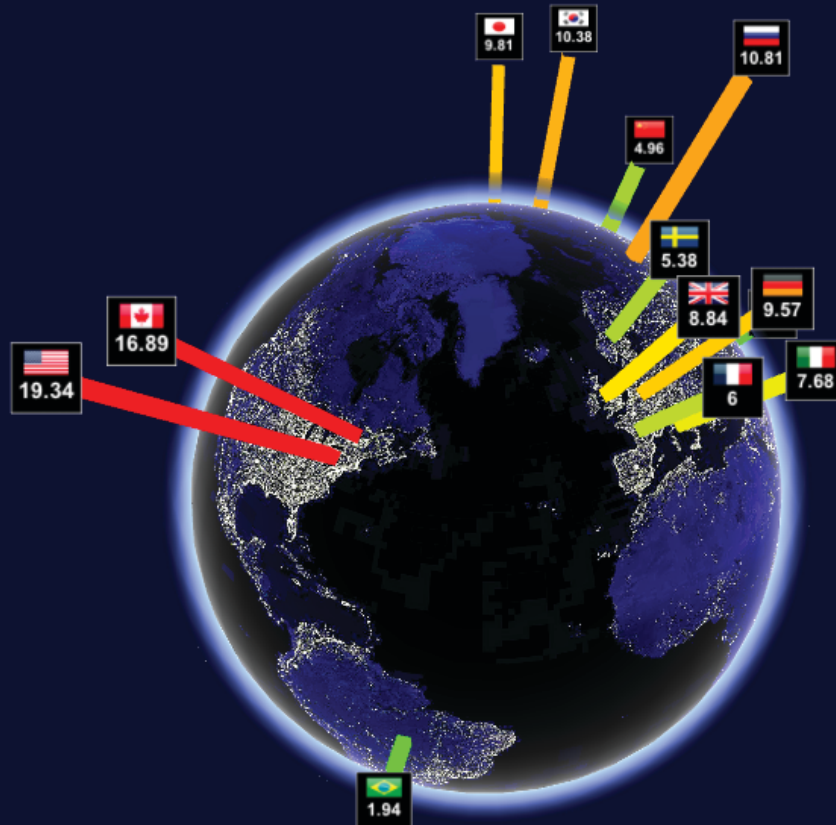


Romain Lacombe
Pierre-Henri Bertin
François Vauglin
Alice Vieillefosse

Ecole des Ponts ParisTech

Rapport remis le 13 juillet 2011 au Ministre de l'Industrie,
de l'Énergie et de l'Économie numérique



Pour une politique ambitieuse des données publiques

Juillet
2011

Les données publiques
au service de l'innovation
et de la transparence



École des Ponts ParisTech. 2011. *Pour une politique ambitieuse des données publiques : Les données publiques au service de l'innovation et de la transparence.* Rapport à la Délégation aux usages de l'Internet, remis au Ministre de l'Industrie, de l'Energie et de l'Economie numérique le 13 juillet 2011.



Proxima
mobile

Ecole des Ponts ParisTech

Rapport à la Délégation aux Usages de l'Internet
dans le cadre du portail Proxima Mobile
de services aux citoyens sur terminaux mobiles

Remis au Ministre de l'Industrie, de l'Energie
et de l'Economie numérique
le 13 juillet 2011

Pour une politique ambitieuse des données publiques

*Les données publiques au service
de l'innovation et de la transparence*

Juillet 2011



École des Ponts ParisTech. 2011. *Pour une politique ambitieuse des données publiques : Les données publiques au service de l'innovation et de la transparence.* Rapport à la Délégation aux usages de l'Internet, remis au Ministre de l'Industrie, de l'Energie et de l'Economie numérique le 13 juillet 2011.

Vous pouvez contacter les auteurs de ce rapport sur
<http://rapport-opendata-2011.fr>

À propos de ce rapport

L'ouverture et la réutilisation des données publiques sont des enjeux majeurs de l'action publique à l'ère numérique. À la croisée des usages innovants d'Internet et des modes de gouvernance collaborative les plus récents, ils ouvrent le champ à des domaines nouveaux de l'action de l'État.

Ce rapport présente une étude sur la réutilisation des données publiques, menée pour la Délégation aux usages de l'Internet du Ministère de l'Enseignement supérieur et de la Recherche dans le cadre du Master d'Action Publique de l'École des Ponts ParisTech.

Il met en perspective la problématique et les enjeux de l'*Open Data*, propose un état des lieux de la réutilisation des données publiques en France, et dessine trois scénarios prospectifs pour l'évolution future de ce mouvement. Elle présente seize propositions pour une politique nationale ambitieuse d'ouverture et de réutilisation des données publiques.

Quatre élèves de l'École des Ponts ParisTech, Pierre-Henri Bertin, Romain Lacombe, François Vauglin et Alice Vieillefosse ont mené cette analyse de septembre 2010 à janvier 2011, en rencontrant les acteurs clés de la réutilisation des données publiques, en prenant part à des colloques internationaux, et en s'appuyant sur la bibliographie existante.

Deux personnalités académiques expertes de l'économie numérique et de ses nouveaux modes de gouvernance, MM. Pierre-Jean Benghozi et Laurent Gille, ont accepté d'encadrer le travail de ce groupe.

M. Benghozi est Directeur de recherche CNRS, Directeur du Pôle de Recherche en Économie et Gestion et professeur d'économie à l'École Polytechnique (titulaire de la Chaire Orange Innovation et Régulation des Services Numériques), et responsable du pôle Innovation du Master Innovation et Régulation de l'Économie Numérique de Télécom ParisTech. M. Laurent Gille, enseignant-chercheur au Département de Sciences économiques et sociales de Télécom ParisTech, est responsable du parcours Industries de Réseaux du Master Innovation et Régulation de l'Économie Numérique.

Les conclusions de l'étude ont été rédigées de novembre à janvier 2011, et présentées au public le 20 janvier 2011 au Ministère de la Recherche et de l'Enseignement supérieur.

Ce rapport a été finalisé par Romain Lacombe entre mars et début mai 2011. Il ne fait qu'évoquer les développements les plus récents, et présente une synthèse des évolutions qui ont eu lieu au printemps 2011 en France, notamment au niveau national.

Les recommandations et propositions exprimées ne reflètent que les conclusions de l'analyse menée par les membres du groupe d'étude, et n'engagent en aucune manière leurs institutions, employeurs ou corps respectifs. Toutes les erreurs ou omissions qui subsisteraient dans ce document ne sont le fait que de ses auteurs.

Remerciements

Nous tenons, très modestement, à témoigner toute notre gratitude envers Monsieur le Ministre de l'Industrie, de l'Energie et de l'Economie numérique, pour avoir bien voulu nous faire l'honneur de recevoir ce rapport sur l'ouverture et la réutilisation des données publiques, ainsi qu'à remercier Séverin Naudet, Directeur de la mission Etalab auprès du Premier ministre, pour avoir accepté d'en présenter la synthèse, le 13 juillet 2011, à l'occasion du lancement du concours européen « Open Data Proxima Mobile ».

Nous remercions M. Bernard Benhamou, Délégué aux Usages de l'Internet, pour sa disponibilité, la confiance qu'il nous a témoignée, et ses conseils avisés et constructifs lors de la rédaction de ce rapport. Nous tenons aussi à remercier pour leur aide l'ensemble des équipes de la Délégation.

