statut de « simple consommateur »²⁶⁵.

Cette démarche consiste alors à solliciter, en premier, les « intermédiaires » qui sont les experts de la donnée (les développeurs, statisticiens, designers, journalistes de données ou tous autres individus possédant des compétences techniques et analytiques avancées) pour développer des services et informations à destination du citoyen/amateur. Cependant, même au niveau des intermédiaires supposés être des spécialistes de la donnée, des défis émergent qui transforment leur environnement et nature de travail. Selon Mazon²⁶⁶, l'approche technique du traitement et représentation graphique des données oblige le data journaliste à acquérir de nouvelles compétences. Alors, l'apprentissage aux outils informatiques est indispensable surtout lorsque les formations des écoles de journalisme développent chez les journalistes des compétences de collecte de données, d'observation des faits, et d'analyse critique et n'accorde pas autant d'importance aux compétences techniques (formation tableur, au développement informatique), indique Mazon.

²⁶⁴ Entretien le 21 juillet 2014

²⁶⁵ Voir à ce sujet l'article de Libertic. "Le Citoyen a-t-il Une Place Dans L'open Data ?" *Libertic*, February 17, 2012. <http://libertic.wordpress.com/2012/02/17/le-citoyen-a-t-il-une-place-dans-lopen-data/>.

²⁶⁶ Entretien le 19 février 2014

En étudiant, les projets de réutilisations des données, nous remarquons des environnements de travail collaboratif et interdisciplinaire. Ces derniers sont composés d'un groupe de spécialistes de plusieurs domaines. Ainsi la méthodologie de création de nouveaux produits ou services d'informations se base sur la complémentarité des compétences et connaissances, donc un renouvellement des modalités de coordination des activités productives. Par exemple, il existe une interdépendance entre le data journaliste et le data designer. Dans le cadre de son travail avec l'équipe du site informationnel OWNI (Objet Web Non Identifié), Marion Boucharlat explique que son job en tant que graphiste était de choisir le mode de représentation graphique des données en fonction du contexte et l'angle d'analyse de l'auteur de l'article. De même Romain Mazon, lui en tant que journaliste, il travaille avec des experts techniques pour valider les traitements de données.

Enfin, l'action est fondée sur la réutilisation des données ouvertes, dans une perspective de création de valeur à partir de l'usage des données publiques. La fonction de la créativité se conjugue à la combinaison de l'intérêt, la connaissance et les compétences. La démarche de la réutilisation pour créer une valeur ajoutée qui va de la collecte des données à la création de nouvelles connaissances, produits ou services qui renforce l'aptitude de l'individu ou l'entreprise à développer simultanément l'innovation et l'amélioration des processus de production. Nous présenterons à la fin de ce chapitre les modèles de réutilisation considérés comme des actions.

Conclusion du chapitre 10

Dans notre cas, il apparaît que la communication et la promotion des données ouvertes est une étape stratégique dans tous les projets open data. Dans l'étude de la stratégie de promotion de l'open data locale, nous avons pris en considération la diversité des données, à la fois aux niveaux des domaines et des propriétés techniques, ce qui élargit le public visé et oblige les collectivités d'utiliser une variété de techniques pour toucher les experts et le grand public. Cette stratégie est fondée sur le territoire. La dimension territoriale est fortement présente car les pratiques de promotion exercées par les collectivités étaient centrées sur le territoire en vue de créer une communauté open data locale.

En plus, trois orientations de techniques de promotion nous ont paru particulièrement fondamentales. La promotion du projet consiste à faire de la publicité sur le projet afin d'inciter les gens à aller sur le portail open data de la collectivité. La promotion de « la culture des données » est incontournable pour atteindre les objectifs définis par la collectivité et tirer pleinement parti des promesses de l'open data qui concerne tous les citoyens. C'est un long processus d'apprentissage qui s'est manifesté par l'organisation d'Infolab, Data Schools et des formations sur les données publiques. Sous la catégorie la « promotion de la réutilisation », nous avons classé les concours, appels à projets, hackathons et ateliers de travail organisés pour principalement pour la création d'applications web et mobiles à partir des données ouvertes.

Les résultats de nos analyses montrent aussi que l'usage conduit à une convergence autour de l'intérêt des réutilisateurs, l'offre des collectivités, l'appropriation des données et l'action de produire de nouveaux produits ou services. Ces éléments aboutiront à une réutilisation optimale ou marginale des données publiques ouvertes.

Chapitre 11. La communauté de réutilisateurs et les modèles de réutilisations

Dans ce chapitre, nous analysons tout d'abord le comportement des usagers face aux données publiques ouvertes. Nous identifions la communauté de réutilisateurs avant de proposer les modèles de réutilisations observés. Dans une approche plus économique, l'offre des données publiques par les collectivités sera opposée à la demande des usagers et au marché des données publiques. L'écosystème des usages est étudié par rapport à cette relation entre l'offre et la demande de données.

Ensuite, nous présentons une vue d'ensemble des réutilisateurs et de leur façon de réutiliser les données publiques. Ainsi, nous précisons le profil des réutilisateurs de l'open data locale et leur compétences nécessaires pour en faire un usage. Nous étudions également l'écosystème et les modalités de réutilisation en tenant compte du « processus de transformation » que subissent les données.

11.1 Les profils des réutilisateurs

Avant de présenter le profil des réutilisateurs, nous rappelons ici que la réutilisation des données ouvertes est structurée à deux niveaux : directe et indirecte. Le réutilisateur direct est l'utilisateur qui utilise les données brutes en tant que « source primaire » tandis que le réutilisateur indirect utilise les produits dérivés des données brutes, donc les « produits informationnels secondaires ». A cette étape, les « médiateurs », qu'ils soient journalistes, programmeurs, des statisticiens, ou designers, qui possèdent normalement des compétences techniques et analytiques, interviennent pour développer des services ou produits secondaires à destination du réutilisateur indirecte qui est souvent un amateur de la donnée.

11.1.1 Les développeurs et les graphistes : deux acteurs mis en vedette

La communauté open data locale la plus visible car médiatisée à travers des listes de diffusion, les réseaux sociaux en ligne, et les concours, est axé sur des jeunes ingénieurs en informatique, des développeurs, des graphistes, des freelances ou des étudiants. La grande majorité des projets des concours a été portée par des individus ou de petits groupes de personnes. Quelques petites entreprises ont remportées des lauréats à l'image de la société de conseil et services en informatique *Prometil* (issue des pépinières de Toulouse Métropole) qui a participé au concours Défi Numérique #2 et a remporté le Prix Digital Place pour l'application "One Bike".

La logique de promotion est marquée par deux faits : d'un côté, les réutilisateurs les plus exposés sont les développeurs et les graphistes indépendants et les startups en informatique. L'objectif est alors d'aider les jeunes innovateurs et les jeunes entreprises à exploiter les données publiques en les transformant en matière compréhensible par tout le monde. De l'autre côté, les compétences des réutilisateurs sont axés sur la visualisation et la technologie informatique.

Les développeurs indépendants

Les développeurs indépendants sont des citoyens qui contribuent à la construction d'outils qui facilitent l'interprétation des données publiques. Programmeurs ou développeurs web ou mobile, ils jouent un rôle important dans l'ouverture des données publiques. En s'intéressant aux données publiques, leur objectif est de créer des outils pour le bien public. Leurs réutilisations techniques sont appliquées dans un contexte civique d'intérêt public. Plus la communauté des développeurs est dynamique et motivée, plus l'usage des données publiques est grand. La communauté des développeurs en France a joué un rôle dans la phase post-ouverture. Par contre, comme nous l'avons déjà signalé, les communautés de « Civic hackers » aux Etats-Unis et en Grande Bretagne étaient des acteurs de pression car ils ont incité les administrations à l'ouverture.

Cyril Mottier²⁶⁷, développeur d'applications mobiles, a créé deux applications urbaines en utilisant des données ouvertes et « non ouvertes » sur les stations de VéloV à Lyon et à Paris. Il montre un engagement personnel pour développer de nouvelles compétences, montrer ce qu'il peut faire et d'apporter de nouvelles applications sans avoir un intérêt commercial car les applications sont téléchargées gratuitement.

Les graphistes

La data visualisation consiste à mettre en image des données chiffrées ou géographiques. C'est un mode de transformation visuelle des données pour les faire parler et les rendre compréhensible. Cette transformation nécessite de la part du graphiste des connaissances dans le domaine d'application pour pouvoir montrer et appréhender de manière synthétique les enjeux essentiels du propos. La visualisation est le résultat d'un processus de mise en visibilité de données brutes parfois complexes et dans des formats illisibles (chiffres, listes, coordonnées de géo-localisation par exemple). Les pratiques de data visualisation et les modes de représentations (Les diagrammes circulaires - les camemberts - les treemaps - les timeline - Les nuages de mots - les illustrations - photo visualisations) ne sont pas nouveaux, par contre ils gagnent plus d'intérêt dans le contexte plus large – la société – que dans un public restreint de spécialistes comme les chercheurs scientifiques et les journalistes hyper spécialisés. Marion Boucharlat²⁶⁸, graphiste indépendante spécialisée en design d'information et cartographie explique que le regain d'intérêt public en data visualisation est lié à deux niveaux d'effets :

- Dans le fond : C'est la manière dont la société mobilise les informations et les données. En effet, le mouvement open data a joué un rôle important à ce niveau. La mise a disposition d'un volume de données publiques par les collectivités et les organismes publiques ainsi que la promotion d'usages des données publiques par le grand

²⁶⁷ Entretien le 18 novembre 2013

²⁶⁸ Propos tenus par Boucharlat lors de la conférence *Dataviz – Infographies : Data + Lyceens / Rhone Alpes*. Lyon, 2014.

- public sont deux facteurs qui incitent à avoir recours à la data visualisation.
- Dans la forme : C'est le développement d'outils plus performants de data-visualisation. En plus, l'émergence d'outils gratuits du web permettant de produire des visualisations de données sans connaissance technique spécifique favorise à généraliser la pratique de visualisation de données par un plus grand nombre de gens.

11.1.2 Les journalistes : réutilisateurs classiques de la donnée publique

En plus des développeurs et des graphistes, d'autres acteurs ayant une expérience classique dans la réutilisation des données publiques ont fait part de leur intérêt pour l'open data. Nous désignons en particulier les data-journalistes. Pour Romain Mazon²⁶⁹, journaliste et rédacteur en chef web de la Gazette des communes, la donnée ouverte est une matière première pour le journaliste. Par la mise à disposition des données des collectivités, le mouvement open data local contribue à l'accessibilité des données publiques « au plus petit échelon territorial », affirme-t-il. En plus, l'ouverture des données permet aux journalistes de renouveler leur champ d'investigation par la reprise d'autonomie vis-à-vis des calendriers officiels. Elle est également un moyen de vérification des faits (fact checking) ce qui apporte une puissance de démonstration par les chiffres. « Alors que le journaliste « traditionnel » protège ses sources, le data-journaliste met un point d'honneur à les mettre à disposition du public » (Chailloux, 2012). Les données ouvertes permettent une nouvelle forme d'exploitation et peuvent apporter une valeur ajoutée aux informations classiques (texte, image, son). « Le traitement de données et la datavisualisation renouvellent l'expérience du lecteur et donnent une nouvelle valeur ajoutée », selon Mazon.

En effet, les données sont transformées par le journaliste en une nouvelle forme de représentation. Plusieurs exemples d'articles basés sur l'open data prennent souvent la forme d'une combinaison d'une analyse sous forme de texte accompagnée d'une visualisation,

²⁶⁹ Entretien le 19 février 2014

interactive ou non. Maxime Vaudano, data journaliste au Monde.fr affirme que : « le data journalisme n'est pas nécessairement que de la data visualisation. Avec le data journalisme, on utilise les données pour rédiger un article, une visualisation ou les deux ensembles, ce qui permet d'améliorer l'efficacité en terme de contenu et en une présentation plus esthétique »²⁷⁰. Les figures 16, 17 et 18 montrent des exemples d'articles du magazine « la gazette des communes » et du journal « le monde » basés sur des données publiques ouvertes.



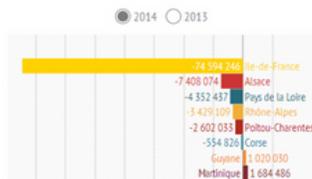
Figure 16. Article de la Gazette des Communes basé sur des données fournies par data.gouv.fr

²⁷⁰ Propos tenus par Vaudano lors de sa présentation sur "En Quoi Les Données et Les Dataviz Changent Le Métier D'un Journaliste ?". *Dataviz – Infographies : Data + Lyceens / Rhone Alpes*. Lyon, 2014.

Fonds régional CVAE : la contribution de l'Île-de-France explose en 2014 !

Publié le 31/03/2014 • Mis à jour le 02/04/2014 • Par Raphaël Richard • dans : [Actu experts finances](#), [Données interactives finances](#)

Fonds CVAE des régions



© La Gazette

RÉFÉRENCES

Consultez notre [application interactive « A qui profite la péréquation ? »](#) (données 2013)

Le fonds de péréquation régional sur la CVAE est passé de 22,8 millions en 2013 à près de 93 millions en 2014, selon les chiffres que vient de mettre en ligne la DGCL. L'Île-de-France est frappée de plein fouet par cette hausse, sa contribution passant de 11,7 millions l'an passé à 74,5 millions cette année !

Très modeste en 2013 - année de sa création -, comme le souhaitaient les régions, le fonds de péréquation sur la cotisation sur la valeur ajoutée des entreprises (CVAE) vient de prendre une toute autre dimension avec quelque 92,940 millions qui vont être redistribués en 2014.

Figure 17. Article paru dans la Gazette des Communes basé sur des données ouvertes par Direction Générale des Collectivités Locales

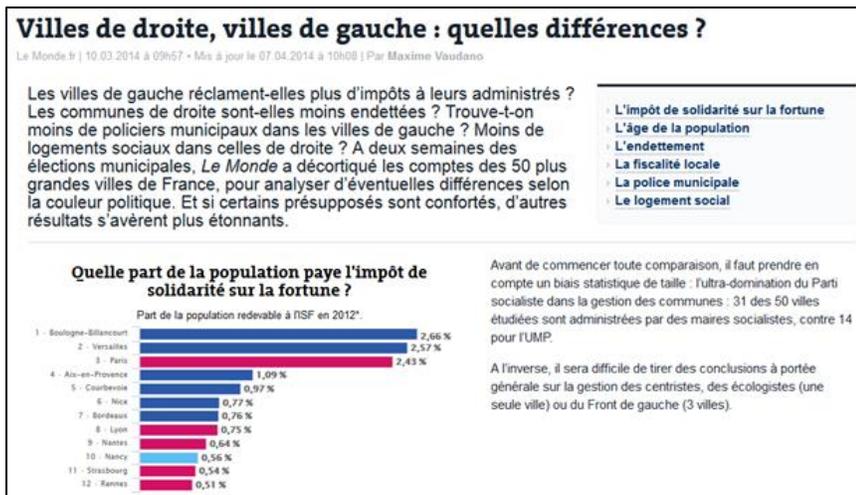


Figure 18. Article du Monde.fr basé sur des données ouvertes par le Ministère de l'Économie et des Finances

11.1.3 Les acteurs économiques : une réutilisation encore floue

Les acteurs économiques se trouvent à la périphérie du réseau de l'open data et le modèle économique issu des données ouvertes reste encore flou. Le premier argument invoqué, par les militants de l'open data, pour expliquer cette situation est que le modèle reste à trouver surtout que nous sommes au début de la phase de post-ouverture des données. Donc, il est encore trop tôt pour en voir les résultats économiques. Dans quelques années, les chercheurs auront plus de recul pour observer les différents écosystèmes liés à l'open data. Selon Sandrine Mathon²⁷¹, sur un territoire local très étroit, le marché des données publiques est rapidement saturé et les idées d'usages innovatrices sont limitées. D'autres facteurs expliquent selon nous l'absence des acteurs économiques. L'exploitation des entreprises des données publiques est certainement sous le seuil de visibilité des réutilisations. Logiquement une entreprise ne va pas dévoiler ses usages et ses sources de données. En plus, le modèle de réutilisation sur le principe du « bien commun » freine l'investissement des entreprises dans les données publiques. Les réutilisations sous la licence ODBL obligent les entreprises à placer les productions réalisées à partir des données gratuites sous la même licence. Quel sera l'intérêt pour une entreprise d'investir dans un projet qui sera forcément ouvert à ses concurrents ? En outre, dans le secteur des informations et des données, les entreprises qui offrent des services basés sur l'exploitation directe et la vente de produits et services de données publiques restent très minoritaires. En effet, il est difficile d'étudier le modèle économique de la réutilisation de l'open data. Il est important de faire la distinction entre les données « moyen » et les données « objets » par rapport aux entreprises. Les données « moyens » sont l'ensemble des données collectées et traitées par une entreprise afin de lui permettre d'améliorer ses services et produits ». Dans ce cas, les données publiques ouvertes peuvent enrichir les analyses de décideurs pour une prise de décision mieux informée sur les enjeux territoriaux. Dans une stratégie globale de traitement de données, l'entreprise fusionne des données provenant

²⁷¹ Entretien le 4 avril 2014.

de sources différentes (internes et externes, y compris les données publiques) pour exploiter pleinement le potentiel des données. Les spécialistes de données incorporent et intègrent les données publiques ouvertes dans la substance de leurs organismes, dans leurs bases de données, pour innover et fournir des produits, des biens et des services de manière à en tirer profit. C'est à ces conditions que la réutilisation des données publiques contribuera à la croissance économique de l'entreprise. Selon Jérémie Valentin, l'open data apporte des « revenus indirectes » à l'entreprise, pourtant ce modèle est encore difficilement quantifiable.

Les données « objets » sont l'ensemble des données collectées, traitées afin de les réutiliser et rediffuser en tant que des services et produits de données. Le principe « Data as a service » (DaaS), apparu avec la montée de Cloud Computing, est un modèle économique basé sur des services de données de qualité, traitées, nettoyées et enrichies. Les données sont utilisées par des applications composites (mashups), mais elles sont maintenant de plus en plus utilisées commercialement. Par exemple, Data publica et Transpacity sont deux entreprises qui proposent différents services sur le marché de la donnée publique.