Nous remercions également M. Pierre-Jean Benghozi, Directeur de recherche au CNRS et professeur d'économie à l'École Polytechnique, et M. Laurent Gille, enseignant

chercheur en sciences économiques et sociales à Télécom ParisTech. Leurs conseils pertinents et le temps qu'ils nous ont consacré, tout au long de notre étude, nous ont été très précieux.

Nous souhaitons également remercier l'équipe du Mastère d'Action Publique de l'École des Ponts ParisTech pour l'opportunité qu'elle nous a accordée : le directeur Dominique Schwartz, le directeur adjoint Vincent Spenlehauer et la secrétaire Sylvie Berte.

Nos remerciements s'adressent enfin à l'ensemble des interlocuteurs qui nous ont reçus ou nous ont fait part de leurs réflexions sur le sujet des données publiques, ainsi qu'à toutes celles qui ont répondu à nos nombreuses questions.

Merci de nous avoir donné la chance de contribuer, modestement, à la réflexion sur un sujet fondamental pour le futur de l'économie numérique et l'avenir de notre démocratie.

Résumé

Les évolutions technologiques récentes ont placé le sujet de la réutilisation des données au cœur de l'action publique. L'administration collecte et produit un champ très vaste de données, depuis les budgets des institutions publiques jusqu'aux mesures de la qualité de l'air que nous respirons, en passant par des statistiques précises sur notre économie, notre société et nos territoires. Ouvrir l'accès le plus large aux informations publiques, c'est tout d'abord améliorer la transparence de l'État et de ses institutions, éclairer le débat démocratique, et rapprocher la puissance publique des citoyens. Favoriser l'ouverture et la réutilisation de ces données (« l'Open Data »), c'est permettre à la communauté des développeurs et des entrepreneurs d'inventer de nouveaux usages et de nouvelles applications utiles aux citoyens. C'est donc encourager l'innovation, donner naissance à la société de l'information, et contribuer au développement des secteurs stratégiques d'avenir pour notre pays, de l'économie numérique au développement durable. L'État doit s'approprier pleinement cette opportunité majeure. Il lui faut prendre le virage de cette nouvelle dimension des politiques publiques, pour la mettre au service de l'ensemble des acteurs du pays.

Mots-clés : Open Data, données publiques, information publique, accès aux documents administratifs, politique publique, gouvernance ouverte, transparence, démocratie, politique industrielle, politique des technologies, innovation, recherche, entrepreneuriat, compétitivité, croissance, emploi, économie numérique, Web sémantique, Internet.

Abstract

Recent evolutions of technology have placed public data at the central focus of public policy. Government collects and produces a vast array of data, from the budgets of our institutions to the quality of the air we breathe, through precise statistics on our economy, our society or our territories. Opening public sector information means improving the transparency of our Government and its institutions, enlightening the democratic debate, and strengthening the bond between the nation and its representatives. Encouraging reuse of this data (« Open Data ») allows the community of Web developers and Internet entrepreneurs to invent new, useful applications for citizens. It is an important public policy lever in favour of innovation, which can ignite the information society and contribute to the development of strategic sectors of the economy, from sustainable development to the growth of the digital economy. Government must fully embrace this major opportunity and this new dimension of public policy, and put its power to work at the service of the nation's every constituent.

Keywords : Open Data, public data, public sector information, freedom of information, public policy, Open Government, transparency, democracy, industrial policy, technology policy, innovation, R&D, entrepreneurship, competitiveness, growth, jobs creation, digital economy, semantic Web, Internet.

Résumé analytique

Les évolutions technologiques récentes ont placé le sujet de la réutilisation des données au cœur de l'action publique.

L'administration collecte et produit un champ très vaste de données, depuis les budgets des institutions publiques jusqu'aux mesures de la qualité de l'air que nous respirons, en passant par des statistiques précises sur notre économie, notre société et nos territoires.

Ouvrir l'accès le plus large aux informations publiques, c'est tout d'abord améliorer la transparence de l'État et de ses institutions, éclairer le débat démocratique, et rapprocher la puissance publique des citoyens.

Favoriser l'ouverture et la réutilisation de ces données (« *Open Data* »), c'est permettre à la communauté des développeurs et des entrepreneurs d'inventer de nouveaux usages. C'est donc encourager l'innovation, et contribuer au développement des secteurs stratégiques d'avenir, de l'économie numérique au développement durable.

L'État doit s'approprier pleinement cette opportunité majeure. Il lui faut prendre le virage de cette nouvelle dimension des politiques publiques, pour la mettre au service de l'ensemble des acteurs du pays.

Une problématique au croisement des évolutions technologiques, politiques et économiques récentes.

L'ouverture et la réutilisation des données publiques constituent désormais une attente évidente de la société civile. Cette problématique a émergé récemment au croisement de plusieurs évolutions profondes.

Une prise de conscience de la valeur économique de la réutilisation des données : chaque année dans l'Union Européenne, la réutilisation de données issues du secteur public générerait un chiffre d'affaires de plus de 27 milliards d'euros. Pour Neelie Kroes, vice-présidente de la Commission européenne chargée de la stratégie numérique, « une grande partie de ces informations, en Europe, sont sous-exploitées. Nous devons examiner l'opportunité de modifier les règles de l'Union en matière de réutilisation, pour libérer pleinement leur potentiel économique ».

Une évolution du cadre législatif : l'Union Européenne a adopté deux directives en la matière, la directive « Informations du secteur public » (PSI) en 2003 et la directive INSPIRE en 2008. Une consultation publique récente sur la directive PSI a notamment porté sur l'opportunité de réviser son champ d'application, les redevances et licences relatives aux données, ainsi que les formats électroniques utilisés.

L'émergence de la notion politique de « Gouvernement Ouvert » : les politiques de diffusion des données publiques ont fortement évolué dans les pays anglo-saxons du fait de l'émergence de la notion d'*Open Government*, thème de campagne du Président Barack Obama. Les États-Unis et la Grande Bretagne ont créé des portails de données publiques réutilisables, *data.gov* et *data.gov.uk*. En France, la mission Etalab vient d'être créée, sous l'autorité du Premier ministre François Fillon, pour coordonner les administrations en matière de données publiques et mettre en place la future plateforme *data.gouv.fr*.

Des initiatives citoyennes en faveur de l'*Open Data* : une série d'initiatives récentes, issues de la société civile et des communautés de développeurs, s'efforcent de promouvoir les concepts de l'*Open Data*, avec le soutien d'associations et de fondations comme la Sunlight Foundation aux États Unis, l'Open Knowledge Foundation (OKFN) en Grande Bretagne, et la FING, LiberTIC et Regards Citoyens en France.

La double rupture du Web 2.0 et des technologies mobiles : les innovations de service rendues possibles par les technologies 2.0 (interfaces dynamiques, cartographie en ligne, ouverture des API) ont amené à la généralisation du *remix* et du phénomène des *mashups* de données. La nouvelle génération d'applications géolocalisées sur *smartphones* repose pour une part importante sur des données géographiques. Ces ruptures technologiques ont révélé le besoin croissant de pouvoir réutiliser des données publiques.

Toutes ces évolutions récentes confèrent une actualité nouvelle à la problématique de la réutilisation des données publiques.

Une opportunité majeure pour l'État et l'administration

L'ouverture des données publiques permet à des entrepreneurs, des chercheurs, des développeurs d'accéder aux données collectées dans le cadre du fonctionnement des administrations.

En les réutilisant, ils inventent de nouveaux usages et peuvent créer de nouvelles applications utiles aux citoyens. Cette coproduction de services au public peut constituer la base d'une nouvelle relation numérique de l'administration à l'utilisateur, et d'un fonctionnement plus efficace de la puissance publique.

L'ouverture des données publiques est de plus 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 fonder les choix politiques sur des analyses éclairées et ancrées dans le réel.

Enfin, un accès plus large aux données publiques peut contribuer à renforcer la confiance des consommateurs, et donc la stabilité de notre économie.

Un outil stratégique pour notre économie

Des conditions de réutilisation plus libres encourageraient l'innovation et pourraient contribuer au développement de l'économie numérique : les nouveaux services et applications mobiles, qui en bénéficieraient particulièrement, connaissent actuellement une très forte croissance. L'économie numérique est un facteur clé de croissance économique et de création d'emplois en France.

L'accès aux données publiques offre de plus les prémisses d'une « infrastructure informationnelle », qui pourrait faciliter le développement durable en permettant aux consommateurs de faire des choix mieux informés.

Enfin, la réutilisation des données publiques constitue un élément de soutien aux objectifs stratégiques d'investissement d'avenir pour l'économie française, de la société numérique aux villes de demain, aux transports intelligents, en passant par l'encouragement à la recherche scientifique et à l'innovation.

Une mise en œuvre complexe qui pourrait être simplifiée

Le cadre juridique est établi pour la majorité des données publiques, mais les types de licences et de contrats se multiplient. Le statut des données issues d'organismes parapublics reste à préciser. L'interopérabilité technique et juridique est cruciale : il convient de clarifier les formats et les licences acceptables dans une optique d'encouragement à l'ouverture et à la réutilisation.

La gratuité pour la réutilisation des données à titre commercial favorise l'innovation et les nouveaux usages. Nous préconisons donc de la généraliser le plus possible, et d'éviter à tout prix la multiplication et le maintien de redevances excessives, qui pourraient conduire à des incohérences et freiner la réutilisation.

Des risques surmontables

Esquissées dans la loi 1978 sur l'accès aux documents administratifs, consacrées par deux directives européennes, l'ouverture et la réutilisation des données publiques s'inscrivent dans une évolution durable de nos sociétés.

Elles induisent une transformation de la culture et du mode de fonctionnement des administrations. Des freins au changement subsistent, et des risques peuvent émerger. Rien n'indique cependant qu'ils ne puissent être surmontés, surtout si l'État s'engage dans une collaboration rapprochée avec la communauté des réutilisateurs.

Trois axes stratégiques pour favoriser la réutilisation des données publiques

Nous proposons les trois axes stratégiques suivants pour favoriser la réutilisation des données publiques et permettre à la société d'en retirer l'ensemble des bénéfices attendus.

Simplifier pour accélérer : réduire les efforts nécessaires d'une part à la diffusion des données par les acteurs publics, et d'autre part à leur réutilisation, afin d'accélérer le développement de services utiles à partir des données publiques.

Faciliter l'expérimentation : adapter les conditions de réutilisation des données publiques pour faciliter l'expérimentation par les citoyens, les associations et la communauté des développeurs et des entrepreneurs.

Favoriser l'émergence d'un écosystème : favoriser l'émergence d'un écosystème de producteurs et de réutilisateurs de données publiques en France, autour de la mission Etalab et du futur portail data.gouv.fr.

Seize propositions pour une politique nationale ambitieuse des données publiques

Ce rapport formule seize propositions, articulées selon les trois axes stratégiques précédents. Elles sont présentées de manière détaillée au chapitre 8, et visent à proposer une feuille de route claire pour mettre en place une politique ambitieuse des données publiques.

1. Formaliser les bonnes pratiques de diffusion des données publiques attendues par la communauté des réutilisateurs :

- Expliciter l'effort minimal attendu des différents acteurs publics.
- Assurer la formation des agents et leur accès à des guides clairs et explicatifs.
- Simplifier les formats de diffusion.

2. Mettre en cohérence les conditions légales de réutilisation des données issues de sources différentes :

- Établir une « licence données ouvertes » pour réutilisation gratuite qui s'appliquerait le plus largement possible.
- Encourager un marquage graphique des droits de réutilisation.

3. Encourager l'utilisation de formats facilement réexploitables, qui respectent les normes d'interopérabilité et de « lisibilité machine » :

- Privilégier les formats ouverts respectant les normes d'interopérabilité.
- Communiquer auprès des agents sur l'importance des formats réexploitables.
- Maintenir une liste de formats correspondants à ces critères.

4. Faire en sorte que chaque contrat public clarifie à l'avenir le financement de l'ouverture et les droits de réutilisation liés aux données produites dans son cadre :

- Clarifier le financement de l'ouverture et les droits de réutilisation de données dans les contrats publics.
- Établir des bonnes pratiques et des clauses types.
- Élargir le périmètre des données réutilisables aux ressources essentielles.

5. Faciliter l'enrichissement des données et l'intégration de corrections éventuelles par la communauté :

- Faciliter l'intégration des métadonnées.
- Permettre l'enrichissement collectif des jeux de données.

6. Impliquer des acteurs publics pilotes dans les projets de recherche sur le Web sémantique.

7. Rendre la réutilisation des données gratuite pour toute réutilisation à titre personnel.

8. Encadrer strictement les cas dans lesquels une redevance est légitime, et charger les producteurs de données d'établir l'absolue nécessité de toute tarification :

- Dépasser une valorisation purement commerciale des données publiques et encourager leur réutilisation la plus large.
- Généraliser la réutilisation libre et gratuite la plus large possible.
- Restreindre la tarification à coûts moyens aux marchés matures.

9. Conditionner toute redevance pour réutilisation de données à valeur ajoutée à la mise à disposition gratuite des données brutes qui les sous-tendent.

10. Explorer des conditions alternatives de réutilisation pour laisser carte blanche à l'expérimentation :

- Privilégier les partenariats avec les nouveaux acteurs de l'innovation.
- Explorer les licences gratuites avec partage des conditions initiales à l'identique (*share alike*).
- Explorer la rémunération par l'accès aux données d'usage.
- Explorer la rémunération par le service client.
- Faciliter l'expérimentation avec les Web Services.

11. Impliquer la communauté des réutilisateurs dans les initiatives de mise à disposition de données publiques :

- Créer un point d'accès unique centralisé et mis à jour automatiquement.
- Développer une offre grand public qui mette en valeur des applications utiles.
- Investir dans l'accompagnement de la communauté.
- Démultiplier le service aux utilisateurs.
- Organiser des séminaires ou ateliers de travail ponctuels sur les sujets techniques d'intérêt pour la communauté.

12. Mettre en place des assises des données publiques et animer la communauté des réutilisateurs :

- Regrouper l'ensemble de l'écosystème pour des assises des données publiques.
- Animer la communauté des réutilisateurs.

13. Encourager le développement d'applications à partir des données publiques :

- Organiser un concours national de développement d'applications.
- Prendre en compte la réutilisation des données publiques dans les appels à projets de R&D numérique.
- Pérenniser l'offre d'applications créées à partir des données publiques.