Selon Data publica, la valeur ajoutée des services de données, , réside dans la quantité et la richesse des données venant de sources variées, la couverture de l'ensemble du territoire, la mise à jour, et la visualisation claire et interactive. Les clients de Data Publica sont « tous les acteurs ayant à produire des analyses, prendre des décisions concernant les territoires ou développer des services sur ces territoires : journalistes, décideurs publics, consultants, société de services Internet, startups, etc. ». Le modèle d'offre de Data publica est double : une grande partie de jeux de données sont accessibles gratuitement, par contre des services plus avancées sont payants comme « Les tableaux de bord des territoires (TBT) ». Ce ont des fiches multi média interactives, accessible en ligne, et construits sur les bases de données, montre des visualisations sur les territoires

géographiques sous des dimensions géographiques, démographiques, économiques ou plus généralement sociétales »²⁷² (voir figure 19).

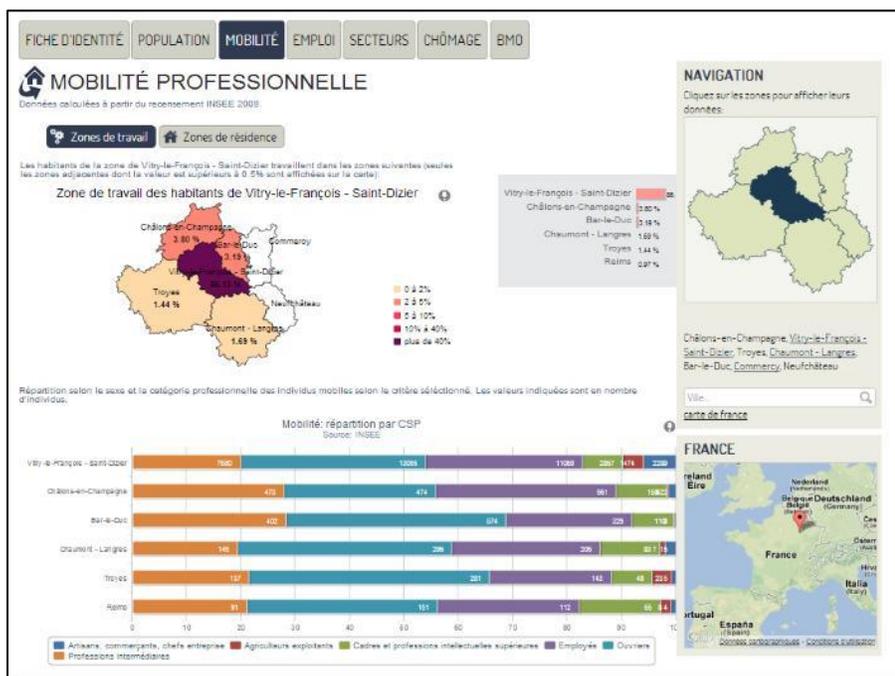


Figure 19. Exemple de Tableaux de bord des territoires (TBT) (Data Publica, 2014)

Du point de vue des clients, ils reçoivent un service incapable de produire eux mêmes à un coût acceptable et qui est adapté à leurs besoins. A l'origine Data publica, créée en 2011, était très focalisé sur la partie open data mais en 2014, la société a adopté une nouvelle approche en s'ouvrant au « Big data ». « L'open data et les données publiques en général sont bien entendu important pour notre métier, mais ce ne sont pas les seules : les données du Web, les données des réseaux sociaux, les données des média (presse, sites spécialisés, forums, etc.) sont aussi importantes », affirme François Bancilhon (2014), le fondateur de Data Publica.

Transpacités constitue un autre exemple des entreprises réutilisateurs de données publiques. Cette société privée, créée en 2006 collecte, structure et analyse les données sur les collectivités dans le but d'offrir des services informationnels et de marketing aux opérateurs

²⁷² Voir : “Les Tableaux de Bord Des Territoires.” Accessed July 20, 2014. <http://www.data-publica.com/content/les-tableaux-de-bord-des-territoires/>.

privés qui s'intéressent à investir dans quatre domaines publics : les déchets, l'eau, le transport en commun et l'électricité. Dans la logique de Transpacités, l'open data est un facilitateur pour traiter d'avantage les sources d'information publique. « Les données collectées et analysées ont une forte valeur économique sur le marché »²⁷³, affirme Didier Gorrion, Associé fondateur de Transpacités. Ses services de données et de Benchmarking aux opérateurs privés sont essentiels pour une « bonne prise de décision » et « réactivité » sur le marché, explique-t-il. Dans un premier retour sur son expérience de réutilisation des données ouvertes, Gorrion indique que les fichiers structurés forment seulement 20 % de l'ensemble de données non structurées, comme les rapports, documents, l'historique des appels d'offre, les résultats des élections, et les articles de presses, collectées pour alimenter la base de données de Transpacités. Ainsi les clients auront accès aux données brutes collectées par un abonnement en ligne à la base de données. Nous avons pu également repérer quelques sociétés d'informatique qui offrent des logiciels de traitements et de visualisation de données publiques. Nous citons Dataville²⁷⁴ et Antidot. Dataville est une plateforme de visualisation de données, permettant d'exploiter des données locales. Un accès gratuit permet d'accéder à un nombre limité de visualisations. Le service premium par abonnement mensuel permet l'accès à l'intégralité de la plateforme. Antidot, editeur de logiciel, a développé l'application web « France musée » accessible gratuitement. Créée à partir des données ouvertes et des données non structurées sur le web, l'application servait principalement pour créer un prototype des services de l'entreprise.

Le tableau ci-dessous résume les écosystèmes des réutilisateurs des données publiques. Divisé entre offre gratuit ou payant ou les deux ensembles, chaque acteur a créé son propre écosystème. Les développeurs indépendants consacrent leurs temps et effort personnel. Certains parmi eux ont bénéficié de subventions directes des organismes publiques suite aux concours organisés par les collectives. Les entreprises qui offrent des services gratuits

²⁷³ Entretien le 13 décembre 2012.

²⁷⁴ <http://www.dataville.fr>

(freemium) et les media ont des revenus de services auxiliaires payants (service premium) ou de la publicité. Un service premium est généralement un service plus élaboré et plus riche de données que la version de base. Le service premium est payant alors que la version de base est gratuite. Ce modèle se rencontre fréquemment sur internet. Enfin les entreprises qui offrent uniquement des services payants s'autofinancent par les abonnements mensuels.

Tableau 17. L'écosystème des réutilisateurs des données publiques

Types de réutilisateurs	Produit gratuit/payant	Type de financement	Produit ou service / producteur
Entreprise privée	Service gratuit	Financement par des services auxiliaires (Dans ce cas des services en relation ne sont pas des services essentiels de l'entreprise)	La plateforme Musée de France / Créé comme un prototype par Antidot
	Service Freemium	Financement par les services premium	Tableaux de bord des territoires / Data publica Plateforme de visualisation / Dataville
	Service Freemium	Financement par les services premium par abonnement et par la publicité	Articles gratuits des media comme LeMonde.fr
	Produit payant	Financement par tarification	Bases de données / Transparency
	Produit payant	Prix de vente - abonnement	Media payant comme Le magazine Gazette des communes
Développeur	Un produit gratuit	Pas de financement Motivation personnelle	Application VéloV / Cyril Mottier
	Un produit gratuit	Financement par des subventions d'institutions publiques comme les applications d'un concours	Applications des concours
Associations	Un produit gratuit	services et de prestations	Nos deputes.fr du collectif

		fournies par l'association, de subventions, de dons	regard citoyen
--	--	---	----------------

11.2 Les modèles de réutilisations

Notre intérêt d'étudier les réutilisations visent principalement à définir des modèles pour le déploiement, la gestion et l'utilisation des données publiques publiées par les collectivités. Nous essayons de montrer le processus de transformation de la donnée par chaque usage, le type de traitement qu'elle a subi et l'acteur intervenant. La réutilisation des données publiques est liée à un « processus de traitement » qui devrait apporter une « valeur ajoutée » aux données utilisées. Selon le rapport de Gaudrat (1992), sur la commercialisation des données publiques, les données automatisée, commentée, systématisée, compilée, et documentée font l'objet de plusieurs types de « plus values » :

1. une plus-value intellectuelle,
2. une plus-value documentaire et
3. une plus-value technique.

Lorsque les données brutes sont commentées, elles portent une valorisation intellectuelle. La sélection, la compilation, le classement, et la corrélation des données constituent une plus-value d'ordre documentaire. La plus-value technique consiste à présenter les données une fois collectée sur un support autre que son support d'origine à des fins d'adaptation et/ou de communication et/ou de conservations particulières. Cette plus-value apporte une plus grande facilité, quantitative ou qualitative, de traitement, pour ceux qui disposent des systèmes de gestion comme une base de données. Avant toute valeur ajoutée de type intellectuelle, documentaire ou technique, le processus de collecte de données brutes constitue déjà une valeur ajoutée. Les données collectées et présentées comme données brutes ou données source forment un état primaire antérieur à tout processus de traitement, mais comme l'indique Gaudrat (1992) : « Le seul fait de collecter et de rassembler des informations isolées constitue déjà une plus-value » puisque par définition « l'information existe déjà au delà de ce rassemblement, mais le rassemblement apporte un plus : une chose est de savoir la température qu'il fait aujourd'hui à midi ; une autre est de savoir la température qu'il a fait chaque jour à midi, dans ce lieu, depuis un demi-siècle ». Il est important de signaler la difficulté de séparer le processus de collecte, des traitements documentaire et technique. En s'inspirant de la

typologie de Gaudrat relative aux traitements, en effet la réutilisation de données publiques correspond à deux fonctions qu'on peut classer selon qu'elle cherche à construire une connaissance, donc une valorisation intellectuelle, ou qu'elle cherche à construire un nouveau produit informationnel. Nous considérons l'activité de production de produits informationnels dérivés des données publiques dans le contexte de l'open data comme une application particulière de l'activité de réutilisation des données publiques. L'approche de réutilisation existait déjà dans le domaine des logiciels. Comme nous l'avons déjà signalé dans la section 1.1, la philosophie du logiciel libre propose le code source comme une base à une nouvelle réutilisation. Ce dernier type de réutilisation a été élargi ensuite aux données publiques dans des documents structurés. Dans la production de nouveaux produits informationnels, les utilisateurs se trouvent dans une situation de création *ex materia*. La création à partir de données préexistantes prend la forme de plusieurs types de réutilisations. En effet, nous avons repéré plusieurs types de réutilisations qui varient de la simple recherche d'une donnée précise aux tâches plus complexes comme le mixage de données de différentes sources.

11.2.1 La recherche d'une donnée précise

Un individu peut faire un usage précis en cherchant une donnée ou un jeu de données précis. Parmi les premiers jeux de données téléchargés sur les portails open data des collectivités nous trouvons des données pratiques que les individus peuvent consulter, par exemple, pour localiser une salle de spectacle, (Guide des salles de spectacles » de la région paca, téléchargé 5,065 fois), se renseigner sur les zones scolaires de la commune de Montpellier (Carte scolaire des écoles maternelles) ou chercher les circuits des balades vertes en Saône-et-Loire (Balade verte, vu 2361 fois). Ce premier type d'usage appartient à la catégorie de réutilisation qui contribue à relever la plus-value intellectuelle par un usage direct des données publiées par la collectivité. En dépit des raisons qui sous-tendent le grand nombre d'utilisateurs illustré par le nombre de téléchargements de tels jeux de données, Tendiero explique cela par le fait que les jeux de données sont accessibles et trouvables par une simple recherche sur les moteurs de recherche.

11.2.2 La production graphique

La transformation des données en visualisations statiques ou dynamiques est une réutilisation courante et encouragée dans le monde de l'open data. La véritable valeur ajoutée réside dans la capacité des visualisations à rendre les données plus compréhensibles. Dans un processus technique, ce type de réutilisation transforme les jeux de données en visualisations. La vulgarisation des données est une pratique exercée en premier lieu par les producteurs de données dans le cadre d'une stratégie qui cible le grand public. En fait, comme nous l'avons déjà constaté en étudiant les portails open data des collectivités, Open data 71 et Smart Data offrent des outils de visualisation sur leurs portails. Une technologie améliorée utilisée dans la modélisation des portails permet une visualisation automatique des données soit en diagrammes circulaires, camemberts ou courbes dans le cas du portail Open data71, soit en cartes pour les données de géo-localisation dans le cas du portail Smart Data.

La transformation des données en visualisations nécessite un traitement technique assuré par des logiciels dans un processus automatique. Pourtant, dans certains cas les données subissent un processus d'extraction pour simplifier la présentation graphique des données. C'est le cas du portail Open data 71 qui publie des jeux de données « abrégés » pour les transformer en visualisations.

Les collectivités qui n'offrent pas eux-mêmes un service de data visualisation se contentent à encourager les usagers à produire des graphes à partir de leurs données. L'organisation des concours et des ateliers de formation présentent au grand public les compétences ainsi que les outils disponibles pour créer des data visualisations.

De même nous trouvons des entreprises qui s'investissent dans le domaine de visualisation des données publiques. Paru comme un nouveau marché prometteur, Dataville est un exemple de sociétés privées qui offrent des services payants de visualisation des données publiques ouvertes.

11.2.3 La production écrite

L'interprétation des jeux de données afin de rédiger un article ou un rapport est une autre forme d'usage observée. Au moment où certains jeux de données existants sont réutilisés dans la reconstruction de nouveau produit

informationnel, la valorisation intellectuelle des données se traduit par une transformation des données structurées en textes. La manipulation et l'analyse des jeux de données créent des nouvelles formes de produits informationnels comme les rapports et les articles de presses. Cette réutilisation est pratiquée par les data-journalistes. Les exemples déjà cités de production journalistique illustrent bien ce type d'usage. Les articles de presse se basent sur des jeux de données publiques ouvertes par les institutions publiques. Nous observons une combinaison de texte et de visualisations dans le même article, donc un double processus de traitements technique et intellectuel qui transforme les données. De même, les cabinets d'analyse du territoire, des sociétés de conseil et de Benchmarking comme Transpacity offrent des services de consultation en se basant sur leur analyse des données publiques.

11.2.4 La production d'interface

Dans un système de traitement de données, l'interface de recherche et de présentation des données agit comme intermédiaire entre le logiciel ou la base de données et l'écran d'affichage. Une application web ou mobile permet aux utilisateurs d'interroger la base de données dans une interface toujours plus conviviale. Cette valeur ajoutée technique apportée aux données publiques est une spécialité des experts informatiques de programmation.

11.2.5 L'extraction ou la combinaison des données

Ce type d'usage des données ouvertes consiste à extraire des fragments de données présentées sous forme de jeux de données pour créer un nouveau jeu de données plus « spécifique ». Cette pratique est observée quand la couverture du jeu de données principal est très large. (plusieurs objets et attributs). La centralisation de production des données sur les territoires locaux dans des bases de données nationales qui agrègent toutes les données sur tous les territoires explique ce type de pratique. Par exemple, l'extraction des données en relation seulement avec le département de Saône et Loire pour créer un nouveau jeu de données est faite à partir des jeux de données de l'Insee qui couvrent toutes les collectivités de France. Dans ce cas d'usage, il ne s'agit pas de créer un nouveau produit mais la valeur ajoutée apportée par le traitement du fichier principal réside dans la

création d'un jeu de données abrégé et spécifique. En plus, la collectivité territoriale est usager et médiateur de données publiques.

Un autre exemple illustre la pratique d'extraction. En effet, à partir des données libres de la base de données OSM, la société Ville à vélo a créé une carte interactive des aménagements cyclables du Grand Lyon²⁷⁵. La société a extrait les données en relation avec les vélos : les pistes cyclables, les voies de bus et de vélo, les parkings vélo, les stations de VéloV, les réparateurs de vélos pour les représenter sur une carte. Dans ce cas, les données extraites ont servi à créer une interface spécialisée sur les vélos sur une zone géographique précise. La manipulation des données par extraction présente une particularité par rapport aux productions graphique et écrite. Ces deux derniers transforment les jeux de données dans le fond et dans la forme. Les produits dérivés (textes et images) ne sont plus réutilisables tandis que les données extraites gardent les spécificités des données sources.

La pratique de combinaison des données est un autre modèle plus complexe de réutilisation. L'extraction et la combinaison peuvent paraître contradictoires. En effet, la combinaison se base sur la collecte de données de plusieurs sources pour les rassembler et les croiser entre eux. Les produits issus des Mashups des données est un exemple typique de cette pratique. Un mashup est une application Web qui utilise un contenu provenant de sources différentes pour créer un nouveau service affiché dans une seule interface. D'après Fichter (2009), le phénomène de mashup n'est pas nouveau. Il a explosé sur le web à partir de l'année 2005 pour deux raisons principales : la mise à disposition des données sur le « cloud » ; et la croissance rapide des outils de création des mashups. La composante fondamentale de l'écosystème de mashup est les fournisseurs de données « ouvertes » dans le « cloud » d'internet. De plus, il existe des données ouvertes, plus les utilisateurs ont de possibilités d'utiliser les mashups pour animer leurs sites web et fournir des services nouveaux et intéressants. La logique des mashups est le mélange et le mixage des données provenant de sources différentes. La valeur ajoutée est de créer un nouveau service d'enrichissement, de concentration de données liées, selon une nouvelle représentation et visualisation dans une seule interface. Les mashups sont un enrichissement, un mixage, une visualisation et une réutilisation des

²⁷⁵ Voir : “Carte Des Aménagements Cyclables Du Grand Lyon.” Accessed August 16, 2014. <http://carte.lavilleavelo.org/>.

données existantes. Pour présenter ce modèle, nous nous référons au Mashup « Les Musées en France »²⁷⁶. Réalisé par Julien Homo, ingénieur informatique et Poupeau, data architect chez Antidot, ce mashup est un exemple d'agrégation de données ouvertes afin de produire une application web. En effet, dans le cas des mashups, les utilisateurs ne sont pas en situation de création de nouvelles données, au contraire, ils utilisent des données existantes sur le web. La recherche des données sur le web ne se limite pas aux données publiques publiées par des institutions publiques mais elle prend en compte toutes les sources de données libres sur le web comme Wikipedia, DBpedia, Openstreetmap et les réseaux sociaux. Ainsi pour la construction du mashup, il s'agit de consolider des données éparses de plusieurs sources. Le mashup « les musées en France », est construit sur les données de bases du fichier Excel de la liste des musées de France disponible sur data.gouv.fr²⁷⁷. Ensuite les données sont enrichies et liées à des données complémentaires provenant de plusieurs sources : répertoire des musées français disponibles sur le site du ministère de la culture, Le fichier fréquentation 2006-2010 des musées de France mise à disposition sur le site data.gouv.fr par le Ministère de la Culture et de la Communication, DBpedia et wikidata avec des données propres à chaque musée, aux œuvres et aux artistes associés, wikimedia commons et le réseau social Instagram via son API pour ajouter des photos sur les musées, wikipedia pour inclure une description de chaque musée, la base de données Freebase²⁷⁸, le site data.bnf.fr de la Bibliothèque National de France (BnF) pour récupérer les ouvrages sur les musées ou dont le musée est l'auteur, la base de données OpenStreetMap pour afficher la géo-localisation du musée sur une carte, les moyens de transports les plus proches, les monuments historiques à proximités, etc. (voir la figure 20). Les bases de données libres comme les portails open data publics, DBpedia, Freebase, OSM sont un exemple des données structurées sur le web. Quand les données ne sont pas publiées dans des formats structurés comme le cas des pages web HTML du ministère de la culture, le développeur du mashup a utilisé la technique de scraping. La collecte de données de ces sources est automatique depuis les différents sites Web.

²⁷⁶ Voir : “Les Musées En France.” Accessed August 15, 2014. <http://labs.antidot.net/museesdefrance/#search>.

²⁷⁷ La liste contient 1224 lignes et inclut le nom du musée, l'adresse, les horaires d'ouverture, le code postale, le site web, la ville, la région, et si le musée est ouvert ou pas.

²⁷⁸ Freebase est une base de données collaborative sous la licence Creative Commons Attribution Licence



Musée des Beaux-Arts

Le musée des beaux-arts de Lyon est un musée municipal fondé en 1801 et situé sur la place des Terreaux dans une ancienne abbaye de bénédictines, l'abbaye des Dames de Saint-Pierre, édifice classé monument historique et plus connu sous le nom de palais Saint-Pierre. De 1988 à 1998 le musée a connu de grands travaux de rénovation qui ont permis une extension importante des surfaces d'expositions. Réparties sur soixante-dix salles, ses collections embrassent une période comprise entre l'Égypte antique et l'art moderne et la plupart des disciplines artistiques y sont représentées, faisant du musée des beaux-arts l'un des plus importants musées français et européens. Des expositions temporaires y sont proposées toute l'année.

[Moins d'informations](#)

Direction
Aucune information disponible

Informations pratiques

Horaires

Ouverture nocturne

Fermeture annuelle

Téléphone

Tarifs

Équipements

Services

Lien

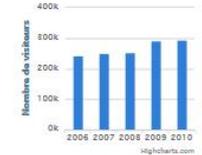


Réseaux sociaux
Aucun réseau social disponible

Localisation
20, Place des Terreaux
69001 Lyon
Rhône-Alpes
Rhône



Fréquentation



Année	Nombre de visiteurs
2006	~250k
2007	~280k
2008	~280k
2009	~300k
2010	~300k

Œuvres (22) Artistes (16) Bibliographie (50) Images de Wikimedia Commons (853) **Images des réseaux sociaux (35)** À proximité (47)

Les photos ci-dessous sont issues des réseaux sociaux. Seules celles en provenance d'Instagram sont disponibles pour le moment. Cliquez sur l'image pour en savoir plus.

Affiner votre recherche : [Par mots-clés](#) [Par filtres](#)

35 images correspondent à ce musée



Figure 20. Capture d'écran du mashup « les musées en France » prise le 15 août 2014 du site web labs.antidot.net.

Conclusion du chapitre 11

En présentant le profil et les structures de la communauté open data locale, les développeurs et les graphistes sont considérés comme deux acteurs mis en vedette. La communauté la plus visible et médiatisée est axée sur des jeunes ingénieurs en informatique, des développeurs et des graphistes. Pourtant, les journalistes gardent leurs places comme réutilisateurs classiques de la donnée publique malgré les défis techniques de manipulation des données. Les acteurs économiques se trouvent à la périphérie du réseau de l'open data et le modèle économique issu des données ouvertes reste encore flou. Les écosystèmes des réutilisateurs des données publiques varient entre offre gratuit ou payant ou les deux ensembles. Chaque acteur a créé son propre écosystème.

Enfin, nous avons présenté le processus de transformation de la donnée par chaque usage, le type de traitement qu'elle a subi et l'acteur intervenant. La réutilisation des données publiques est liée à un « processus de traitement » qui devrait apporter une « valeur ajoutée » aux données utilisées. Le premier modèle de réutilisation cherche à construire une connaissance. Un individu peut faire un usage précis en cherchant une donnée ou un jeu de données précis. La transformation des données en visualisations statiques ou dynamiques est une réutilisation courante et encouragée dans le monde de l'open data pour rendre les données plus compréhensibles au grand public. La production écrite est une interprétation des jeux de données qui les transforme en article ou rapport écrit qui combine souvent le texte et les visualisations pour une meilleure exploitation des données. Une réutilisation technique des données par les experts informatiques produit des interfaces de recherche et de présentation des données. Le dernier modèle de réutilisation repose sur la logique d'extraction ou de combinaison des données. L'extraction consiste à extraire des fragments de données présentées sous forme de jeux de données pour créer un nouveau jeu de données plus « spécifique », tandis que la combinaison est une pratique plus complexe qui se base sur la collecte de données de plusieurs sources pour les rassembler et les croiser entre eux.

CONCLUSION GENERALE ET PERSPECTIVES

Open data est un phénomène récent qui connaît une expansion rapide mais qui est encore en phase d'expérimentation et d'évaluation. C'est un objet complexe que les différents acteurs l'interprètent d'une approche nuancée et modulée. A défaut d'une définition universelle ou juridique reconnue de l'open data, les acteurs créent un « open data hybride » face aux contours encore flous de cet objet en développement. Dans cette conclusion générale, nous présentons nos principales réflexions, que nous articulons par le biais de nos hypothèses de recherche, ainsi que nos propres perspectives de recherche.

Synthèse de nos contributions

Si de nombreux questionnements ont été soulevés lors de ce travail de recherche, et si nous avons déjà exposé les réponses à nos principales problématiques dans la troisième partie, nous souhaitons ici résumer nos réflexions issues de nos chapitres en les articulant avec nos hypothèses.

La première hypothèse suppose que l'open data donne lieu à une nouvelle distribution des rôles des acteurs impliqués tout au long du processus d'ouverture. En étudiant dans la troisième partie les principales phases du processus d'ouverture, nous avons repéré une multitude d'acteurs en identifiant leur place et leur degré d'influence, directe ou indirecte, sur le projet open data de la collectivité territoriale. Dans le tableau 18 nous proposons de restituer les principaux acteurs en fonction de toutes les phases du processus open data ainsi que leurs apports que nous avons pu dégager. Nous avons distingué un positionnement dynamique des acteurs et conclu que le partage des rôles n'est pas clair. Les pratiques de chaque acteur ne se limitent pas à une étape du processus d'ouverture des données - que ce soit de réflexion, production, de diffusion, de promotion ou d'usage - ce qui crée un système mixte d'open data. Par son propre rôle de fédérateur, la collectivité rassemble des producteurs et des utilisateurs autour d'une plateforme de diffusion. C'est de cette manière qu'un processus d'ouverture dépasse en premier le stade de collecte interne et parvient à créer un écosystème d'acteurs autour du projet. Les quelques partenariats de production exposés mettent en exergue les différents producteurs

externes associées au projet dans une logique de « participation dirigée ou encadrée ». Toute production suppose de la part de la collectivité une stratégie d'ouverture mais aussi un certain contrôle. La collectivité se charge aussi de mobiliser la réutilisation de ses données en suivant deux logiques : la promotion et la négociation. Cette situation de départ est perçue comme satisfaisante afin de créer un lien direct entre la collectivité et les utilisateurs de données.

Le mouvement associatif doté d'une mission de promotion est un acteur concerné de façon remarquable dans l'open data. Dans la phase de réflexion, quelques collectifs ont réussi à développer une dynamique sociale et attiré l'attention des élus à l'open data. Le mouvement associatif a également accompagné le processus d'open data par la vulgarisation et la promotion du concept d'open Data en organisant plusieurs activités et événements mais rarement comme réutilisateurs de données publiques.

La deuxième hypothèse met en question le système d'offre des collectivités et de son adéquation avec les besoins des utilisateurs. En effet, la question de l'ouverture des données reste une initiative volontaire prise par les élus en l'absence d'une injonction juridique. Les collectivités sont conscientes des avantages politiques, sociaux et économiques, mais aussi des risques de l'open data. Nous avons montré que malgré les efforts de créer un système multi-acteur, la mise en œuvre de l'identification et de la collecte des données est éclectique et instable et ne construit pas un processus solide de l'Open data. Malgré l'urgence suggérée par le contexte politique et sociétal, l'ouverture et le choix des données sont généralement déterminés par des considérations pragmatiques qui imposent une ouverture lente mais progressive sans prétendre à l'exhaustivité. Si l'ouverture est une conquête progressive, il y a des « degrés d'ouverture » selon l'effort et les techniques disponibles. La collectivité qui lance un portail open data ne signifie pas automatiquement un accès illimité et inconditionnel à ses données. Dans notre étude, les collectivités adoptent la définition opérationnelle de l'open data qui laisse à chaque acteur de l'expliquer en considérant ses objectifs et ses compétences.

En étudiant l'offre des collectivités territoriales en matière de données ouvertes et la morphologie et la ligne éditoriale des plateformes, nous avons discerné une double approche expert/amateur au regard des utilisateurs ciblés. Loin d'être uniquement un outil et une source pour un public restreint

d'experts de la donnée, les plateformes open data s'adressent également aux amateurs de la donnée donc au grand public.

En s'interrogeant sur les besoins des utilisateurs, l'open data nous apparaît comme construit sur une logique d'offre que de demande pour deux raisons. Premièrement, l'open data est un mouvement accompagné voire stimulé. La stratégie des collectivités consiste, dans une étape de post-publication, à faire connaître l'open data tant auprès des acteurs publics que des citoyens. La diversité des données, à la fois aux niveaux des domaines et des propriétés techniques, ce qui élargit le public visé et oblige les collectivités à utiliser une variété de techniques de promotion pour toucher les experts et le grand public. Les orientations stratégiques de la promotion cible, en premier, « le projet » pour inciter les gens à aller sur le portail open data de la collectivité, ensuite « la culture des données » pour la sensibilisation du grand public à l'exploitation des données publiques, et enfin la « promotion de la réutilisation », pour mettre en action la réutilisation des données ouvertes par l'organisation des concours et hackathons. Deuxièmement, ce fait est dû à la difficulté de repérer les réutilisateurs et d'identifier la nature de leurs besoins. Les usages se manifestent difficilement et la « mesure de l'usage », en comptant le nombre de fois un fichier est téléchargé, semble le seul indicateur pertinent pour mesurer et évaluer l'usage. A partir des usages observés, nous avons pu distinguer quatre éléments fondamentaux à la réutilisation qui sont l'intérêt des utilisateurs, l'offre de données par les collectivités, le degré d'appropriation des données par les utilisateurs et enfin l'action où les modèles aboutis des réutilisations. Les usages observés se structurent essentiellement autour de la recherche d'une donnée précise, la production graphique et analytique et la production d'interface de recherche technique. L'écosystème de la réutilisation par les acteurs économiques reste encore floue. Pourtant nous avons proposé un modèle du système économique issue des exemples de réutilisation observées. C'est un modèle qui distingue entre les données « moyen » et les données « objets » par rapport aux entreprises et entre l'offre gratuit celui qui est payant.

La troisième hypothèse s'intéresse particulièrement aux plateformes open data considérées comme des « dispositifs socio-techniques » émergents du web social dans leurs aspects de médiations, partage, collaboration et de coproduction de données. Dans ce cadre, nos observations nous permettent de mettre en exergue, de questionner et d'analyser les plateformes open data dont leur architecture principale nous amènent à les qualifier

«dispositifs socio-techniques ». Ils sont des outils sociaux car ils donnent lieu à des formes d'action. Autour d'un catalogue de jeux de données, il subsiste des espaces virtuels, des blogs, des forums d'interactions entre les producteurs et les réutilisateurs des données ouvertes. La structure et le fonctionnement du dispositif disposent d'une panoplie d'indicateurs pour connaître voire gérer leurs utilisateurs : nombre de téléchargement, le nombre de jeux de données. Ces indicateurs de mesure, devenus prolifiques, ne sont pas purement techniques car ils sont des signes de distinction sociale hautement symbolique.

Les limites de la recherche

Le premier argument invoqué par les acteurs de l'open data pour expliquer leurs pratiques face au phénomène open data est que le modèle reste à trouver surtout que nous expérimentons le début de la phase de post-ouverture des données. Donc, il est encore relativement tôt pour en voir les résultats sociaux, politiques ou économiques de l'open data. Malgré notre tentative d'examiner les enjeux de l'implémentation de l'open data et les réactions des acteurs, il nous semble que dans quelques années les chercheurs auront plus de recul pour observer les différentes pratiques et écosystèmes liés à l'open data. La situation actuelle doit être considérée comme transitoire avant d'étudier de nouvelles conditions de diffusion et de réutilisations.

Les projets open data s'inscrivent alors dans des configurations instables et provisoires. Dans ce contexte, la dynamique de la législation française (la dernière loi apparue en décembre 2015 sur la gratuité des données publiques) et l'évolution des technologies et des contenus des plateformes open data en parallèle avec la rédaction de notre recherche nécessitaient de notre part un travail de veille systématique afin de garantir la fraîcheur et la mise à jour de nos analyses.

Prolongements possibles de cette recherche

Si nous adoptons dans cette recherche une approche globale du projet open data visant à associer les modes de production et de collecte de données, les méthodes de diffusion et les dispositifs utilisés, et enfin les modèles d'usage des données publiques numériques dans un processus d'ouverture, c'est que ce travail constitue la première étape pour des études plus approfondies sur chaque étape du processus. L'open data comme objet vaste et hybride à appréhender nous a amené dans un premier temps à essayer de l'embrasser dans toute sa complexité. Les prolongements possibles de cette recherche ouvrent de nouvelles pistes d'approfondissement à la fois de nature théorique et pratique.

Tout d'abord, le fait d'aborder la notion de données par rapport à des réflexions théoriques menées sur la notion de « document numérique » nous ouvre la voie à une réflexion critique sur les nouvelles formes d'un document sur le web. De nombreux travaux restent à mener dans la lignée des recherches menées par exemple par Salaün (2012).

Comme nous l'avons montré, la mise en œuvre de l'identification et de la collecte des données est éclectique et instable et ne construit pas un processus solide de l'Open data. Si la doctrine fondamentale de l'open data est la diffusion proactive (ou open data par défaut), il nous semble alors essentiel de se questionner plus avant sur un système de diffusion possibles et les méthodes d'application d'un tel système. La proposition d'un système modèle pourrait à notre avis aider les organisations vers un open data par défaut.

Enfin, l'ambition de notre approche sur les usages des données est de préciser l'observation sur une catégorie spécifique des usagers qui est les data journalistes. Il s'agit d'observer les pratiques d'usages des données publiques locales pour appréhender le lien entre l'open data et les usages effectifs de ces données afin d'exploiter en quoi, et comment, celles-ci orientent réellement les comportements des journalistes au plan local.

Tableau 18. L'apport des acteurs dans le processus open data

REFLEXION			
Acteurs internes	Acteurs d'autorité	Elus locaux	Prendre des décisions stratégiques. Par exemple adopter ou de rejeter un projet d'ouverture de données. Des décisions stratégiques relatives au fonctionnement d'ensemble du projet (la licence, la mutualisation, les partenariats) sont prises par le conseil général de la collectivité.
	Acteurs opérationnels	Directeurs des services	Montrer l'engagement de la collectivité dans la politique open data
		Services responsables du projet open data	Piloter et mettre en œuvre du projet
Acteurs externes	Mouvement associatif	Collectifs citoyens et activistes	Exercer une pression et développer une dynamique locale d'ouverture
IDENFIFICATION			
Acteurs internes	Acteurs opérationnels	Directeurs des services	Produire des données associées qu territoire et sélectionner des données à ouvrir
		Services responsables du projet open data	Identifier et collecter les données publiques produites par les services de la collectivité
Acteurs externes	Offices et instituts nationaux	Insee, IGN, ONEMA	Produire des données associées à un territoire défini. Des collectivités extraient des données de l'Insee pour les

			publier sur leurs portails open data
	Entreprises chargées d'un service public	Groupe Bolloré, société Keolis	Produire des données associées à un territoire défini. Par exemple, ces deux entreprises publient, en partenariat avec Grand Lyon, de données associées au territoire.
	Crowdsourcing	Communauté locale et contributeurs d'OSM	Produire des données associées à un territoire défini. Par exemple, des collectivités publient des données extraites de la base de données cartographique d'OSM et organise des cartoparties en partenariat avec la communauté locale
DISPOSITIFS			
Acteurs internes	Les acteurs opérationnels	Services responsables du projet open data	Administrer le contenu des dispositifs
Acteurs externes	Les collectivités locales	Collectivités du même département ou région	Créer des catalogues communs
	Utilisateurs	Individus	Evaluer, commenter et proposer des jeux de données.
PROMOTION			
Acteurs internes	Les acteurs opérationnels	Les services responsables du projet open data	Promouvoir le projet, la culture des données et la réutilisation
Acteurs externes	Mouvement	Fing, LiberTic	Organiser des événements

	associatif		
	Educateurs	Écoles	Promouvoir la culture des données
USAGE			
Acteurs externes	Utilisateurs	Individus	Rechercher une donnée précise
		Infographiste	Produire des graphiques
		Journaliste et consultants	Produire des analyses
		Ingénieurs informatiques	Produire des interfaces
		Ingénieurs informatiques	Combiner ou extraire les données

BIBLIOGRAPHIE

Textes de lois et rapports officiels

Communauté urbaine de Grand Toulouse. "Délibération N° DEL-11-337 : Ouverture Des Données Publiques de La Communauté Urbaine de Grand Toulouse," September 29, 2011. <http://data.toulouse-metropole.fr/documents/10157/18fb9822-c587-4f42-8f27-bb3b401a265d>.