14. Intégrer au débat démocratique les données publiques et leur réutilisation :

- Communiquer aux administrations l'importance des données pour mettre en lumière leurs métiers.

- Illustrer la réutilisation des données publiques par le développement de services aux citoyens.
- Faire des données et de leur réutilisation un outil d'éclairage du débat public.

15. Développer une stratégie nationale à partir du rôle de coordination des institutions publiques de la mission Etalab :

- Mutualiser des solutions de mise en ligne pour les administrations et les collectivités.
- Développer une stratégie nationale pour les données publiques.

16. Élargir le champ de l'Open Data :

- Élargir le champ de la libre réutilisation des données publiques.
- Explorer l'application de l'Open Data aux entreprises privées.
- Développer une stratégie européenne et internationale.

Table des matières

Résumé analytique	7
Table des matières	11
Introduction	13
Partie 1. « Open Data » : Contexte et enjeux d’une question sociétale	17
Chapitre I. Données publiques : émergence d’une question sociétale	19
1. La donnée au cœur des nouveaux usages	19
2. Les données publiques et leur réutilisation	21
Chapitre II. La libre réutilisation des données publiques à l’international	25
1. Le mouvement Open Data	25
2. États-Unis : initiatives locales, puis démarche fédérale	26
3. Grande Bretagne : une impulsion transpartisane du gouvernement	28
4. Situations contrastées en Europe	30
Chapitre III. Les enjeux de la réutilisation des données publiques	33
1. Démocratie, transparence et débat public	33
2. Modernisation de l’État	34
3. Innovation et soutien à l’économie numérique	35
4. Des données stratégiques pour l’économie	38
5. Un nouveau mode d’action publique	42
Partie 2. État des lieux de la réutilisation des données publiques en France	47
Chapitre IV. Le cadre de la réutilisation des données publiques en France	49
1. Le cadre juridique pose en principe la liberté de réutilisation	49
2. Les acteurs de la réutilisation de données publiques	53

Chapitre V. Premières expériences d'ouverture des données publiques en France _____	57
1. La réutilisation des données dans les collectivités territoriales _____	57
2. L'ouverture des données au niveau national : trois études de cas _____	62
3. La mise en place d'une politique nationale des données publiques _____	65
Partie 3. Comment favoriser la réutilisation des données publiques ? _____	69
Chapitre VI. Abaisser les barrières à la réutilisation _____	71
1. Répondre aux contraintes techniques _____	71
2. Gérer la complexité juridique _____	72
3. Anticiper l'impact organisationnel _____	74
4. Établir des modèles économiques propices à l'innovation _____	75
Chapitre VII. Prospective : Trois scénarios pour l'avenir _____	83
1. « Inertie » : Développement de l'offre de données publiques à son rythme historique _____	83
2. « Capture » : Des monopoles informationnels se substituent à la puissance publique _____	85
3. « Symbiose » : Catalyser le développement d'un écosystème _____	87
Chapitre VIII. Seize propositions pour une politique ambitieuse des données publiques _____	91
1. Simplifier pour accélérer _____	91
2. Faciliter l'expérimentation pour encourager l'innovation _____	94
3. Soutenir l'émergence d'un écosystème _____	96
Bibliographie _____	101
Textes règlementaires _____	105
Interlocuteurs _____	107
Liens et références _____	109

Introduction

« Si l'on partage des données sur le Web, des données publiques, des données scientifiques, des données citoyennes, quelles qu'elles soient, d'autres que nous sauront en tirer des créations merveilleuses que nous n'aurions jamais imaginées. »

Tim Berners-Lee, Inventeur du Web

Hal Varian, Économiste en Chef chez Google, le précise : « Je répète sans cesse que la profession sexy dans les dix prochaines années, ce sera statisticien. Et ce n'est pas de l'humour.¹ »

On peut émettre des réserves quant au pouvoir de séduction d'un tableur informatique ; il reste que l'attrait de la science des données s'est considérablement accru ces dernières années. Ses algorithmes ont gagné en puissance et ses applications jouent un rôle central dans la société de l'information.

À mesure que nos outils informatiques deviennent plus mobiles, leurs applications gagnent en importance dans nos vies quotidiennes. En recoupant des données numériques diverses, en les adaptant à notre usage à travers une ergonomie simple, les *smartphones* et les tablettes tactiles simplifient bon nombre des tâches de la vie quotidienne et professionnelle.

La réutilisation des données publiques

Un mouvement de réutilisation des données publiques² a vu le jour depuis plusieurs années. Concrètement, il s'agit pour des développeurs informatiques ou tout autre citoyen d'accéder à des données produites initialement dans le cadre du fonctionnement de l'État, et de leur

inventer de nouveaux usages. Tous types d'objets numériques sont concernés : des tableaux chiffrés, des mesures numériques, des bases de données, des cartes ou encore des jeux de données géolocalisées.

Les applications en question vont des services les plus simples aux plus complexes, de l'expérimentation au produit professionnel : des étudiants développent une application mobile, des citoyens quantifient l'action de leurs députés, des journalistes créent de nouvelles infographies plus instructives, une entreprise calcule des itinéraires de transports publics au service des internautes.

Les développements technologiques récents ont largement accéléré le développement de ce phénomène, et les attentes d'une partie du public sont aujourd'hui très fortes.

Une problématique nouvelle et complexe pour le secteur public

La réutilisation des données publiques offre des possibilités nouvelles à la sphère publique. Elle appelle une évolution de la culture et du mode de fonctionnement des administrations. Elle soulève surtout de nombreuses questions.

Quelles opportunités ? Quels risques ? Quelles pistes pour en faire profiter au mieux la société ? Quels enseignements tirer de l'expérience d'autres pays en la matière ? Quel(s) public(s) cette politique vise-t-elle, avec quels objectifs et quels moyens ? Quelles limites et quelles garanties donner au droit à la réutilisation ?

¹ Steve Lohr, 5 août 2009. « For Today's Graduate, One Word: Statistics ». The New York Times.
<http://www.nytimes.com/2009/08/06/technology/06stats.html>

² À l'exclusion bien sûr des données personnelles et des données confidentielles (loi de 1978 sur la liberté d'accès aux documents administratifs, modifiée par Ordonnance n°2005-650 du 6 juin 2005).

Une opportunité importante pour le développement du numérique

Grâce aux réseaux mobiles à haut débit, la connectivité devient ubiquitaire. Les nouvelles plateformes mobiles permettent désormais d'accéder en direct à des services ou des informations dans le contexte dans lequel ceux-ci sont le plus utiles. Ces évolutions ont donné naissance à une industrie nouvelle et en forte croissance, celle des applications mobiles.

Des acteurs internationaux du Web ont su prendre des positions fortes sur ce marché. La France possède des atouts pour réussir dans ce secteur. Faciliter l'accès aux données publiques en vue d'une réutilisation commerciale pourrait favoriser la création d'activités économiques novatrices.

Un débat passionné

Opportunité de développement économique et avancée en matière de transparence d'une part, érosion potentielle de la capacité d'action de l'État et de ses ressources propres de l'autre : la tarification de la réutilisation des données publiques est l'objet d'âpres débats, quand elle ne donne pas lieu à des contentieux.

L'ouverture des données elle-même peut faire peur, quand ce n'est pas l'incompréhension quant à l'intérêt d'un droit à la réutilisation de données qui interpelle.