États-Unis. Executive office of the President memorandum for the heads of executive departments and agencies. 2009.
http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf

États-Unis. *Open Government Act*, 2007.
<http://www.justice.gov/oip/amendment-s2488.pdf>.

États-Unis. *The Freedom of Information Act*, 1966.
<http://foia.state.gov/Learn/FOIA.aspx>.

France. Assemblée Nationale. Projet de loi sur la société de l'information. 2001. <http://www.assemblee-nationale.fr/11/projets/pl3143.asp>

France. *Loi N° 78-753 Du 17 Juillet 1978 Portant Diverses Mesures D'amélioration Des Relations Entre L'administration et Le Public et Diverses Dispositions D'ordre Administratif, Social et Fiscal / Modifié Par Ordonnance n°2005-650 Du 6 Juin 2005*, 2014.
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006068643&dateTexte=vig>.

France. *Loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public*.
<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000031701525&dateTexte=&categorieLien=id>

France. Commission générale de terminologie et de néologie. *Vocabulaire de L'informatique et Du Droit*, 2014.
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000028890784&dateTexte=&categorieLien=id>.

France. *Décret N° 2010-31 Du 11 Janvier 2010 Relatif À La Direction de L'information Légale et Administrative. 2010-31*, 2010.
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000021658499categorieLien=id>.

France. Délégation interministérielle à la réforme de l'Etat. *Le Développement des sites internet des services de l'Etat : évaluation 2001*. France : La documentation française, 2001.
<http://www.ladocumentationfrancaise.fr/rapports-publics/014000796/index.shtml>

France. Premier ministre, *Circulaire Du 14 Février 1994 Relative À La Diffusion Des Données Publiques*, 1994.
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000181007>.

France. Premier ministre. *Circulaire Du 14 Février 1994 Relative À La Diffusion Des Données Publiques*, 1994.
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000181007>.

France. Premier ministre. *Circulaire Du 20 Mars 1998 Relative À L'activité Éditoriale Des Administrations et Des Établissements Publics de l'Etat*.
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000204563>.

France. Premier Ministre. Secrétariat général pour la modernisation de l'action publique. "Remise Du Rapport Trojette Sur L'ouverture Des Données Publiques" *Le Portail de La Modernisation de L'action Publique*, May 11, 2013. <http://www.modernisation.gouv.fr/laction-publique-se-transforme/en-ouvrant-les-donnees-publiques/remise-du-rapport-trojette-sur-louverture-des-donnees-publiques>.

Lacombe, Romain, Francois Vauglin, Alice Vieillefosse, and France. Délégation aux usages de l'Internet. *Les Données Publiques Au Service de L'innovation : Pour Une Politique Ambitieuse de Réutilisation Des Données Publiques : Rapport À La Délégation Aux Usages de l'Internet*. Paris: Ecole des Ponts ParisTech, January 7, 2011.
<http://www.proximamobile.fr/sites/default/files/RapportDonneesPubliques2011.pdf>.

Levy-Rosenwald, Marianne. *Rapport du Médiateur de l'édition publique pour l'année 2011*. Rapport public. France: Ministère de la culture et de la communication, 2012. <http://www.ladocumentationfrancaise.fr/rapports-publics/134000139/index.shtml>.

Mandelkern, Dieudonné, Bertrand Du Marais, and France. Commissariat général du plan. *Diffusion des données publiques et révolution numérique: rapport de l'atelier "Des moyens nouveaux au service de la diffusion des données publiques."* Paris, France: la Documentation française, 1999. <http://www.ladocumentationfrancaise.fr/rapports-publics/994001620/>.

Trojette, Mohammed Adnène. *Rapport Au Premier Ministre : Ouverture Des Données Publiques : Les Exceptions Au Principe de Gratuité Sont Elles Toutes Légitimes?*, 2013. <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/134000739/0000.pdf>.

Ouvrages

Balmisse, Gilles. *Gestion des connaissances : outils et applications du knowledge management*. Paris: Vuibert, 2005.

Bluenove. *Open Data : Quels Enjeux et Opportunités Pour L'entreprise : Une Initiative de Bluenove, En Partenariat Avec SNCF, Le Groupe La Poste, SUEZ Environnement, Groupe Poult Sur La Base de L'étude Bluenove - BVA*, 2011. http://www.bva.fr/data/actualite/actualite_fiche/329/fichier_download22386.pdf.

Bollier, David. *The Promise and Peril of Big Data*. Washington: The Aspen Institute, 2010. http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf.

Breton, Philippe. *L'utopie de La Communication*. Paris: la Découverte, 2004.

Chesbrough, Henry. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston: Harvard Business Press, 2003.

Chesbrough, Henry. *Open Business Model : How to Thrive in the New Innovation Landscape*. Cambridge: Harvard Business School Press, 2006.

Chesbrough, Henry, Wim Vanhaverbeke, and Joel West, eds. *Open Innovation: Researching a New Paradigm*. Oxford: Oxford University Press, 2006.

Chignard, Simon. *Open data : comprendre l'ouverture des données publiques*. Limoges: EdFyp, 2012.

Commission Européenne. Direction générale de la société de l'information. "eEurope 2002 : Créer Un Cadre Communautaire Pour L'exploitation de L'information Émanant Du Secteur Public" Commission Européenne, 2002.

Commission européenne. *L'information Emanant Du Secteur Public : Une Ressource Clef Pour l'Europe : Livre Vert Sur L'information Emanant Du Secteur Public Dans La Société de L'information*, 1998.
ftp://ftp.cordis.europa.eu/pub/econtent/docs/gp_fr.pdf.

Cross, Harold L. *The People's Right to Know: Legal Access to Public Records and Proceedings*. Columbia University Press, 1953.

Crozier, M. et Friedberg, E. *L'acteur et le système. Les contraintes de l'action collective*. Paris : Éditions du Seuil, 1992.

Décider ensemble. *Ouverture Des Données Publiques et Participation : Enjeux Démocratiques*. Paris, France: Décider ensemble, 2012.
<http://www.deciderensemble.com/nos-etudes/ouverture-des-donnees-publiques-et-participation-quels-enjeux-democratiques>.

Delort, Pierre. *Le Big Data*. Presses Universitaires de France, 2015.

Even, B. "La notion de document administratif ". *AJDA*, 1985. p. 529.

Favre, Cédric, and Yann Mareschal. *Données Publiques : Guide Pratique et Juridique*. Bordeaux: Aquitaine Europe Communication, 2010.
<http://www.aecom.org/Vous-informer/Juridique-TIC/Guides-juridiques/Les-donnees-publiques-decembre-2010>.

Flichy Patrice. *L'innovation technique*. Paris : La découverte, 1995.

Fondation Internet Nouvelle Génération. "Guide Pratique de L'ouverture Des Données Publiques Territoriales." *Fing*, 2011. <http://fing.org/?Guide-pratique-de-l-ouverture-des&lang=fr>.

Gaudrat, Philippe. *Commercialisation des données publiques*. La Documentation française, 1992.

Gitelman, Lisa. *"Raw Data" Is an Oxymoron*. London: MIT Press, 2013.

Glaser B.G. and Strauss, A.L. *The discovery of Grounded Theory : Strategies for qualitative research*. Chicago : Aldine Publishing Company, 1967.

Glott, Ruediger, Philipp Schmidt, and Rishab Ghosh. *Wikipedia Survey – Overview of Results*. Netherlands: United Nations University-Merit, 2010.
http://www.ris.org/uploadi/editor/1305050082Wikipedia_Overview_15March_2010-FINAL.pdf.

Hogge, Becky, and Open Society Foundations. *Open Data Study*. London, UK: Open Society Foundation, 2010.
<http://www.opensocietyfoundations.org/reports/open-data-study>.

Institut National de recherche en informatique et en automatique. *Pérenniser Le Document Numérique*. Paris: ADBS, 2006.

Kaplan, Daniel, and Thierry Marcou. *La ville 2.0, plateforme d'innovation ouverte*. Limoges, France: Fyp éd., 2009.
http://doc.openfing.org/FING/LAFING/PUBLICATIONS/La%20ville_2.0_plateforme.pdf.

Millerand, Florence. *Usages des NTIC : les approches de la diffusion, de l'innovation et de l'appropriation (2ème partie)*. Vol. 2, 1999.

Sassen, Saskia, and Robert Latham. *Digital Formations: IT and New Architectures in the Global Realm*. Princeton, N.J.: Princeton university press, 2005.

Le Coadic, Yves-François. *La science de l'information*. Paris : Presse Universitaire de France, 1994.

Le Coadic, Yves-François. *Usages et usagers de l'information*. Paris, France: ADBS : Nathan, 1997.

Le Marec, Joëlle. *Publics et musées: la confiance éprouvée*. Paris, France: L'Harmattan, DL 2008.

Levy, P. *L'intelligence collective. Pour une anthropologie du cyberspace*. Paris : Editions La Découverte, 1994.

Libaert, Thierry. *La Transparence en trompe-l'oeil*. Paris: Descartes & Cie, 2003.

Marcon, C., and N. Moinet. *La Stratégie-Réseau*. Paris: Edition ZéroHeure, 2000.

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big data: a revolution that will transform how we live, work and think*. New-York, Etats-Unis: Eamon Dolan books-Houghton Mifflin harcourt, 2013.

McCarthy, E. Jerome. *Basic marketing: a global-managerial approach*. Boston, Etats-Unis: McGraw-Hill : Irwin, 2005.

McKinsey Global Institute. "Big Data: The next Frontier for Innovation, Competition, and Productivity," 2011.
http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

McKinsey Global Institute. "Game Changers : Five Opportunities for US Growth and Renewal," 2013.
http://www.mckinsey.com/insights/americas/us_game_changers.

Millerand, F, Proulx, S. & Rueff, J. *Web social – Mutation de la communication*. Québec : Presses de l'université du Québec, 2010.

Musso, P.. *Télécommunications et Philosophie des Réseaux. La postérité paradoxale de Saint-Simon*. Paris : PUF., 1997.

Olivesi, Stéphane. *Introduction à la recherche en SIC*. Grenoble : Presses universitaires de Grenoble, 2007.

Open Knowledge Foundation, *Le Manuel de L'opendata*. Open Knowledge Foundation, 2012. <http://opendatahandbook.org/fr/>.

Pollitt, Christopher, and Geert Bouckaert. *Public Management Reform: A Comparative Analysis - New Public Management, Governance, and the Neo-Weberian State*. Oxford: Oxford University Press, 2011.

Rebillard, Franck. *Le web 2.0 en perspective : Une analyse socio-économique de l'internet*. Editions L'Harmattan, 2007.

Roberts, Alasdair. *Blacked Out: Government Secrecy in the Information Age*. Cambridge ; New York: Cambridge University Press, 2006.

Rowley, Jennifer, and Richard Hartley. *Organizing Knowledge*. 4 edition. Aldershot, England ; Burlington, VT: Ashgate, 2008.

Shuen, Amy. *Web 2.0: A Strategy Guide*. Sebastopol, CA: O'Reilly Media, 2008.

Tauberer, Joshua. *Open Government Data*, 2012. <http://opengovdata.io>.

Vidal, Geneviève, ed. *La sociologie des usages, continuités et transformations*. Paris: Lavoisier, 2012.

Vickery, Graham. *Review of Recent Studies on PSI Re-Use and Related Market Developments*. European Commission, 2011.

Wallace, Danny P. *Knowledge Management: Historical and Cross-Disciplinary Themes*. USA: Libraries Unlimited, 2007. p. 1-14.

Chapitres d'ouvrages

Alberede, Marine, and Fabien Eychenne. "Ville 2.0 : favoriser l'accès à la ville et faire du numérique un facteur de lien social." In *Bibliothèques 2.0 à l'heure des médias sociaux*, edited by Muriel Amar and Véronique Mesguich, Nouvelle édition. Bibliothèques. Paris: Éd. du Cercle de la librairie, 2012.

Bullich, V et Guignard, T. « Les plates-formes d'accès aux contenus : des dispositifs au cœur de la reconfiguration des filières communicationnelles » In *La richesse des réseaux numériques*, edited by S. Agostinelli, D. Augey et F. Laurie . Aix en Provence: Presses Universitaires d'Aix-Marseille, 2012. pp. 145-160.

Fichter, Darlene. "What Is a Mashup?" In *Library Mashups: Exploring New Ways to Deliver Library Data*, edited by Nicole C Engard. Medford, N.J.: Information Today, Inc., 2009.
<http://books.infotoday.com/books/Engard/Engard-Sample-Chapter.pdf>.

Jouët, Josiane, and Coralie Le Caroff. "L'observation ethnographique en ligne." In *Manuel d'analyse du web en sciences humaines et sociales*, edited by Christine Barats. Paris: A.Colin, 2013.

O'Reilly, Tim. "Government as a Platform." In *Open Government*, edited by Daniel Lathrop and Laurel Ruma. Farnham: O'Reilly, 2010.
<http://chimera.labs.oreilly.com/books/1234000000774/ch02.html>.

Merzeau, Louise. « L'intelligence de l'utilisateur ». In *L'utilisateur numérique*. Paris : INRIA, 2010. <https://halshs.archives-ouvertes.fr/halshs-00526527/document>

Proulx, S. « Communautés virtuelles : ce qui fait lien », in Proulx, S., Poissant, L., Senecal, M. (Dir.). *Communautés virtuelles : penser et agir en réseau*. Québec : Presses de l'Université Laval, 2006. pp. 13-26.
<http://sergePROULX.uqam.ca/wp-content/uploads/2010/12/2006-PROULX-communautes-vi-36.pdf>.

Roberts, Alasdair. "Dashed Expectations: Governmental Adaptation to Transparency Rules." In *Transparency: The Key to Better Governance?*, edited by Christopher Hood and David Heald. Oxford ; New York: Oxford University Press, 2006.

Rosenberg, Daniel. "Data before the Fact." In *Raw Data Is an Oxymoron*, edited by Lisa Gitelman. London: MIT Press, 2013.

Van de Walle, Steven, and Alasdair S. Roberts. "Publishing Performance Information : An Illusion of Control?" In *Performance Information in the Public Sector*, edited by Wouter Van Dooren and Steven Van de Walle, 211–26. Houndmills: Palgrave Macmillan, 2008.

Articles de revues

Anadon, M. et Guillemette, F. « La recherche qualitative est-elle nécessairement inductive ? », *Recherches qualitatives*, Hors-série, n° 5, 2007.

Anthes, Gary. "Invasion of the Mobile Apps." *Communications of the ACM* 54, no. 9 (September 2011): 16–18. doi:10.1145/1995376.1995383.

Apter, David. "Une approche interdisciplinaire de l'interdisciplinarité." *Revue internationale des sciences sociales*, no. 196 (October 1, 2013): 7–19.

Association des professionnels de l'information et de la documentation. « Le programme gouvernemental pour une société de l'information ». *Documentaliste* 35, no. 1 (January 1, 1998): 27–27. <http://www.adbs.fr/le-programme-gouvernemental-pour-une-societe-de-l-information-13481.htm?RH=REVUE>.

Bachimont, Bruno, et al. "Enjeux et Technologies : Des Données Au Sens." *Documentaliste-Sciences de l'Information* 48, no. 4 (2011): 24–41.

Badouard, Romain. "La mise en technologie des projets politiques. Une approche « orientée design » de la participation en ligne." *Participations* 8, no. 1 (May 22, 2014): 31–54.

Balcoü, Jean-Pierre, Stéphane Cottin, and Olivier Iteanu. "Droit de l'information". *Documentaliste-Sciences de l'Information* Vol. 48, no. 4 (December 1, 2011): 16–19.

Baskarada, Sasa, and Andy Koronios. "Data, Information, Knowledge, Wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of the Hierarchy and Its Quality Dimension." *Australasian Journal of Information Systems* 18, no. 1 (2013): 5–24.

Battisti, Michèle. "La Réutilisation Des Données Publiques : Un Enjeu Majeur Pour La Société Européenne de L'information." *Documentaliste-Sciences de l'Information* 41, no. 6 (2004): 349–55.

Berthault, Denis, and Bernard Marx. "Tarification de la réutilisation des informations publiques en France." *Les Cahiers du numérique* Vol. 9, no. 1 (April 1, 2013): 39–53.
http://www.cairn.info/article.php?ID_ARTICLE=LCN_091_0039.

Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked Data - The Story So Far." *International Journal on Semantic Web and Information Systems* 5, no. 3 (2009): 1–22. <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>.

Blumberg, Robert, and Shaku Atre. "The Problem with Unstructured Data." *DM Review*, no. February (2003): 42–46.

http://soquelgroup.com/Articles/dmreview_0203_problem.pdf

Boisot, M., and Canals A., « Data, information and knowledge : have we got it right ? ». *Journal of Evolutionary Economics* 14 (2004): 43-67

Boisnier-Bardou, Sylvie, and Isabelle Pailliar. "Information publique : stratégies de production, dispositifs de diffusion et usages sociaux." *Les Enjeux de l'information et de la communication* Volume 2012, no. 1 (December 1, 2012): 149–57.

Boudreau, Christian. "Qualité, efficience et efficacité de l'administration numérique à l'ère des réseaux : l'exemple québécois." *Revue française d'administration publique*, no. 131 (January 6, 2010): 527–39.

Boustani, Joumana. "Accès et Réutilisation Des Données Publiques. Etat Des Lieux En France." *Les Cahiers Du Numérique* 9, no. 1 (2013): 21–38.

Bowker, Geoffry C. "Biodiversity Datadiversity." *Social Studies of Science* 30, no. 5 (2000): 643–83.

Bowker, Geoffry C. "The Theory/Data Thing Commentary." *International Journal of Communication* 8, no. 2043 (2014): 1795–99.

Boyd, Danah, and Kate Crawford. "Critical Questions for Big Data : Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15, no. 5 (2012).

<http://www.tandfonline.com/doi/pdf/10.1080/1369118X.2012.678878>.

Burger-Helmchen, Thierry, and Julien Penin. "Crowdsourcing : Définition, Enjeux, Typologie." *Management & Avenir* 1, no. 41 (2011): 254–69.

Calzada Prado, Javier, and Miguel Ángel Marzal. "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents". *Libri: International Journal of Libraries Information Services* 63, no. 2 (June 2013): 123–34. doi:10.1515/libri-2013-0010.

Chanal, Valérie, and Marie-Laurence Caron-Fasan. "The Difficulties Involved in Developing Business Models Open to Innovation Communities: The Case of a Crowdsourcing Platform." *M@n@gement* 13, no. 4 (2010): 318–41.

Chaudiron, S. et Ihadjadene, M. « De la recherche de l'information aux pratiques informationnelles ». *Études de communication* 35 (2010) : 13-30.
<http://edc.revues.org/index2257.html>.

Chesbrough, H.W., and M.M. Appleyard, "Open innovation and strategy". *California Management Review* 50 (2007) : 57-76.

Chikh, Azzeddine. "Une approche méthodologique de réutilisation en ingénierie de document." *Document numérique* Vol. 7, no. 1 (March 1, 2003): 59-88.

Chun, S. A., R. Sandoval, S. Shulman, and E. Hovy. "Government 2.0: Making Connections between Citizens, Data and Government." *Information Polity* 15, no. 1-2 (2010): 1-9.

Cook, Fay Lomax, Lawrence R. Jacobs, and Dukhong Kim. "Trusting What You Know : Information, Knowledge, and Confidence in Social Security." *The Journal of Politics* 72, no. 02 (2010): 397-412.

Cukier, Kenneth, and Viktor Mayer-Schoenberger. "The Rise of Big Data." *Foreign Affairs* 92, no. 3 (June 5, 2013): 27-40.

Dumbill, E. "Making Sense of Big Data". *Big Data* 2013.
<http://online.liebertpub.com/doi/pdf/10.1089/big.2012.1503>

Flichy, Patrice, and Éric Dagiral. "L'administration électronique : une difficile mise en cohérence des acteurs." *Revue française d'administration publique* no110, no. 2 (June 1, 2004): 245-55.

Florescu, D., A. Levy and A. Mendelzon, « Database techniques for the World-Wide Web : a survey », *SIGMOD Record*, vol. 27, n° 3, p. 59-74, 1998.

Gardiès, Cécile, Isabelle Fabre, and Viviane Couzinet. "Re-Questionner Les Pratiques Informationnelles." *Études de Communication* 35 (2010): 121-32.
<https://edc.revues.org/2241>.

Gensollen, Michel. "Économie non rivale et communautés d'information." *Réseaux* no 124, no. 2 (April 1, 2004): 141-206.
http://www.cairn.info/resume.php?ID_ARTICLE=RES_124_0141.

Grimmelikhuijsen, Stephan. "Relier transparence, connaissances et confiance des citoyens dans l'État : expérience." *Revue Internationale des Sciences Administratives* 78, no. 1 (March 1, 2012): 55–78.

Grimmelikhuijsen, Stephan G., and Eric W. Welch. "Developing and Testing a Theoretical Framework for Computer-Mediated Transparency of Local Governments." *Public Administration Review* 72, no. 4 (July 1, 2012): 562–71.

Henry, Nicholas L. « Knowledge Management: A New Concern for Public Administration ». *Public Administration Review* 34, no. 3 (May-June 1974): 189-196.

Huyghe, François-Bernard. "« Think Tanks », Penser Pour Peser." *Médium* 1, no. 14 (2008): 27–41.

Jeanneret, Yves. « Usages de l'usage, figures de la médiatisation ». *Communication et langages* n°151 (2007): 3-19.

Labelle, Sarah, and Jean-Baptiste Le Corf. "Modalités de diffusion et processus documentaires, conditions du « détachement » des informations publiques. Analyse des discours législatifs et des portails open data territoriaux." *Les Enjeux de l'information et de la communication* Volume 2012, no. 1 (December 1, 2012): 209–22.

Le Marec, Joëlle. « L'usage et ses modèles : quelques réflexions méthodologiques ». *Spirale* no. 281 (2001): 105-122.
http://archivesic.ccsd.cnrs.fr/documents/archives_0/00/00/01/32

Lemos, A. « Les trois lois de la cyberculture. Libération de l'émission, connexion au réseau et reconfiguration culturelle », *Sociétés*, 91 (2006), 37-48.

Lemos, André, Rezende Pedro A. D., « Le Brésil dans la société de l'information : gouvernement Lula, *copyleft* et logiciels libres. », *Hermès, La Revue* 2/2005 (n° 42) , p. 198-205.

Lev-On, Azi, and Manin Bernard, « Internet : la main invisible de la délibération. », *Esprit* 5/2006 (Mai): 195-212.

Miège, Bernard. "Les Apports À La Recherche Des Sciences de L'information et de La Communication." *Réseaux* 18, no. 100 (2000): 547–68.

Moon, M. Jae". "The Evolution of E-government Among Municipalities: Rhetoric or Reality". *Public Administration Review* 62, no.4 (2002): 424–435.

Oliveri, Nicolas. « Logiciel libre et open source : une culture du don technologique », *Quaderni* 76 (2011): 111-119.

Pailliart, Isabelle. "La société de l'information : une société de contradictions?" *Revue européenne des sciences sociales. European Journal of Social Sciences*, no. XL- 123 (June 1, 2002): 55–63.

Pédauque, R. T. *Document : forme, signe et médium, les re-formulations du numérique*. Hal archives ouvertes, 2003.

http://archivesic.ccsd.cnrs.fr/sic_00000511

Peraya, Daniel et Bonfils, Philippe. « Nouveaux dispositifs médiatiques : comportements et usages émergents ». *Distances et médiations des savoirs* (2012). <http://dms.revues.org/126>

Phillips, Jeffrey. "Open Innovation Typology." *International Journal of Innovation Science* 2, no. 4 (2010).

Plantin, Jean-Christophe, and Jérémie Valentin. "Données ouvertes et cartographie libre." *Les Cahiers du numérique* Vol. 9, no. 1 (April 1, 2013): 85–107. http://www.cairn.info/article.php?ID_ARTICLE=LCN_091_0085.

Proulx, S., Latzko-Toth, G. « la virtualité comme catégorie pour penser le social », *Sociologie et sociétés*, Vol. 32, n° 2, pp. 99-122, 2000. p.105 <http://sergeproulx.uqam.ca/wp-content/uploads/2011/02/2000-proulx-la-virtualite-77.pdf>

Rauzier, Jean-Michel et al. "Métiers et Compétences." *Documentaliste-Sciences de l'Information* 48, no. 4 (2011): 4–9.

Renault, Sophie. "Crowdsourcing : La Nébuleuse Des Frontières de L'organisation et Du Travail." *RIMHE : Revue Interdisciplinaire Management, Homme(s) & Entreprise* 2, no. 11 (2014): 23–40.

Rieder, Bernhard. « De la communauté à l'écume : quels concepts de sociabilité pour le « web social » ? », *Tic&société*, Vol. 4, n° 1, 2010.
<http://ticetsociete.revues.org/822>

Rouvroy, Antoinette, and Thomas Berns. "Gouvernementalité Algorithmique et Perspectives D'émancipation." *Réseaux*, no. 177 (2013): 163–96.

Rowley, Jennifer. "The Wisdom Hierarchy: Representations of the DIKW Hierarchy." *Journal of Information Science* 33, no. 2 (April 1, 2007): 163–80. doi:10.1177/0165551506070706

Sakkab, N. "Connect and Develop Complement Research and Develop at P&G." *ResearchTechnology Management*, no. March-April (2002): 38–46.

Salaün, Jean-Michel. "Chronique Inachevée D'une Réflexion Collective Sur Le Document." *Communication & Langages*, 2004.
http://archivesic.ccsd.cnrs.fr/sic_00001028/document.

Shannon, C. E. "A Mathematical Theory of Communications." *The Bell System Technical Journal* 27 (1948).

Souchier, Emmanuël, Yves Jeanneret, and Joëlle Le Marec. *Lire, écrire, récrire : Objets, signes et pratiques des médias informatisés*. Paris: Bibliothèque Publique d'Information, 2003.

Soufron, Jean-Baptiste. « Standards ouverts, open source, logiciels et contenus libres : l'émergence du modèle du libre », *Esprit* 2009/3 (Mars/avril), p. 128-136.

Stiller, Henri. « Le portail, outil fédérateur d'information et de connaissances ». *Documentaliste-Sciences de l'Information*, no. 1 Vol. 38 (2001): 39-42

Thompson, Kristi, and Daniel M. Edelstein. "A Reference Model for Providing Statistical Consulting Services in an Academic Library Setting". *IASSIST Quarterly*, Summer/Fall (2004).
http://www.iassistdata.org/downloads/iqv0282_3thompson.pdf.

West, Joel, and Scott Gallagher. "Challenges of Open Innovation : The Paradox of Firm Investment in Open Source Software." *R&D Management* 36, no. 3 (2006).

Yu, Harlan, and David Robinson. "The New Ambiguity of 'Open Government.'" *59 UCLA Law Review Discourse*, no. 178 (2012). <http://papers.ssrn.com/abstract=2012489>.

Zins, Chaim. "Conceptual Approaches for Defining Data, Information, and Knowledge." *Journal of the American Society for Information Science and Technology* 58, no. 4 (2007): 479–93.

Articles de magazine, blogs et sites web

"2014/City Events", *Wiki International Open Data Hackathon 2014*. Accessed March 25, 2014. http://wiki.opendataday.org/2014/City_Events.

Anderson, Chris. "The End of Theory : The Data Deluge Makes the Scientific Method Obsolete." *Wired Magazine*, 2008. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory.

"Appel À Projet - Données Publiques Ouvertes - Montpellier Territoire Numérique." *Montpellier Territoire Numérique - Données Publiques Ouvertes*. Accessed July 19, 2014. <http://opendata.montpelliernumerique.fr/Appel-a-projet>.

Bancilhon, Francois. La startup du jour : Data Publica utilise le Big Data pour faire de la vente prédictive, February 3, 2014. <http://frenchweb.fr/la-startup-du-jour-data-publica-utilise-le-big-data-pour-faire-de-la-ventepredictive/140349>.

Berdot, Vincent. "Pourquoi Les Collectivités Rechignent À Mutualiser Leurs Plates-Formes Open Data". *01Business*, April 6, 2012. <http://pro.01net.com/editorial/567381/pourquoi-les-collectivites-rechignent-a-mutualiser-leurs-plates-formes-open-data/>.

Berners-Lee, Tim, James Hendler et Ora Lassila, « The Semantic Web », *Scientific American*, mai 2001.

Berners-Lee, Tim. "Sir Tim Berners-Lee: Raw Data, Now!," In *Wired UK*, September 11, 2012. <http://www.wired.co.uk/news/archive/2012-11/09/raw-data>.

Blanc, Sabine. "L'État Plateforme, Vraie Source de Services Publics Innovants Ou Cache-Misère?" *Lagazette.fr*, 2015.

<http://www.lagazettedescommunes.com/323547/letat-plate-forme-vraie-source-de-services-publics-innovants-ou-cache-misere/>.

Blanc, Sabine. "Le Fouillis Des Licences Open Data S'éclaircit [Fiche Pratique]." *Lagazette.fr*, November 25, 2013.
<http://www.lagazettedescommunes.com/208893/le-fouilli-des-licences-open-data-seclaircit-fiche-pratique/>.

Blanc, Sabine. "L'ouverture Des Données Publiques En France Reste Perfectible." *Lagazette.fr*, October 30, 2013.
<http://www.lagazettedescommunes.com/204891/louverture-des-donnees-publiques-en-france-reste-perfectible/>.

"Bordeaux, Cité Digitale : Votre Ville Résolument 'Open Data.'" *Open Data Ville de Bordeaux*. Accessed July 10, 2014.
<http://opendata.bordeaux.fr/demarche>.

"Browser statistics and trends". W3schools. Accessed May 20, 2015.
http://www.w3schools.com/browsers/browsers_stats.asp.

Bruguière, Jean-Michel. "Droit Des Données Publiques". *Éduscol Portail National Des Professionnels de L'éducation*, 2004.
<http://eduscol.education.fr/chrgt/droitddp.pdf>.

Cardon, Dominique. "Zoomer Ou Dézoomer? Les Enjeux Politiques Des Données Ouvertes." *OWNI*, février 2011. <http://owni.fr/2011/02/21/zoomer-ou-dezoomer-les-enjeux-politiques-des-donnees-ouvertes/>.

Casilli, Antonio. « Small data vs. Big Data ». Séminaire EHESS Étudier les cultures du numérique : approches théoriques et empiriques. 15 février 2012
<http://www.bodyspacesociety.eu/2012/02/10/small-data-vs-big-data-seminaire-antonio-casilli-ehess-15-fevr-2012-17h/>

Chantry, Marion. "L'OpenDataLab : Un Lieu de Réflexion et de Tests En Mode Collaboratif." *Aquitaine Europe Communication. Agence Des Initiatives Numériques*, 2012. <http://www.aecom.org/Auberge-Numerique-incubateur-AEC/L-OpenDataLab>.

"Les Chefs d'Etat du G8 signent une Charte pour l'Ouverture des Données Publiques." *etalab*. Accessed June 23, 2014.
<http://www.etalab.gouv.fr/article-les-chefs-d-etat-reunis-a-loughe-erne>

[signent-une-charte-du-g8-pour-l-ouverture-des-donnees-publique-118576420.html](#).

Chemouil, David. "Opendata Toulouse". Accessed July 4, 2014. <http://opendata-toulouse.numahel.net/doku.php/>.

Civic Impulse, 'GovTrack.us - About', *GovTrack.us*. Accessed June 11, 2014. <https://www.govtrack.us/about>.

CNIL - Commission nationale de l'informatique et des libertés. "Open Data, Quels Enjeux Pour La Protection Des Données Personnelles?" In *CNIL*, 2013. <http://www.cnil.fr/les-themes/collectivites-locales/article/article/open-dataet-donnees-personnelles-lancement-dune-consultation-des-acteurs/>.

CNIL - Commission nationale de l'informatique et des libertés. "Open Data, Quels Enjeux Pour La Protection Des Données Personnelles?" In *CNIL*, 2013. <http://www.cnil.fr/les-themes/collectivites-locales/article/article/open-dataet-donnees-personnelles-lancement-dune-consultation-des-acteurs/>.

Collectif Open Data Rennes. "Nouveau Départ Pour l'Open Data À Rennes Métropole?" *Collectif Open Data Rennes*, April 13, 2014. <http://blog.cod-rennes.fr/2014/04/13/nouveau-depart-pour-lopen-data-a-rennes-metropole/>.

Commission d'accès aux documents administratifs, "Caisse Nationale D'assurance Maladie Des Travailleurs Salariés (CNAMTS de Paris) : Avis 20140831 Séance Du 27/03/2014," *CADA*, March 27, 2014, <http://www.cada.fr/avis-20140831,20140831.html>.

Commission d'accès aux documents administratifs, "Directeur Général de L'agence Francaise de Sécurité Sanitaire Des Aliments (AFSSA) : Conseil 20020741 Séance Du 14/03/2002," *CADA*, March 14, 2002, <http://www.cada.fr/conseil-20020741,20020741.html>.

Commission d'accès aux documents administratifs, "La Notion de Document Administratif," *Cada*. Accessed May 24, 2014. <http://www.cada.fr/la-notion-de-document-administratif,56.html>

Commission d'accès aux documents administratifs. "L'existence Du Document." *Commission D'accès Aux Documents Administratifs*. Accessed March 24, 2014. <http://www.cada.fr/l-existence-du-document,6082.html>.

Commission Européenne. Direction générale de la société de l'information. "Stratégie Numérique: Valoriser Les Données Des Administrations Publiques," December 12, 2011. http://europa.eu/rapid/press-release_IP-11-1524_fr.htm?locale=en.

Commission nationale de l'informatique et des libertés. "Missions". *CNIL*. Accessed July 7, 2014. <http://www.cnil.fr/linstitution/>.

"Concours Open Data Défi Numérique." *Data.toulouse-Metropole.fr*. Accessed July 19, 2014. <https://data.toulouse-metropole.fr/web/guest/concours-opendata>.

Conseil National du Numérique. "L'ouverture Des Données Publiques : Un Enjeu Majeur Qui Doit Etre Protégé Par Une Volonté Politique Déterminée," 2012. http://www.cnnumerique.fr/wp-content/uploads/2012/06/12-06-05_CP_OpenData.pdf.

Conway, Drew. "The Data Science Venn Diagram." *Drew Conway*, September 30, 2010. <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>.

Court, David. "Getting Big Impact from Big Data." *McKinsey Quarterly*, 2015. http://www.mckinsey.com/insights/business_technology/getting_big_impact_from_big_data.

Data Publica. "« L'open Data Est Très Mal Estimé » Interview de Claire Gallon de L'association Libertic". *Data Publica*, January 2, 2014. <http://www.data-publica.com/content/2012/11/lopen-data-est-tres-mal-estime-interview-de-claire-gallon-de-lassociation-libertic/>.

Data Publica. "Les Tableaux de Bord Des Territoires." Accessed July 20, 2014. <http://www.data-publica.com/content/les-tableaux-de-bord-des-territoires/>.