Un mouvement engagé sur le long terme

La mise à disposition des données publiques pour leur réutilisation n'est pourtant pas un effet de mode : il s'agit d'un droit opposable consacré par la loi, basé sur une Directive de 2003 que la Commission Européenne envisage d'étendre et de renforcer.

Cette évolution engagée confronte chaque administration centrale, chaque collectivité locale, chaque établissement public à des questions difficiles. Qu'attend-on d'eux exactement, et comment bien faire ?

Contexte et enjeux d'un sujet de société

Le premier chapitre s'attelle à esquisser un portrait synthétique des grandes évolutions qui ont amené la question des données publiques à devenir, en l'espace de quelques années, un véritable sujet de société. Nous passons notamment en revue les évolutions technologiques récentes qui l'ont portée sur le devant de la scène (*chapitre 1*).

Le second chapitre dresse un état des lieux de la réutilisation des données publiques en

2011 dans le monde, et un historique des évolutions récentes dans les pays occidentaux (*chapitre 2*).

Nous nous attachons ensuite à cerner les bénéfices que l'on peut attendre d'une politique de soutien à la réutilisation des données publiques. Nous évoquons les enjeux liés à la participation démocratique et à l'efficacité de l'action publique. Nous explicitons les différents mécanismes à travers lesquels la donnée publique peut contribuer au développement économique (*chapitre 3*).

La réutilisation des données publiques en France

Une deuxième partie fait ensuite le point sur la réutilisation des données publiques en France aujourd'hui (*chapitre 4*).

Le chapitre suivant vise à fournir aux agents publics des pistes et références à partir desquelles construire leur réflexion. Nous cherchons en particulier à comprendre le contexte de cette politique, son cadre juridique et technique, et les acteurs qu'elle met en jeu (*chapitre 5*).

Comment favoriser la réutilisation des données publiques ?

Le mouvement de réutilisation de données étant durablement amorcé, il convient de s'interroger sur les difficultés auxquelles il risque de se heurter dans l'administration, les établissements publics et les collectivités : difficultés d'ordre juridique, organisationnel, technique et financier. Nous abordons la question de la tarification optimale, et des modèles économiques qui doivent s'appliquer aux données publiques (*chapitre 6*).

À partir des dynamiques en présence, nous avons imaginé trois scénarios futurs pour la réutilisation des données publiques : l'Inertie, la Capture et la Symbiose. Nous détaillons les caractéristiques et les implications de chacun de ces futurs possibles (*chapitre 7*).

Conclusions et recommandations

Le scénario de la Symbiose et du développement d'un écosystème présente de loin le plus d'avantages : travailler à l'ouverture des données publiques en collaboration avec la communauté de ceux qui les réutilisent est une nécessité pragmatique plus qu'un choix idéologique.

Pour que la société puisse tirer l'ensemble des bénéfices potentiels de la réutilisation des données publiques, il convient cependant

d'encourager son développement au rythme le plus rapide.

Nous proposons enfin (*chapitre 8*) seize propositions groupées en trois axes stratégiques, visant à établir une feuille de route claire vers une politique ambitieuse des données publiques.

Simplifier pour accélérer : réduire les efforts nécessaires d'une part à la diffusion des données par les acteurs publics, et d'autre part à leur réutilisation, afin d'accélérer le

développement de services utiles à partir des données publiques.

Faciliter l'expérimentation : adapter les conditions de réutilisation des données publiques pour faciliter l'expérimentation par les citoyens, les associations et la communauté des développeurs et des entrepreneurs.

Favoriser l'émergence d'un écosystème : favoriser l'émergence d'un écosystème de producteurs et de réutilisateurs de données publiques en France, autour de la mission Etalab et du futur portail data.gouv.fr.

Partie 1. « *Open Data* » :
Contexte et enjeux d'une question sociale

Chapitre I. Données publiques : émergence d'une question sociétale

Les évolutions technologiques récentes ont rendu possible le partage rapide et aisé de tous types de données. Elles réactivent, sous un jour nouveau, le débat déjà ancien sur l'accès aux documents administratifs, maintenant sous forme électronique et bien plus aisément diffusables.

L'Internet a fait apparaître une économie des données. Les stratégies de croissance des acteurs du Web, via l'ouverture de leurs interfaces de programmation (API), ont illustré le potentiel de la réutilisation des données, et suscité des attentes similaires des développeurs vis-à-vis du secteur public. Ils exercent une pression pour que les États, en facilitant l'accès à leurs informations, s'insèrent et tiennent leur rang dans ce « Web des Données ».

Ces ruptures technologiques ont accompagné par ailleurs une prise de conscience de l'importance des données produites par l'administration et le secteur public. La possibilité novatrice de réutiliser ces données sous forme brute, pour leur trouver de nouveaux usages utiles et instructifs, a abouti à l'émergence de la question des données publiques en tant que véritable question sociétale.

1. La donnée au cœur des nouveaux usages

Internet et les ruptures technologiques récentes ont fait apparaître une nouvelle « économie des données ».

(i) Le Web est structuré autour de l'échange de données.

Le génie informatique (*computer science* dans le monde anglo-saxon) est défini comme la manipulation algorithmique de données et de leurs structures de représentation. Le Web, avant d'être un phénomène social, est une architecture logicielle³ conçue par un chercheur britannique en physique des hautes énergies au CERN, Sir Tim Berners-Lee⁴. Elle

avait pour but de faciliter l'accès et le référencement de documents en symbolisant leur manipulation sous forme de phrases, liant verbes intuitifs (lire, afficher, remplacer, effacer) et métadonnées d'adressage universel (URL ou *Uniform Resource Locator*⁵).

Bien que basées sur un réseau d'échange de documents, les techniques du Web ont naturellement développé un tropisme de plus en plus marqué vers les données elles-mêmes. L'introduction des bases de données relationnelles dans les outils de développement a permis la création de sites dynamiques, c'est-à-dire dont les pages s'adaptent aux circonstances (évolutions de prix, listes d'objets changeants, « unes » de journaux électroniques mises à jour fréquemment). À partir de 2004, une technique émergente a vu le jour qui consiste à utiliser les fonctionnalités de *scripting*⁶ présentes dans les navigateurs pour permettre la mise à jour en temps réel de pages déjà chargées, en allant chercher en tâche de fond des données remises à jour. Nommée AJAX⁷ pour JavaScript Asynchrone et XML (*Asynchronous Javascript And XML*), à partir du langage de *scripting* JavaScript et du format d'échanges de données XML, cette technique à la croisée de différentes technologies Web est reconnaissable derrière des applications Web hautement réactives dont les pages ont rarement besoin d'être rafraîchies – telles que Gmail⁸, Quora⁹ ou Facebook¹⁰.

<http://www.w3.org/People/Berners-Lee/>

⁵ World Wide Web Consortium (W3C). 1994. *Uniform Resource Locators (URL) : A Syntax for the Expression of Access Information of Objects on the Network*.

<http://www.w3.org/Addressing/URL/url-spec.txt>

⁶ Capacité d'un navigateur à faire tourner dynamiquement un programme qui agit en temps réel sur la page affichée. Par exemple, une page qui affiche l'heure repose sur un script qui demande au navigateur de mettre à jour l'heure, les minutes et les secondes à mesure qu'elles s'écoulent. Dans le cas d'un service de courrier en ligne, c'est souvent un script qui réagit au click sur l'intitulé d'un email et change l'affichage pour détailler le contenu du mail plutôt que celui de la boîte de réception. C'est aussi un script qui vérifie automatiquement l'arrivée de nouveaux courriers en tâche de fond, et réintègre éventuellement ces nouvelles données dans le corps de la page affichée.