Décider ensemble. "Ouverture Des Données Publiques et Participation". Accessed July 4, 2014. <http://www.deciderensemble.com/nos-etudes/ouverture-des-donnees-publiques-et-participation-quels-enjeux-democratiques>.

Deldicque, Benoit. "Concours Opendata : Les Collectivités S'engagent (ou Pas)." *Benoit Deldicque*, December 12, 2013. <http://www.benoit-deldicque.com/blog/concours-opendata-les-collectivites-sengagent-ou-pas/>.

Département de la Gironde. "Mise En Relation Des Données et Des Services". *Ressourcerie Datalocale*. Accessed July 10, 2014. <http://www.datalocale.fr/drupal7/content/mise-en-relation-des-donn%C3%A9es-et-des-services>.

Département de Saône-et-Loire. "Chef de Projet Open Data - Usages Numériques Innovants". *Département de Saône-et-Loire*, March 24, 2014. http://www.cg71.fr/le-departement/offres-demploi/detail-de-loffre-demploi/emploi/chef-de-projet-open-data-usa/listepointer/tx_eoffresdemploi_pi1%255BuidPage%255D%3D2%26cHash%3D28ec99955e9ed3ad8b94200f08d6e9d9/#.UzV4tYWF8jo.

Le Département de Saône-et-Loire, 'Retour Sur Le Comité D'éthique Open Data Du 27 Mars 2013', *Open data 71*, 2013. Accessed 12 June 2014. <http://www.opendata71.fr/toutes-les-actualites/90-comite-ethique-mars-2013>.

Didier Frochot, "Les Données Publiques, Les Documents Administratifs et Les Autres," *Les-Infostrateges.com*, November 16, 2013, <http://www.les-infostrateges.com/article/0311107/les-donnees-publiques-les-documents-administratifs-et-les-autres>.

Eaves, David. "Let's Do an International Open Data Hackathon." *Eaves.ca*, October 29, 2010. <http://eaves.ca/2010/10/29/lets-do-an-international-open-data-hackathon/>.

Favre, Cédric. "Interview de Spécialiste : Cédric Favre, Juriste Rédacteur Du Guide Pratique Des Données Publiques." *Data Publica*, Juillet 2011. <http://www.data-publica.com/content/2011/07/interview-de-specialiste-cedric-favre-juriste-redacteur-du-guide-pratique-des-donnees-publiques/>.

"La feuille de route du Gouvernement en matière d'ouverture et de partage des données publiques." Etalab. *OverBlog*, February 28, 2013. <http://www.etalab.gouv.fr/article-la-feuille-de-route-du-gouvernement-en-matiere-d-ouverture-et-de-partage-des-donnees-publiques-115767801.html>.

Fondation Internet Nouvelle Génération, and Amandine Bruguière. "Les Compétences 'Data.'" *FING*, September 2, 2013. <http://www.fing.org/?Les-competences-Datalang=fr>.

Fondation Internet Nouvelle Génération. "Une Campagne Autour de l'OpenData". *Fing*. Accessed May 31, 2014. <http://fing.org/?Une-campagne-autour-de-l-OpenDatalang=fr>.

France. Commission d'accès aux documents administratifs. "Le Rôle de La Cada". Accessed July 7, 2014. <http://www.cada.fr/le-role-de-la-cada,6077.html>.

Free Software Foundation, 'Qu'est-Ce Que Le Logiciel Libre?', *Le système d'exploitation GNU*, 2013. Accessed May 27, 2014. <http://www.gnu.org/philosophy/free-sw.fr.html>.

Fréquence Ecoles, and Violaine Lemaitre. "Découvrez Ici L'intégralité Du Dispositif Pédagogique #dataviz_RRA | Fréquence Écoles." *Fréquence Ecoles*, June 25, 2014. http://frequence-ecoles.org/2014/06/25/decouvrez-ici-lintegralite-du-dispositif-pedagogique-dataviz_rra/.

Frochot, Didier. "Données Publiques – 1 : De L'accès Aux Documents Administratifs À L'open Data." *Les Infostrateges.com*, February 21, 2014. <http://www.les-infostrateges.com/actu/14021770/donnees-publiques-1-de-laces-aux-documents-administratifs-a-lopen-data>.

Galy, Suzanne. "Les Outils de Data Visualisation" *AEC Agence Des Initiatives Numérique*, June 13, 2013. <http://www.aecom.org/Vous-informer/Actualites2/Les-outils-de-data-visualisation>.

Galy, Suzanne. "Open Data School : 130 Étudiants Pour Un Maximum de Créativité" *AEC Agence Des Initiatives Numérique*, February 3, 2012. <http://www.aecom.org/Vous-informer/Actualites2/Open-Data-School-130-etudiants-pour-un-maximum-de-creativite>.

Galy, Suzanne. "Une Expérience de Data Journalism Est Lancée À Bordeaux." *AEC Agence Des Initiatives Numérique*, January 23, 2012. <http://www.aecom.org/Vous-informer/Actualites2/Une-experience-de-data-journalisme-est-lancee-a-Bordeaux>.

Guimard, Emmanuel. "« Open Data » : Les Collectivités Sont En Quête D'un Modèle Économique." *Lesechos.fr*, January 6, 2012. <http://www.lesechos.fr/01/06/2012/LesEchos/21196-025-ECH --open-data----les-collectivites-sont-en-quete-d-un-modele-economique.htm>.

Harbert, Tam. "Big Data, Big Jobs?" *Computer World*, September 20, 2012. http://www.computerworld.com/s/article/9231445/Big_data_big_jobs?pageNumber=2.

"Un Hackathon Késako?". Accessed July 13, 2014. <http://hackdatapaca.net/post/25996058806/hackathon>.

Initiative de Budapest pour l'accès ouvert. "Accès Libre." *Libre Accès À L'information Scientifique et Technique*, February 14, 2002. <http://openaccess.inist.fr/?+-Libre-Acces-+>.

"Les Interconnectés : Une Association Indépendante Qui a Pour Objectif de Favoriser La Diffusion Des Usages Des Nouvelles Technologies Dans Les Collectivités Locales.," February 6, 2013. <http://www.interconnectes.com/fr/presentation/>.

"Jérémy, Chef de Projet Open Data Pour La Ville de Montpellier : 'Je Ne Suis Pas Informaticien, Je Passe Peu de Temps Devant Mon Ordinateur.'" *L'étudiant.fr*, juin 2013. <http://www.letudiant.fr/metiers/metiers-du-numerique-des-opportunités-pour-les-jeunes-diplomes/jeremie-chef-de-projet-open-data-pour-la-ville-de-montpellier.html>.

Kaplan, Daniel. "Big Data, Grande Illusion?" *InternetActu.net*, November 4, 2012. <http://www.internetactu.net/2012/04/11/big-data-grande-illusion/>.

"Keolis Rennes.com : Les Données Transport de Rennes Métropole En Accès Libre." *Keolis Rennes.com*, May 8, 2013. <http://data.keolis-rennes.com/fr/accueil.html>.

Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety". *MetaGroup*. 2001. <http://blogs.gartner.com/douglaney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

Libertic. "Le Citoyen a T-Il Une Place Dans L'open Data ?" *Libertic*, February 17, 2012. <http://libertic.wordpress.com/2012/02/17/le-citoyen-a-t-il-une-place-dans-lopen-data/>.

Libertic. "Lettre Ouverte En Faveur de L'ouverture Des Données Publiques de Nantes," December 21, 2010. <http://libertic.wordpress.com/2010/12/21/petition-en-faveur-de-louverture-des-donnees-publiques-de-nantes/>.

Ménégo, Karine. "Open Data : La CU de Bordeaux Change de Licence, Espérant Ainsi Générer Des Emplois - Lagazette.fr." *Lagazette.fr*, January 25, 2013. <http://www.lagazettedescommunes.com/151375/open-data-la-cu-de-bordeaux-change-de-licence-esperant-ainsi-generer-des-emplois/#fn-151375-1>.

Mazon, Romain. "L'ouverture Des Données Publiques Devient Une Obligation Pour Les Collectivités Locales." *Lagazette.fr*, April 19, 2013. <http://www.lagazettedescommunes.com/157970/l%E2%80%99ouverture-des-donnees-publiques-devient-une-obligation-pour-les-collectivites-locales/#fn-157970-1>.

Naudet, Severin. "Les Nouveaux Usages Appellent Une Nouvelle Gouvernance," May 19, 2011. <http://severinnaudet.com/article-les-nouveaux-usages-appellent-une-nouvelle-gouvernance-74177859.html>.

"Mozilla at a glance". Mozilla blogs. Accessed Juin 20, 2015. <https://blog.mozilla.org/press/ata glance/>

OECD. *Incidence de La Société de L'information Sur Le Processus D'élaboration Des Politiques et Sur La Qualité de La Démocratie*, 1999. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=PU/MA\(98\)15&docLanguage=fr](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=PU/MA(98)15&docLanguage=fr).

"Open Data, Ouverture Des Données Publiques". Espace Mendès-France: culture médiation scientifiques, March 16, 2013. <http://emf.fr/14557/open-data-ouverture-des-donnees-publiques/>.

Open Knowledge Foundation, 'Open Government Data Camp 2011', *Open Government Data Camp 2011*, 2011 .<http://ogdcamp.org>.

Open Knowledge Foundation. "Définition Du Savoir Libre | Open Definition,". Accessed February 22, 2013. <http://opendefinition.org/okd/francais/>.

Open Knowledge Foundation. "Définition Du Savoir Libre." *Open Definition*, February 22, 2013. <http://opendefinition.org/okd/francais/>.

Open Knowledge Foundation. "Ecole Des Données" Accessed July 16, 2014. <http://ecoledesdonnees.org/>.

Open Knowledge Foundation. "International Open Data Hackathon2014". Accessed 25 March 2014. <http://opendataday.org/>.

Open Knowledge Foundation. "Open Data Index," October 28, 2013. <https://index.okfn.org/country/>.

Open Knowledge Foundation. "Open Government Data." *Open Government Data*. Accessed June 8, 2014. <http://opengovernmentdata.org/>.

Open Society Justice Initiative. "Constitutional Provisions, Laws and Regulations," *Right2Info.org*. Accessed May 22, 2014. <http://www.right2info.org/laws/constitutional-provisions-laws-and-regulations#section-1>.

"Opendata Saint-Brieuc". *Création Du Collectif Saint-Brieuc OpenData*. Accessed July 4, 2014. <http://opendata.la-matrice.org/>.

Opendata Toulouse. "Lettre Au Grand Toulouse". *Opendata Toulouse*, October 22, 2011. <http://opendata-toulouse.numahel.net/doku.php/lettre>.

OpenData Touraine #37. "Lettre Ouverte Aux Élus d'Indre-et-Loire En Faveur de La Libération Des Données Publiques". *Opendata37: Pour La Libération Des Données Publiques d'Indre-et-Loire (37)*. Accessed July 4, 2014. <http://opendata37.fr/>.

O'Reilly, Tim. "Web 2.0: Compact Definition?" *O'Reilly Radar*, 2005. <http://radar.oreilly.com/2005/10/web-20-compact-definition.html>.

"Le Premier ministre réaffirme son attachement à l'ouverture et à la mise à disposition des données publiques", *etalab*, 2013. <http://www.etalab.gouv.fr/article-le-premier-ministre-reaffirme-son->

attachement-a-l-ouverture-et-a-la-mise-a-disposition-des-donnees-p-118492716.html.

Proxima Mobile. "Les Perspectives Des Services Mobiles Pour Les Collectivités Locales." *Proxima Mobile*, May 5, 2010. <http://www.proximamobile.fr/article/les-perspectives-des-services-mobiles-pour-les-collectivites-locales>.

Regards Citoyens. "Licences OpenData : L'APIE Grille La Priorité À ÉtaLab et Invente Le Pseudo-Libre," July 19, 2011. <http://www.regardscitoyens.org/licences-opendata-lapie-grille-la-priorite-a-etalab-et-invente-le-pseudo-libre/>.

"Rencontre Avec Philippe Grosvalet, Président Du Conseil Général de Loire Atlantique". *EntrepreNantes*, March 16, 2012. <http://www.entreprenantes.com/interviews/298-rencontre-avec-philippe-grosvalet-president-du-conseil-general-de-loire-atlantique.html>.

"Répartition Des Compétences." *Collectivités Locales*, 2014. http://www.collectivites-locales.gouv.fr/files/files/tableau_repartition_compетенces_valideCILapublier%281%29.pdf.

Rudloff, Pierre. "Un Premier Pas Vers l'Open Data À Strasbourg". Accessed July 4, 2014. https://secure.avaaz.org/fr/petition/Ouverture_des_donnees_de_la_Communaute_Urbaine_de_Strasbourg/.

Stewart, Darin. "Big Content: The Unstructured Side of Big Data." *Gartner*, May 1, 2013. <http://blogs.gartner.com/darin-stewart/2013/05/01/big-content-the-unstructured-side-of-big-data/>.

Sunlight Foundation. "Open Data Policy Guidelines." *Sunlight Foundation*, 2014. <http://sunlightfoundation.com/opendataguidelines/>.

"Top 500 Global Sites", *Alexa Internet*. Accessed May 28, 2014. <http://www.alexa.com/topsites>.

Tronc, Jean-Noël. 'Le Rôle Des Pouvoirs Publics Dans L'accès À L'information Faciliter L'accès Aux Réseaux et Services', *Unesco*, 2000. <http://www.google.fr/url?sa=trct=jq=esrc=ssource=webcd=3ved=0CD8QFj>

ACurl=http%3A%2F%2Fwebworld.unesco.org%2Finfoethics2000%2Fdocuments%2Fpaper_tronc.rtf&rtf=7bWJUX4AcfJOFS3gPgKusg=AFQjCNFLMLdr58OIlfPyHprXlx03ihMgsAsig2=otl12QzLrQ1J7A7VngOtZw.

Valentin, Jérémie. "Cartographie Contributive Quatrième Journée - Cartopartie #4 - Données Publiques Ouvertes - Montpellier Territoire Numérique." *Montpellier Territoire Numérique - Données Publiques Ouvertes*, February 10, 2012. <http://opendata.montpelliernumerique.fr/Cartographie-contributive,95>.

Valentin, Jérémie. "Infolab # 6 : Découverte Par Les Lycéens Du Monde Des Données." - *Données Publiques Ouvertes - Montpellier Territoire Numérique*, June 13, 2014. <http://opendata.montpelliernumerique.fr/Infolab>.

Valentin, Jérémie. "Premier Retour Sur La Mise À Disposition En Open Data et En Temps Réel Des Résultats Des Élections - Données Publiques Ouvertes - Montpellier Territoire Numérique." *Montpellier Territoire Numérique - Données Publiques Ouvertes*, March 26, 2014. <http://opendata.montpelliernumerique.fr/Premier-retour-sur-la-mise-a>.

Xavier, Crouan. "Libération Des Données Publiques : Faits, Enjeux et Perspectives." *Le Blog Du Modérateur*. Accessed February 20, 2014. <http://www.blogdumoderateur.com/liberation-des-donnees-publiques-faits-enjeux-et-perspectives-par-xavier-crouan/>.

Zimmer, Michael. "More On the 'Anonymity' of the Facebook Dataset – It's Harvard College," 2008. <http://www.michaelzimmer.org/2008/10/03/more-on-the-anonymity-of-the-facebook-dataset-its-harvard-college/>.

Articles de journaux

Leloup, Damien, and Eric Numès. "Nathalie Kosciusko-Morizet veut une opération 'transparence des données publiques.'" *Le Monde.fr*, November 16, 2011. http://www.lemonde.fr/politique/article/2011/11/16/nathalie-kosciusko-morizet-veut-une-operation-transparence-des-donnees-publiques_1604708_823448.html.

"Des statistiques d'Etat à géométrie variable." *Le Monde.fr*, Décembre 12, 2011. http://www.lemonde.fr/idees/article/2011/12/14/des-statistiques-d-etat-a-geometrie-variable_1618346_3232.html.

Thèses et Mémoires

Chailloux, Alexis. "Data-Hari : Le Journalisme Sacrifié Sur L'autel Des Données?" Mémoire de Master 2, Université de Paris IV - Sorbonne, 2012. http://www.slideshare.net/alexis_chailloux/mmoire-datajournalisme?ref=http://projetjournal.org/un-journalisme-sacrifie/.

Goeta, Samuel. "Open Data : Qu'ouvre-T-on Avec Des Données Publiques ?" Mémoire de Master 2, Celsa, 2012. <http://fr.scribd.com/doc/98398429/Memoire-Open-data-qu-ouvre-t-on-avec-des-donnees-publiques-Samuel-Goeta-Celsa-2012>.

Rekik, Y. "Modélisation et Manipulation Des Documents Structurés : Une Approche Modulaire, Flexible et Évolutive." Thèse de doctorat, EPFL, 2001.

Teresi, Laurent. "La commercialisation des données publiques." Thèse de doctorat, Université Paul Cézanne, 2007.

Sites web

"Agence Du Patrimoine Immatériel de L'état - APIE," Accessed August 3, 2013. <http://www.economie.gouv.fr/apie>.

Association Libertic. "Blog de L'association Libertic,". Accessed June 17, 2014, <http://libertic.wordpress.com/>.

"Blog." *Données Publiques Ouvertes - Montpellier Territoire Numérique*. <http://opendata.montpelliernumerique.fr/Blog>.

Brest, '@ Brest - Relier et Partager Autour Du Web'. Accessed 7 June 2014. <http://www.a-brest.net/>.

"Data Ville." Accessed July 20, 2014. <http://www.dataville.fr/index/premium>.

"Datacatalogs.org". Accessed March 19, 2014. <http://datacatalogs.org/>

"Fabrique d'Objets Libres," 2013. <http://www.fablab-lyon.fr/>.

"Fondation Internet Nouvelle Génération". Accessed July 4, 2014. <http://fing.org/?lang=fr>.

"GFII - Les Acteurs Du Marché de L'information et de La Connaissance". Accessed July 4, 2014. <http://www.gfii.fr/fr/>.