⁷ Garrett, J.J. 18 février 2005. « Ajax : A New Approach To Web Applications ».

<http://www.adaptivepath.com/ideas/essays/archives/000385.php>

⁸ Gmail, service de boîte de courrier électronique en ligne de Google, Inc. : <http://www.gmail.com>

³ World Wide Web Consortium (W3C), 2010, HTTP – Hypertext Transfer Protocol Overview – draft 09.

<http://www.w3.org/Protocols/>

⁴ World Wide Web Consortium (W3C). Sir Tim Berners-Lee: Bio. Téléchargée le 10 janvier 2011.

Pour supporter cette technologie, de plus en plus d'applications en ligne ont développé à la fois deux interfaces : une interface graphique classique accessible aux utilisateurs à travers leur navigateur, et une interface exclusivement consacrée aux machines, qui permet justement au moteur de *scripting* du navigateur de récupérer des données à la volée. C'est cette interface, qu'on peut se représenter comme un protocole d'échange de données de machine à machine, qu'on appelle interface de programmation ou API (sigle anglophone pour *Application Programming Interface*).

(ii) La donnée, monnaie d'échange et levier d'influence sur le Web.

Les API sont devenues un véritable outil stratégique pour les sociétés de services en ligne.¹¹ Par exemple, c'est dans une large mesure grâce à l'ouverture de leur API¹² que Twitter a pu permettre à une communauté de développeurs de se créer, et de multiplier les applications permettant d'utiliser le service en dehors du site Web principal (notamment dans les premiers jours de l'App Store d'Apple durant l'été 2008). Ce grand nombre d'applications a servi de « force de vente » virtuelle pour Twitter, et est un des facteurs principaux qui expliquent son succès populaire.

Bien documentées et basées sur des standards acceptés (tels que JSON ou XML) pour éviter les problèmes d'interopérabilité, ces interfaces s'adressent à des développeurs informaticiens. Selon le degré d'ouverture (fonction des choix stratégiques de la société) et les conditions de licences (souvent conçues pour encourager la réutilisation) consentis à ces derniers, ils peuvent extraire, éventuellement modifier, et transformer les données du service Web en question, pour les réutiliser, notamment à la construction de nouveaux services ou de nouvelles applications.

L'échange est mutuellement bénéfique : le nouveau service, consommateur de données,

gagne en utilité pour ses utilisateurs, et le système producteur voit sa position compétitive renforcée par la dépendance accrue du marché à ses services. En l'essence, c'est une transaction de marché où l'accès à des données est échangé contre leur diffusion à un nouveau public ou dans un nouveau contexte.

L'exposition d'une interface API est aujourd'hui une étape presque obligée du développement d'un nouveau produit en ligne. La généralisation de ce phénomène est d'une grande importance pour la question des données publiques, car elle a permis de prendre conscience de l'omniprésence des données derrière les services qu'utilisent quotidiennement les internautes.

En un mot, « l'économie des données » s'est construite autour de l'échange et de la fluidité sur le Web : les données ont souvent plus de valeur comme outil d'échange et d'influence qu'isolées au sein des entrepôts de données des entreprises¹³.

(iii) Avec le Web 2.0, la donnée devient collective

Un corollaire de la montée en puissance de la donnée est son rôle de liant social sur le Web. En particulier, avec l'apparition de phénomènes de collaboration de masse tels que Wikipedia, on s'aperçoit que l'échelle d'échanges rendue possible par Internet ouvre des possibilités insoupçonnées jusqu'alors.

Des services tels que Delicious (site de partage de bookmarks, qui agrège les catégories dans lesquelles les utilisateurs rangent différents liens, et permet ensuite de trouver des liens par catégorie d'intérêt) ou Flickr (qui permet de chercher des photographies libres de droits sur un sujet particulier, en se basant sur les *tags* de catégorisation produits par l'ensemble des utilisateurs) ont ainsi profité de ce phénomène de *folksonomy*, la taxonomie par les foules ou classification par l'intelligence collective¹⁴.

D'autres services tels qu'Amazon ou Netflix ont, eux, profité de la masse

⁹ Quora, Inc., site de questions et réponses dynamique caractérisé par l'expertise pointue de sa communauté d'utilisateurs, fondé en 2009 et qui a connu une très forte croissance en 2010 : <http://www.quora.com>

¹⁰ Facebook, Inc., le réseau social dominant qu'on ne présente plus, fondé par Mark Zuckerberg en 2004 et approchant 600 millions d'utilisateurs début 2011 : <http://www.facebook.com>

¹¹ Shah S., 28 August 2010, « Cannibalize Business Development by Popularizing your API » : <http://shaivalshah.com/cannabilize-business-development-by-populariz>

¹² Entretien le 12 novembre 2010 avec Henri Verdier, Président de Cap Digital.

¹³ Ed Parsons - Geospatial Technologist for Europe, Middle East and Africa chez Google, au séminaire « Innovations et développement du secteur de l'information géographique » organisé le 5 octobre 2009 à la Grande Arche de la Défense par le Conseil National de l'Information Géographique (CNIG), l'AFIGÉO et l'Institut des Sciences et des Techniques de l'Équipement pour le Développement (ISTED)

¹⁴ Tapscott, D. and Williams, A. 2006. *Wikinomics : How Mass Collaboration Changes Everything*.

d'utilisateurs se servant de leurs sites pour en extraire des modèles de préférences des consommateurs, qui leur permettent de faire des recommandations hautement personnalisées selon les goûts de leurs clients.

L'un des enseignements de la nouvelle incarnation sociale et collaborative du Web, c'est que la valeur des données n'est pas intrinsèque, mais dérive de leur agrégation, de leur recoupement, de leur analyse et de la réutilisation qui en est faite.

Le Web a ainsi mis à jour une économie des données, qui joue un rôle de monnaie d'échange parmi les entreprises d'Internet et les services en ligne.

(iv) Données géographiques et « mash-ups »

L'autre développement technologique qui a conduit à prendre conscience de l'importance des données, c'est la démocratisation de l'information géographique.

Google indique par exemple que son approche est « d'utiliser la géographie pour organiser toute l'information mondiale, qu'elle soit géographique ou non. C'est un principe d'organisation qui sera utilisé pour mieux comprendre l'information. »¹⁵.

Si les briques technologiques sont disponibles pour utiliser et manipuler la donnée géographique à loisir, ou même pour l'utiliser pour structurer d'autres données, le défi qui reste posé est celui de la constitution d'une véritable infrastructure des données géographiques. Ces données sont aujourd'hui accessibles en nombre et de qualité, mais les repérer et les réutiliser peut rester obscur et complexe¹⁶.

Conséquence récente des évolutions du Web, des applications cartographiques puissantes réservées jusque-là à des experts sont devenues accessibles à tous. Le monde du Web collaboratif a largement repris à son compte la possibilité de représenter des données sur des cartes. Grâce à l'ouverture gratuite (dans des limites relativement larges de volume) de son API, Google Maps est devenu une base de choix pour la création de « mashups », ces applications Web combinant

¹⁵ Ed Parsons – Ibid.