Institut National de la Statistique et des Etudes Economiques. Accessed January 01, 2010. <http://www.insee.fr/fr/default.asp>.

"Montpellier Territoire Numérique." *Montpellier.fr*. Accessed July 23, 2014. <http://www.montpellier.fr/3421-montpellier-territoire-numerique.htm>.

"Les Musées En France." Accessed August 15, 2014. <http://labs.antidot.net/museesdefrance/#search>.

Open data 69. "Opendata69 : Collectif Pour L'ouverture Des Données Dans Le Rhône". Accessed July 4, 2014. <http://opendata69.org/>.

"Open Data Bourgogne". Accessed July 4, 2014. <http://opendatabourgogne.org/>.

"Open Government." *Data.gov*. <https://www.data.gov/open-gov/>.

Open Government Partnership. "What Is the Open Government Partnership?" *Open Government Partnership*. Accessed October 26, 2015. <http://www.opengovpartnership.org/>.

"Opendata France". Accessed June 17, 2014. <http://opendatafrance.net/>.

"Opendata France". <http://opendatafrance.net/>.

"The Tropical Disease Initiative." Accessed May 28, 2014. <http://sgt.cnag.cat/TDI/>.

StrasWeb. "Open Data". Accessed July 4, 2014. <https://strasweb.fr/category/open-data/>.

Union européenne. *Portail des données ouvertes de l'Union européenne*. Accessed June 24, 2014. <https://open-data.europa.eu/fr/data>

Ville à Vélo. "Carte Des Aménagements Cyclables Du Grand Lyon." Accessed August 16, 2014. <http://carte.lavilleavelo.org/>.

Présentations et actes de conférences

Berners-Lee, Tim. "The next Web." Présenté à *Ted2009*. Long Beach, California, 2009.
http://www.ted.com/talks/tim_berniers_lee_on_the_next_web/transcript.

Berners-Lee, Tim. "Open, Linked Data for a Global Community." Présenté à *Gov2.0 expo*, Washington, DC, 2010.

Berthault, Denis. "PSI re-use in France : position of the industry". Présenté à *11th PSI Group meeting*. Luxembourg, 2008.
www.epsiplatform.eu/sites/default/files/ezpublish_media/berthault.pps

Bertrand, Paul. "Les digital humanities sont-elles solubles dans le Steampunk?" Présenté à *THATCamp Paris*. Paris 2012. Paris: Éditions de la Maison des sciences de l'homme, 2012.
<http://books.openedition.org/editionsmsh/327>.

Boucharlat, Marion. "Datavisualisation : Définition et Historique." Présenté à *Dataviz – Infographies : Data + Lyceens / Rhone Alpes*. Lyon, 2014.

Denis, Jérôme, and Samuel Goëta. "La Fabrique Des Données Brutes. Le Travail En Coulisses de L'open Data." Dans *Penser L'écosystème Des Données. Les Enjeux Scientifiques et Politiques Des Données Numériques*. Paris, 2013. <https://halshs.archives-ouvertes.fr/halshs-00990771/document>.

Népote, Charles. "Le Potentiel de La Réutilisation Des Données Publiques Pour L'innovation Sociale". Dans *L'ouverture Des Informations Publiques : Un Enjeu Pour Le Développement de L'économie Numérique*. Paris: GFII, 2010.
<http://www.gfii.fr/uploads/docs/20-mai-2010-le-potentiel-de-la-reutilisation-des-donnees-publiques-pour-l-innovation-sociale.pdf?symfony=ptfpnqk88rfpla3sifo92u64e7>.

Népote, Charles. "Un Nouveau Monde de Données". Présenté à *Dataviz – Infographies : Data + Lycéens / Rhône-Alpes*. Lyon, 2014.

O'Reilly, Tim, and John Battelle. "Web Squared: Web 2.0 Five Years On." Présenté à *Web 2.0 Summit*. San Francisco, Calif., 2009.
<http://www.web2summit.com/web2009/public/schedule/detail/10194>.

Open government working group. "The 8 Principles of Open Government Data." Dans *Open Government Data Workshop*. Sebastopol, CA: opengovdata.org, 2007. <http://opengovdata.org/>.

Poupeau, Gauthier. "A La Découverte Des Musées de France, Exemple D'agrégation de Données Ouvertes Pour La Réalisation D'une Application Web." Présenté à *Organisation et Traitement Des Open Data*. Lyon, 2014.

Ronai, Maurice. "Identification Des Gisements et Des Causes de Non Exploitation" Dans *Diffusion des données publiques*. Paris, 2002.
<http://www.gfii.fr/fr/document/colloque-diffusion-des-donnees-publiques#maurice%20ronai>.

THATCamp Saint-Malo 2013 : Non actes de la non conférence. Paris : Éditions de la Maison des sciences de l'homme, 2014.
<http://books.openedition.org/editionsmsmh/2181>.

Vaudano, Maxime. "En Quoi Les Données et Les Dataviz Changent Le Métier D'un Journaliste ?" Présenté à *Dataviz – Infographies : Data + Lyceens / Rhone Alpes*. Lyon, 2014.

Vanhaverbeke, Wim. "L'open Innovation, Qu'est-Ce Que C'est?" Présenté à *Open Innovation Day*. Lyon, 2013.

Zaslavsky, Arkady, Charith Perera, and Dimitrios Georgakopoulos. "Sensing as a Service and Big Data." Dans *Proceedings of the International Conference on Advances in Cloud Computing (ACC)*. Bangalore, India, 2012.
<http://arxiv.org/abs/1301.0159>.

Encyclopédies et dictionnaires

"DBpedia." *Wikipédia*, July 20, 2014.
<http://fr.wikipedia.org/w/index.php?title=DBpedia&oldid=105149058>.

Online Dictionary for Library and Information Science, "Data," accessed June 11, 2014, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx>.

Annexes

Annexe A : Liste des conférences et séminaires auxquels nous avons assistés.

27 Juin 2014 - Lyon

L'open data : nouvelle richesse du web, mais pour qui ? / Programme Big Open Linked Data – BOLD (Université Lyon 3) et Programme Matinées de l'Open Data en Rhône-Alpes – Modra (Elico, Équipe de recherche de Lyon en sciences de l'Information et de la Communication)

Parmi les intervenants :

- Vincent Mandinaud, Chef de projet, Agence nationale pour l'amélioration des conditions de travail. «Territoire et attractivité des entreprises».
- François Perrussel-Morin, Directeur interrégional, Direction centre-est Institut national de l'information géographique et forestière (IGN). « L'Evolution de l'offre de service public IGN ».

19 Mai 2014 - Lyon

Dataviz – Infographies : Data + Lycéens, Journée-débat autour du data journalisme et de l'open data / Région Rhône Alpes et Fréquence Ecoles

Parmi les intervenants :

- Marion Boucharlat, graphiste, OWNI. "Datavisualisation : Définition et Historique".
- Maxime Vaudano, Data journaliste, le Monde.fr. "En Quoi Les Données et Les Dataviz Changent Le Métier D'un Journaliste ?"
- Charles Népote, Chef de projet "Partage des données publiques" à la FING. "Un Nouveau Monde de Données".

14 Mai 2014 - Lyon

Organisation et traitement des Open data / Elico

Les intervenants :

- Gauthier Poupeau , Concepteur/Réalisateur, Antidot. « A la découverte des musées de France, exemple d'agrégation de données ouvertes pour la réalisation d'une application web ».
- Laurent Chevereau, Centre d'études sur les réseaux, les transports, l'urbanisme et les constructions publiques (CERTU). « Open data et mobilité : Stratégies des collectivités et enjeux d'interopérabilité ».

14 février 2014 - Lyon

L'open data peut-il servir le patrimoine ? / Elico

Les intervenants :

- Cyril Longin, Directeur des Archives Municipales de Saint-Etienne et Chef du projet Open data. « Open data et archives publiques : les frères ennemis ? »
- Karine Feige, Directeur de Projet Sitra. « Le développement des stratégies d'open data dans le secteur du tourisme institutionnel ».

31 Janvier 2014 - Lyon

Les données de l'Open Data : Quels usages pour quels publics? / Elico

Les intervenants :

- Thomas Lacroix, Directeur Technique – CTO, CoSMo Company
- Romain Mazon, Rédacteur en chef web, Pôle collectivités locales, La Gazette des communes. « Un enrichissement du travail journalistique pour la Gazette des communes ».
- Charles Népote, "Partage des données publiques", Fing

20 Décembre 2013 - Lyon

Les données de l'Open Data : hétérogénéité des données / Elico

Les intervenants :

- Catherine Bouteiller, Laboratoire d'Economie des Transports
- Christine Buisson, LICIT - Laboratoire Ingénierie Circulation Transports. « Collecter des données rares pour les partager : le cas de MOCOPo »
- Didier Gorrion, Transpacity

3 Décembre 2013 - Lyon

Journée Open Data / Association Les Interconnectés

Parmi les intervenants :

- Simon Chignard, Consultant Open data. « La donnée, une matière première bien étrange ».
- Jean-Marie Bourgogne, Directeur du programme « Montpellier Territoire Numérique ».
- Sophie Tendeiro, Chef du projet Open Data, Region PACA.
- Jean-Christophe Elineau conseiller municipal chargé de l'innovation numérique et de la communication, Ville de Brocas
- Gildas Laéron, conseiller municipal en charge des TIC, ville de Grenoble et . Karine Devenyns, Business & Decision Eolas. « Présentation du projet Datalyse ».
- Sarah Ladon et Grégory Blanc, Grand Lyon. « Présentation du projet SmartData du Grand Lyon »

2 et 3 Décembre 2013 - Lyon

10^{eme} forum des Interconnectés / Association Les Interconnectés

Parmi les intervenants :

Quel avenir numérique ? Rétrospective / Prospective - 10 ans d'innovation dans les territoires

- Christophe Bernard, Secrétaire général, AdCF
- Céline Colucci, Déléguée générale, Réseau des territoires innovants
- Constance Mollat, conseillère municipale en charge du développement de la Cité numérique, Bordeaux
- David Hecq, vice président, Communauté Urbaine d'Arras
- Joël Duquenoy, président de la Communauté d'agglomération de St Omer
- Bruno Marzloff, Sociologue, société Chronos

Smart City : peut-on planifier la cité désirable ?

- Karine Dognin-Sauze - vice-présidente du Grand Lyon Innovation et nouvelles technologies
- Florence Barale - Ville de Nice et Métropole Nice Côte d'Azur
- Béatrice de François - vice-présidente de la Communauté urbaine de Bordeaux
- Francis Jutand - Directeur Scientifique de l'institut Mines-Telecom

La collectivité face à l'homo numericus : chef de file, animateur, garant ?

- Emile Hooge - Nova 7
- Sébastien Semeril - Rennes Agglomération
- Eric Legale - Issy-les-moulineaux
- Akim Oural - Communauté urbaine de Lille et membre du CNN

21 Novembre 2013 - Lyon

La politique des données personnelles : Big data ou contrôle individuel ? / ENS de Lyon

Les intervenants :

- Yves-Alexandre de Montjoye, MediaLab, MIT (Boston) Big but personal data
- François Taïani, Professeur à l'ESIR / Université de Rennes 1, spécialiste de la décentralisation des données
- Antoinette Rouvroy, Chercheur qualifié FNRS (philosophie du droit) au Centre de Recherche en Information, Droit et Société (CRIDS) de l'Université de Namur.
- Daniel Kaplan, Fondation Internet Nouvelle Génération (Fing)
- Dominique Cardon, sociologue (Orange Lab)
- Stéphane Grumbach, chercheur à INRIA Lyon

15 Novembre 2013 - Lyon

Territoire et producteurs de données : Open Data à la carte / Elico

Les intervenants :

- Florence Martinez, Responsable de l'unité Diffusion des données et géoservices, service Information géographique, Grand Lyon. « Grand Lyon SmartData : une plateforme territoriale de données au service de l'innovation - Contexte et mise en œuvre au Grand Lyon ».
- Florian Farge, OpenStreetMap. « La réappropriation des cartes par la contribution participative »

7-8 novembre 2013 - Clermont-Ferrand

Colloque international « Open Data : enjeux communicationnels et sociétaux » / Laboratoire Communication et Solidarité (EA 4647).

Parmi les intervenants :

- Simon Chignard, consultant Open data. « L'ouverture des données publiques : un état des lieux, une approche des enjeux à venir ».
- Richard Swetenham, Head of Unit - Access to Information (E4) DG Information Society European Commission. "Les enjeux de l'Open Data en Europe".
- Valérie Peugeot, Prospectiviste chez Orange Labs, Présidente Vecam, Vice-Présidente Conseil National du Numérique. « L'Open data entre transparence et pouvoir d'agir : quel rôle en démocratie ? ».
- Christophe Boutet, Membre du Conseil d'Administration de Regards citoyens,
- Patrice Carré, Président du Conseil scientifique de Décider Ensemble.
- Mathieu Jeandron, Chef du service stratégie et urbanisation, adjoint au directeur, Direction interministérielle des systèmes d'information et de communication
- Sophie Tendeiro, Chef de projet Services numériques innovants et mutualisation des données publiques, Service Innovation et Economie Numériques, Région PACA. « Open PACA : la démarche partenariale

de la Région Provence-Alpes-Côte d'Azur ».

- Evelyne Leroy, Chargée de mission Développement numérique, Mission Développement numérique, Conseil Régional d'Auvergne
- Philippe Benmoussa, Responsable du Marché Collectivités, Société Sopra Group. « Open Data : quelles ambitions pour le Secteur Privé et ses Entreprises ».
- Pierre Valentin, Community Manager France, Moovit Application Mobile - TranzMate et Samy Guesmi, Enseignant Chercheur, Ecole Supérieure de Commerce Idrac
- Romain Lalanne, Responsable Open Data à la SNCF. « Open Transport : repenser la mobilité avec les données ».
- Thierry Chanier, Laboratoire de Recherche sur le Langage (LRL), Département de Linguistique, Université Blaise Pascal (Clermont 2). OpenData : au-delà des publications, le partage des données de la recherche en sciences humaines. Pour qui ? Comment ?
- Jérémy Valentin, Docteur en Géographie, chef de projet Open Data sur Montpellier
- Sabine Blanc, Journaliste web freelance, ancienne d'OWNI, média spécialisé dans le data journalisme. « Le data journalisme, un exercice difficile »

26 Juin 2013 - Lyon

Journée Open Innovation / ARDI Rhône-Alpes et OpenAlps

Parmi les intervenants :

- Wim Vanhaverbeke, Professeur, University of Hasselt . "L'open Innovation, Qu'est-Ce Que C'est?"
- Martin Duval, Bluenove. « Open data : quel enjeux et opportunités pour l'entreprise ».
- Mohand-Said Hacid, Chercheur, Laboratoire d'Informatique en Imagerie et Systèmes d'information (LIRIS)
- Fidélia Ibekwe-SanJuan, Maitre de Conférences, Elico, Université de Lyon 3

10 Mai 2013 - Lyon

Matinale Les 4V du BigData : Volume, Variété, Vitesse, Valeur / Doc Forum et le Cluster Edit

Les intervenants :

- Christian Dubourg , Directeur Ever Team, coordinateur du livre blanc Aproped Content Analytics
- Pierre Col, Antidot
- Virgile Devaux, Chef de projet Big Data, unité Business & Strategic Developments, Atos Worldline
- Henri Verdier, Directeur d'Etalab, Service du Premier Ministre chargé de l'ouverture des données publiques et du développement de la plateforme française Open Data
- Grégory Bourbon, responsable du service Web & Organisation du Grand Lyon
- Mohand Saïd Hacid et Jean-François Boulicaut laboratoire LIRIS. « La donnée, c'est l'or noir du 21e siècle »
- Yves de Montcheuil, AFDEL et VP Marketing chez Talend

29 Novembre 2012 - Paris

Huitièmes Rencontres de Décider ensemble : « Ouverture des données publiques et participation : Quels enjeux démocratiques? » / Association Décider ensemble

Les intervenants :

- Claire Gallon, administratrice de LiberTIC. « L'Open Data en deux mots et une vidéo »
- Laurence Monnoyer-Smith, professeure à l'Université de Technologie de Compiègne (UTC). « Présentation d'une étude réalisée par Décider ensemble »
- Patrice Carré, président du Conseil scientifique de Décider ensemble. « En quête d'une ouverture de l'information et des "données" dans le temps long : quelle histoire possible? »
- Simon Chignard, consultant, auteur de L'open data, comprendre

l'ouverture des données publiques

- Florence Durand-Tornare, directrice associée de La Suite dans les Idées, fondatrice et déléguée de Villes Internet. « L'Open Data à la française : un outil de transparence ou de participation ? »
- Béatrice de François, vice-présidente de la Communauté Urbaine de Bordeaux
- Cyril Lage, responsable de Démocratie ouverte
- Alban Martin, directeur des médias sociaux d'Orange, Celsa Paris IV Sorbonne
- Benjamin Ooghe-Tabanou, co-fondateur de Regards Citoyens
- Alexandre Quintard Kaigre, chargé de mission, Etalab
- Yann Moullet-Boutang, professeur à l'UTC. « L'écosystème de l'open data : peut-on parler d'innovation ascendante ? »
- Daniel Kaplan, délégué général de la Fondation Internet Nouvelle Génération (FING). « Vivre dans un monde de données : quels enjeux politiques ? »
- Amandine Brugière, chef de projet « Partage des données publiques », FING
- Caroline Goulard, co-fondatrice, Dataveyes
- Jean-Marc Lazard, CEO, OpenDataSoft
- Gaël Musquet, président d'Open Street Map
- Valérie Peugeot, prospectiviste, Orange Labs
- Sébastien Sémeril, vice-président de Rennes Métropole en charge du numérique

17-18 Octobre 2011 - Lyon

Biennale du numérique 2011 : médiations, savoirs, innovations / École Nationale Supérieure des Sciences de l'Information et des Bibliothèques (Enssib)

Parmi les intervenants :

- Joumana Boustany, Maître de conférences à l'université Paris Descartes. « L'open data en France ».
- Jean-Baptiste Gabellieri, Co-fondateur de l'association Regards Citoyens et Joumana Boustany, Maître de conférences à l'université Paris Descartes. « Open data et création de services ».