¹⁶ Voir l'intervention de Max Craglia, coordinateur technique d'Inspire pour la Commission Européenne (centre commun de recherche) au séminaire « Innovations et développement du secteur de l'information géographique » organisé le 5 octobre 2009 à la Grande Arche de la Défense par le Conseil National de l'Information Géographique (CNIG), l'AFIGÉO et l'Institut des Sciences et des Techniques de l'Équipement pour le Développement (ISTED).

« différentes sources au sein d'une seule plateforme, (...) l'une des facettes du design collaboratif sur le Web¹⁷ ». Cette nouvelle tendance a connu un tel succès au tournant des années 2006 à 2008 que le prestigieux Museum of Modern Art de New York y a consacré une exposition phare¹⁸.

Au-delà de l'incidence culturelle de ce phénomène¹⁹, la réappropriation du local et de l'espace géographique par le biais du Web et du mobile annonçait l'importance croissante de la donnée informatique géographique, et, au-delà, l'émergence du phénomène culturel du *remix* et de la réutilisation des données.

2. Les données publiques et leur réutilisation

La génération des *digital natives*, qui a grandi dans l'habitude de la réutilisation, du *remix* et de la recombinaison permanente, devait nécessairement s'intéresser aux informations collectées par l'État. L'émergence de la réutilisation des données publiques comme sujet de société, poussée par les nouveaux usages, a révélé une chaîne de valeur fondée sur ces données qui, lorsqu'elles sont mises à disposition de façon brute et réexploitable, permet aux développeurs et aux entrepreneurs de créer de nouveaux services.

(i) Les données numériques ont vocation à la diffusion la plus large

Contrairement aux documents papiers dont l'impression engendre des coûts variables conséquents et qui nécessitent des délais de traitement et d'acheminement, les données numériques ont un coût marginal de diffusion nul ou presque²⁰, et sont accessibles quasiment instantanément.

Le travail éventuel de mise à disposition de données numériques est donc un coût fixe : la diffusion la plus large, en distribuant ce coût, est donc le modèle le plus efficace d'un point de vue économique.

¹⁷ *Google Earth mashups 5 of 14*, in « Design and the Elastic Mind. » Museum of Modern Art. New York, NY. 2008.

http://www.moma.org/interactives/exhibitions/2008/elastic_mind/#/98/

¹⁸ « Design and the Elastic Mind. » Museum of Modern Art. New York, NY. 2008.

http://www.moma.org/interactives/exhibitions/2008/elastic_mind/

¹⁹ Lessig, L. 2009. *Remix : Making art and commerce thrive in the hybrid economy*. <http://remix.lessig.org/>

²⁰ Les volumes qui peuvent être atteints font que le coût de diffusion n'est en revanche pas négligeable et peut dans certains cas devenir important.

L'instantanéité de diffusion et l'ubiquité d'accès offrent de plus des perspectives tout à fait nouvelles en terme de périmètre de diffusion et d'audience potentielle, notamment au regard de la démocratisation de l'Internet.

(ii) Combiner les données brutes amplifie leur potentiel

Marqué par les traditions et les habitudes d'un monde de communication par le papier, l'État diffuse le plus souvent ses informations sous forme de « produits finis », tels que des analyses, des pages Web explicatives ou des rapports écrits – en quelque sorte des données « retravaillées » afin d'être intelligibles par tous. À l'inverse, les données brutes qui sous-tendent ces études sont rarement mises à disposition du public, car leur intérêt pour une réutilisation potentielle est encore très inégalement perçu par les producteurs de données publiques.

Aujourd'hui, les technologies de l'information permettent une mise à disposition des données à l'état brut. Loin d'être un obstacle à leur réutilisation ultérieurement à leur mise à disposition par l'administration, leur caractère brut permet au contraire de les recombinaison plus facilement, ce qui amplifie leur potentiel et permet la création de services impossibles à réaliser sans ces croisements.

Par exemple, des mesures locales de bruit permettent de mettre en perspective la corrélation positive mesurée entre les dépenses énergétiques d'un ménage et son taux d'équipement en doubles vitrages, paradoxale au premier abord. Dans le cadre d'un concours organisé par la ville de Washington, D.C., des données sur les vols, croisées avec les tracés des pistes cyclables, ont aussi permis de créer une carte des zones à éviter pour les cyclistes.

Parce qu'elle favorise ces croisements et ces combinaisons fertiles sur des sujets recoupant des champs extrêmement vastes de l'économie, de la société et de l'environnement, la mise à disposition des données brutes de l'État et des collectivités est un terreau fertile pour des réutilisateurs, qui peuvent inventer de nouveaux usages à ces données.

(iii) Les données publiques et leur réutilisation sont à l'origine d'une chaîne de valeur

L'État et le secteur public au sens large, dans le cadre de leurs opérations, touchent à l'ensemble des secteurs de notre économie, notre société et notre environnement. Les données qu'il collecte sont donc naturellement

riches et très variées ; certaines ont une forte valeur économique. Le marché de l'information est estimé en France à 3,7 milliards d'euros dont 60 % sont des données publiques²¹. Au niveau européen, ce chiffre représenterait près de 27 milliards d'euros²² par an.

La valorisation passe par différentes étapes nécessaires pour rendre les données publiques utilisables ou réutilisables. D'une part l'institution productrice doit définir précisément quelles données elle doit diffuser en ligne et de quelle manière, par un travail d'analyse et de collecte. S'imposent à ce stade le respect de critères de bon sens comme la protection de la vie privée, de la propriété intellectuelle des tiers, ou de la sécurité nationale.

D'autre part, pour faciliter la réutilisation, l'institution productrice doit s'assurer de fournir une donnée qualifiée et mise à jour. Pour cela, il semble important de créer des lots de données cohérents et de s'assurer de la présence d'une légende et plus généralement de la présence des métadonnées.

Il faut les mettre à disposition via un support adapté mais en réalisant un catalogage précis : il ne suffit pas de mettre à disposition des données, encore faut-il qu'elles puissent être découvertes et facilement accessibles grâce à un portail de données publiques ou un moteur de recherche.

Enfin, les données publiques numériques mises à disposition peuvent être réutilisées par la communauté des développeurs et des entrepreneurs, qui peuvent leur inventer de nouveaux usages. Ces données représentent aussi un outil interne au service des administrations, qui peut leur permettre des échanges d'information plus fluides et efficaces.

(iv) La visualisation de données joue un rôle structurant

Les données géographiques ont un rôle structurant dans la chaîne de valeur. Elles rendent possible la spatialisation et la visualisation sur carte d'autres jeux de données, notamment géolocalisées. Selon l'Association Française pour l'Information

²¹[http://www.rgpp.modernisation.gouv.fr/index.php?id=51&tx_ttnews\[tt_news\]=531&tx_ttnews\[backPid\]=11&cHash=89b9c6970f](http://www.rgpp.modernisation.gouv.fr/index.php?id=51&tx_ttnews[tt_news]=531&tx_ttnews[backPid]=11&cHash=89b9c6970f)

²² Dekkers, M. et al., 2006, « MEPSIR, Measuring Public Sector Information Resources. Final Report of Study on Exploitation of public sector information », http://ec.europa.eu/information_society/policy/psi/docs/pdf/s/mepsir/final_report.pdf

Géographique (AFIGÉO)²³, 80 % des informations qui sont utilisées possèdent une dimension géographique. Cela leur confère donc une importance particulière à la fois d'un point de vue technique, mais aussi en tant qu'outil de recoupement et de recombinaison de données.