Annexe B : Guide d'entretien avec les chefs du projet open data dans les collectivités

I. L'intérêt de l'Open data locale

1. Quel est l'intérêt de l'open data dans votre collectivité?
2. Quels sont vos objectifs à court terme et à long terme du projet Open data ?
3. Quels sont les facteurs clés qui contribuent et influent votre engagement dans un projet Open data ?
4. Est-ce l'Open data dans votre collectivité est une initiative prise par les élus ou un service au sein de la collectivité?
5. Est-ce que tous les acteurs internes (élus, directeur général des services, responsables des services municipaux, chef du projet open data) ont la même vision de l'open data ?
6. Quel est l'effet du mouvement Open data sur la mission, le mode de fonctionnement et l'organisation interne des collectivités? Parle-t-on d'une mutation ?
7. Pensez-vous que les collectivités sont en **situation** de **sous-exploitation de leurs données** ?

II. La mise en œuvre du projet

8. Quelles sont les étapes clés dans votre projet open data ? Quelle est la durée de chaque étape ? Avez-vous créé un guide pratique ou un plan d'action ?
9. Quels sont les points et les dates remarquables du déroulement du projet ?
10. Quel est le **coût** d'investissement du projet? Est-ce que le budget d'un projet open data fait partie d'autre ligne budgétaire au sein de la collectivité ? Comment est-il divisé? Comment le budget a évolué avec le temps?
11. Quelle licence avez-vous choisi? Et pourquoi ?

Equipe Open data

12. A quel service municipal est confié le projet Open data ? Qui est le personnel responsable du projet ? Quelles sont leurs tâches quotidiennes ?
13. Quel est le profil du chef de projet Open Data ?

14. Ya t-il des compétences requises par les responsables du projet?
Avez-vous eu des formations ?
15. y-a-t-il un rôle accordé au PRADA (personne responsable à l'accès aux documents administratifs)?

Collecte et sources de données

16. Quelle est la modalité de collecte des données ?
17. Quelles sont vos sources de données ?
18. Qui décide quelles données à ouvrir ?
19. Quels services municipaux sont concernées par l'open data?
Sont-ils convaincus de l'intérêt de la démarche ?
20. Existe-il d'autres sources de données sur la collectivité ?
21. Quel est le degré d'implication des citoyens dans la collecte et la production de données (crowdsourcing)? Questionnez-vous la fiabilité des données produites par les citoyens ?
22. Comment améliorer la collecte et la pertinence des données ?
23. Quelle est votre politique de mise à jour des données ?

Quelles sont les données ouvertes ?

24. Quels sont les domaines de vos données ouvertes ? Quel domaine est prioritaire ? Pourquoi ?
25. Avez-vous ouvert des données qui ont jusqu'alors gardées inaccessible ? Si oui, donnez un exemple.
26. Selon vous, quelles sont les données qui apportent plus de transparence aux administrations ? sont-elles ouvertes ? Si non pourquoi ?

Ouverture mutualisée / coopération intercommunale

27. Est-ce que vous êtes contre ou pour un hébergement mutualisé de vos données sur la plateforme data.gouv.fr ou sur les plateformes des communautés urbaines ? Pourquoi ?
28. Quelles relations existent entre les initiatives des communes, agglomérations, départements et régions?
29. Si c'est le cas d'un hébergement mutualisé, qui est le gestionnaire des données ? Est-ce un tiers ?

Aspects techniques de l'ouverture

30. Quel traitement subissent les fichiers de données avant de les mettre en ligne ? (Contrôle, changement de format, indexation, numérisation)
31. Quels outils utiliser pour la diffusion des données ? Qui est le gestionnaire des données ? Qui peut ajouter un jeu de données ? y-a-t-il plusieurs fournisseurs ?
32. Quel est l'intérêt des solutions libres et open source dans l'open data ?
33. Quel est le format de publication des données ? (téléchargement en vrac, API, visualisation?) Est-ce que l'open data va au-delà de l'ouverture de données brutes et des formats structurés ?

Les acteurs externes du projet open data

34. Avez-vous créé des partenariats avec des acteurs externes dans le cadre du projet open data ? si oui quelle est la nature de ces partenariats ?
35. Existe-t-il un collectif citoyen ou association dans votre collectivité militant pour l'Open data ? Si oui quel est son rôle dans la démarche open data ?
36. Coopérez-vous avec des opérateurs privés en charge d'un service public dans votre collectivité ? Quel type de contrat règle la propriété des données et le droit de diffusion ?
37. Quelle est votre relation en tant que projet communal avec Etalab, APIE? Selon vous quel rôle joue Etalab ?
38. Pourriez-vous décrire votre situation par rapport à d'autres projets open data en France ? y-a-t-il une concurrence entre les collectivités territoriales ?

III. Un retour sur l'usage des données publiées

39. Ciblez-vous une catégorie déterminées d'utilisateurs? Les startups ? Développeurs ? Grand public ?
40. Qui sont les réutilisateurs de vos données ?
41. Avez-vous des statistiques sur l'usage ? nombre de téléchargement ? les fichiers les plus utilisés ?
42. Avez-vous mené des études pour identifier les besoins des usagers?

43. Recevez-vous des demandes d'accès et de réutilisation de données spécifiques ? Comment vous répondez à ces demandes ?
44. Avez-vous exercé un travail de promotion de vos données ? (concours, Hackaton)
45. On est dans une logique d'offre ou de demande de données?
46. Développez-vous une stratégie d'interaction avec la population à travers le web 2.0. Avec quels outils ?

IV. Évaluation du projet

47. Êtes-vous satisfait de l'avancement du projet et quels sont vos attentes ?
48. Comment vous pouvez mesurer l'effet du service open data au niveau social, politique et économique?
49. Quels sont les points forts et les points faibles du projet ?
50. Quelles sont les conditions de réussite d'un projet open data ?
51. Quels sont les freins identifiés et comment les surmonter?
52. Quels sont vos conseils aux autres chefs de projet Open data dans d'autres collectivités ?

Annexe C : Guide d'entretien avec Moïse Mayo, Directeur Régional de l'Insee – Bourgogne

1. Quel était le rôle du comité dans la démarche d'ouverture ?
2. Est-ce que tous les acteurs au niveau local ont la même vision de l'open data ? avez-vous repéré des conflits entre eux ?
3. Les collectivités ne sont pas les seuls producteurs de données sur leurs territoires, quelle est la position et le poids des données produites par l'Insee par rapport à la production de données locales ?
4. Quel est l'effet du mouvement open data sur l'Insee ?
5. Comment expliquez-vous le référencement des données de l'Insee sur les portails open data des collectivités ? est-ce qu'ils ont subi un traitement ?
6. Pensez-vous que les collectivités sont en situation *de* sous-exploitation de leurs données ?
7. Pensez-vous qu'on est dans une logique d'offre ou de demande de données ?

Annexe D : Guide d'entretien avec Cyril Longin, Chef du projet Open data et directeur des archives publiques de Saint Etienne

1. Quelles sont les lois et les tendances qui régissent l'ouverture et la réutilisation des archives publiques en France
2. Quel changement apporte le mouvement open data sur les modes de communication des archives publiques ?
3. Quel est l'effet du mouvement Open data sur la mission et le mode de fonctionnement des archives publiques ? Parle-t-on d'une mutation ?
4. Quels sont les challenges pour que archives publiques appliquent les principes de l'open data : accès en ligne complet, gratuit, sans inscription aux données en format réutilisable
5. Distinguez-vous entre les notions : accès et réutilisation ?
6. Pourquoi les données culturelles forment une exception à la réutilisation ?
7. Les collections des archives publiques sont-elles considérées comme des « données » à ouvrir ?
8. Pensez-vous que les collectivités sont en situation *de* sous-exploitation de leurs données culturelles ?

9. La redevance est-elle un frein à l'utilisation ou sorte de protection aux mauvais usages ?
10. Comment éviter la « privatisation » des données issues des archives publiques ? et comment assurer a win-win situation entre archives publiques et réutilisateurs?
11. Comment protéger les données personnelles?

Annexe E : Guide d'entretien avec Romain Mazon, Data journaliste, rédacteur en chef du site de la gazette des communes.

En tant qu'observateur-évaluateur

1. Dans une démarche volontaire des collectivités, est-ce qu'ils ont réussi à surmonter les freins et passer vers une culture d'ouverture ? Quels sont les facteurs clés qui contribuent et influent leurs engagement dans un projet Open data ? y-a-t-il une concurrence entre les collectivités territoriales ?
2. Comment vous pouvez mesurer l'effet du service open data au niveau social, politique et économique?
3. « L'open data remet en cause même la production, et la manière de faire un service publique » Quel changement apporte l'open data sur les pratiques des politiciens et des administrateurs ?
4. Selon vous, quelles sont les données qui apportent plus de transparence aux administrations ? sont-elles ouvertes ? Si non pourquoi ?
5. Est-ce que vous pouvez mesurer les attentes des citoyens ? On est dans la logique d'offre ou de demande ?

Data mining

6. Quelle relation existe entre le data journalisme et l'open data ?
7. Qu'est ce qu'une donnée, une donnée publique, une donnée ouverte ?
8. Quelles données vous cherchez ? (données alphanumériques ? (puissance des) chiffres ? des statistiques ? un fait ? une base de données ?)
9. Quelle est l'importance du volume des données ?

10. Quel est le degré d'implication des citoyens dans la production de données structurées ou non sur les collectivités (crowdsourcing)?
Questionnez-vous la fiabilité des données produites par les citoyens ?
peut-on les considérer comme données « publiques » ?
11. Questionnez-vous la qualité des données ouvertes par les collectivités ?
12. Ou peut-on trouver la donnée ? Qui sont vos sources traditionnelles d'informations sur les collectivités ? est-ce que l'open data a créé d'autres sources ou canaux de données ?
13. Existe-t-il des données publiques qui sont soumises à redevance ?
la redevance est-elle un frein à l'usage ?
14. Y-a-t-il des données qui sont jusqu'alors gardées secrètes ou inaccessibles ?
15. Que pensez-vous de l'hébergement mutualisé ?
16. Quelles sont vos attentes en tant que journaliste de données publiées par les collectivités ?
17. Quel est le format idéal de libération et d'exploitation ?
18. Quel est la licence idéale? quels sont les conséquences sur vos usages ?

Traitement et présentation de données

19. Quelle procédure de traitement subissent les données depuis la collecte à la publication ?
20. Quels sont les produits ou services que vous produisez à partir des données publiques ?

Annexe F : Guide d'entretien avec Cyril Mottier, développeur d'applications mobiles

1. On remarque une explosion d'applications (jeux, page statiques, applications avec des moteurs de recherche...) avec l'augmentation des usages des outils de nomades (tablette, Smartphone, quel est la place des applications mobiles urbaines ?
2. Quels sont les facteurs clés qui contribuent et influent votre engagement dans la création de telles applications ?
3. Quelles sont les difficultés rencontrées?
4. Quel est le profile des développeurs des applications urbaines?
5. Quel est l'intérêt de l'open data pour un développeur ?

6. Comment vous qualifiez les initiatives faites par les collectivités ?
7. Quelle est la modalité de collecte d'extraction des données ? Qui sont vos sources pour créer les applications urbaines ?
8. Quelle est la valeur ajoutée de vos applications ?
9. Quel est le retour des utilisateurs des applications ?

Annexe G : Guide d'entretien avec Didier Gorrion président et fondateur de l'entreprise Transpacity

1. Comment définir l'information territoriale ? Les notions d'informations et de données sont-ils prises comme synonymes ?
2. Pourquoi choisir l'échelle des collectivités territoriales ? (communes, départements, et régions, municipalités ??)
3. Votre société a été créée en 2006. Dans quel contexte vous l'avez créée et pour quel but ?
4. Quels sont les facteurs clés qui contribuent et influent la façon dont vous offrez vos services ? Existe-t-il des tendances sociales ou enjeux sociaux qui affectent vos services ?
5. Pouvez-vous décrire l'évolution de votre travail par rapport à vos relations établies avec les différents acteurs :
 - Les collectivités
 - Les entreprises ou opérateurs privés ou clients
 - Les habitants ou usagers
6. Qui sont vos sources d'informations publiques ?
7. Vous avez créé un partenariat étroit avec les collectivités comme étant les sources et producteurs de données publiques. Quel type de partenariat s'agit-il ?
8. Comment décrire votre relation avec les collectivités ?
9. D'après vous « Les collectivités territoriales connaissent une mutation profonde » Vous asseyez d' « orienter les collectivités dans leurs choix et d'être perçu comme une force de proposition! » Quel type de mutation vous en parlez et quel sera l'intérêt des collectivités de libérer ces données ? Comment faire face à la solide tradition de service public en opposition à la mise à disposition des données au secteur privé?

10. Selon quels critères choisissez-vous l'échantillon de panélistes ? (le groupe de décideurs territoriaux interrogés régulièrement sur leurs opinions ou leurs attitudes)
11. Quelle est votre méthodologie pour collecter les données?
12. Vous exercez un travail de veille. Quels types de données vous capter et pourquoi ? Quelle est la spécificité des données collectées ? en quel format ? des données brutes ?
13. Utilisez-vous les données publiées par l'Institut National de la Statistique et des Études Économiques qui a comme une mission de collecter, traiter, et de diffusion de l'information?
14. Avez-Vous eu recours aux données de la plateforme data.gouv ? data publica ? des applications ? que pensez-vous du travail élaboré par ETALAB ?
15. En principe, la loi du 17 juillet 1978 donne aux citoyens le droit d'accès aux documents administratifs et l'ordonnance n° 2005-650 du 6 juin 2005 permet l'usage commercial des données collectées dans le cadre légal de l'accès aux documents administratifs. Existe-t-il en pratique des difficultés à accéder et à capter les données publiques émanant des services publics?
16. Quel est l'intérêt des différents clients à exploiter les données territoriales?
17. Vous parlez des données publiques en tant que moyen pour « sécuriser les investissement commercial! » Expliquez.
18. Quelle est la finalité des données ou les différents usages possibles pour chaque client?
(dimension sociale pour le fonctionnement démocratique, ou dimension économique à forte valeur économique?)
19. Vos données traitées sont produites à des fins internes de l'organisme ou bien elles peuvent être médiatisées à destination du grand public? Si oui dans quels cas ?
20. Vous menez également des études auprès des habitants pour mesurer, à l'échelon local, leur perception d'un service, leurs attentes et leur degré d'adhésion aux projets communaux ou intercommunaux. Comment choisir votre échantillon de la population ? Quels sont les critères ?
21. Que pensez-vous du mouvement Open data qui se base sur l'ouverture des données et la réutilisation **gratuite** des données brutes émanant par tout service publique. vous proposez des services

payants, la logique de commercialisation des données publiques s'oppose-t-elle à ce mouvement ?

22. L'opérateur privé est censé rediffuser l'information collectée gratuitement à des fins commerciales ?
23. Y a-t-il un risque de concurrence entre les opérateurs publics et les opérateurs privés ?
24. Le mouvement open data encourage les entreprises à libérer leurs données pour permettre l'innovation et un nouveau dynamisme économique. Que pensez-vous de ce modèle économique ?
25. Pourriez-vous décrire le travail d'indexation, d'analyse et de structuration des données collectées que vous pratiquez ?
26. A votre avis, quelle est la valeur ajoutée des données rediffusées ?
27. Que désignez-vous par « information diffusée à usage immédiat » et « données qualifiées et dynamiques » ?
28. Quelle est la place des nouvelles technologies d'information et de communication dans votre travail ?
29. Quels types de logiciel utilisez-vous ? Logiciels de traitement et d'analyse de données ? Logiciels cartographiques ?
30. En quel format les données traitées sont représentées aux clients ? données sous forme de cartes, histogrammes, courbes et tableaux ?

Annexe H : Guide d'entretien avec Florian Farge, contributeur actif dans OpenStreetMap

1. Pourquoi une base de données cartographiques libre ?
2. Quelle est la différence entre OSM et d'autres producteurs de données géographiques ?
3. Quelle est la valeur ajoutée d'OSM ?
4. Pourquoi la licence OdbL a remplacé la licence CC-BY-SA ?
5. Qui sont les contributeurs d'OSM ? Que pensez-vous du concept de crowdsourcing ?
6. Qui sont les usagers d'OSM ?
7. Quelles sont les limites du libre ?
8. Qui va vérifier la légitimité et la validité des données ?
9. Avez-vous monté un partenariat avec les collectivités ?

Annexe I : Guide d'entretien avec Sami Rabih, Membre du collectif OpenData69

1. Quand le collectif a été créé ? Dans quel contexte vous l'avez créée et pour quel but ?
2. Quel est le rôle des organisations et fondations qui militent en faveur de l'ouverture des données ?
3. Quels sont les facteurs clés qui contribuent et influent le mouvement Open data? Existe-t-il des tendances sociales ou enjeux sociaux qui favorisent l'émergence de ce mouvement?
4. Pouvez-vous décrire vos relations établies avec les différents acteurs :
 - Les collectivités
 - Les entreprises ou opérateurs privés ou clients
 - Les habitants ou usagers
5. Pouvez-vous décrire la situation ou donner un état des lieux de l'open data de la région Rhône alpes ?
6. Quels sont les composants et les conditions de réussite d'une ouverture de données ?
7. Quel est le rôle des citoyens dans la génération ou collecte de données ?
8. Quelles données ouvrir? Y-a-t-il des domaines prioritaires ?
9. En quel format les données traitées doivent être représentées?
10. Que pensez-vous du travail élaboré par ETALAB et de la politique nationale de l'open data ?
11. Existe-t-il des barrières à l'exploitation des données publiques?
12. Quelle est la finalité des données ou les différents usages possibles?
13. Comment vous pouvez mesurer la perception d'un service informationnel, les attentes et le degré d'implication des citoyens ?
14. Que pensez-vous de la commercialisation des données ?