Ainsi, dès 1854, John Snow détermina les causes de la propagation du choléra grâce à la visualisation de données sur une carte. À cette époque, on pensait que la maladie se propageait par inhalation. Le médecin superposa sur une carte les adresses des victimes du choléra à Exeter (quartier de Londres) durant un épisode aigu, et constata qu'elles coïncidaient avec l'accès à la pompe à eau de la Broad Street. Par déduction, il en conclut au rôle de l'eau et des conditions sanitaires dans la propagation de l'épidémie, ce qui permit de la contrôler, et de réduire drastiquement le taux de mortalité lié à cette pathologie dans les décennies qui suivirent.

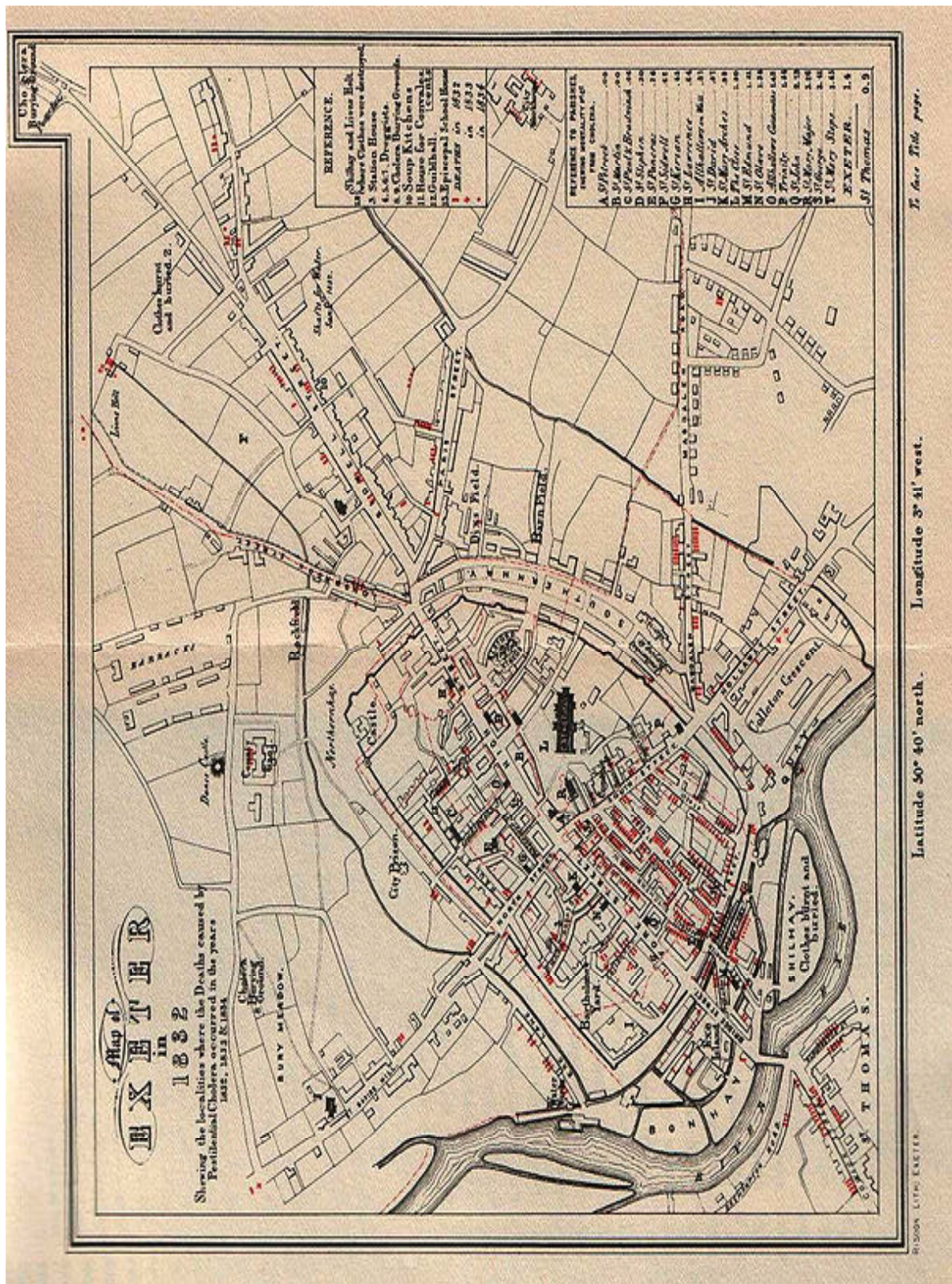
De manière plus générale, le succès de l'infographie en ligne et des nouveaux modes de visualisation dynamique ont amené les graphistes et designers à rechercher de nouvelles informations et de nouvelles sources de données à mettre en scène dans leurs productions.

Le processus de recherche, d'analyse puis de représentation graphique synthétique des données sous forme interactive, qui s'apparente à du journalisme et qu'on appelle parfois « *datajournalism* » ou journalisme de données, a participé à créer une demande pour plus de transparence de la part des gouvernements, et à l'émergence du sujet de la réutilisation des données publiques.

Les données publiques offrent, par le biais des réutilisations qui peuvent en être faites, un fort potentiel d'innovation, tant à titre commercial que pour la recherche, le développement de nouveaux services aux citoyens ou l'amélioration du travail de l'administration.

Les nouvelles possibilités de recombinaison et de réutilisation offertes par les technologies de l'information ont révélé l'importance de leur réutilisation, et contribué à en faire une question sociétale.

²³ <http://www.afigeo.asso.fr/>



Carte d'Exeter tirée de *History of the Cholera in Exeter in 1832*, T. Shapter.

Chapitre II. La libre réutilisation des données publiques à l'international

Le sujet de la réutilisation des données publiques et l'appel à la mise à disposition d'informations publiques sous forme brute a émergé dans le contexte des ruptures technologiques et économiques de ces dernières années.

C'est pourtant par deux ruptures politiques, au plus haut niveau de l'État, d'abord aux États-Unis puis en Grande-Bretagne, chaque fois dans un contexte d'alternance politique, qu'il a été concrétisé par une politique ambitieuse d'ouverture des données publiques.

Le chapitre suivant présente un tour d'horizon de ce mouvement d'ouverture des données publiques, et l'historique récent de ces deux principales initiatives gouvernementales, qui l'ont concrétisée à l'international.

1. Le mouvement *Open Data*

Le terme *Open Data*, littéralement « données ouvertes » ou « données libres » (au sens « libres de droits »), est utilisé pour désigner des données en libre accès et ouvertes à la réutilisation. Il trouve son origine en 1995 dans une publication du Comité sur les Données géophysiques et environnementales du Conseil national de la Recherche aux États-Unis intitulée « De l'échange complet et ouvert des données scientifiques »²⁴ :

« Les programmes internationaux de recherche sur le changement global et la surveillance environnementale dépendent du principe d'un échange complet et ouvert des données (i.e. les données et les informations sont rendues disponibles sans restriction, sur une base non discriminatoire, pour un coût réduit à celui des besoins de reproduction et de distribution). »

Le terme, qui désigne donc à l'origine le processus de collaboration des scientifiques autour de données d'expériences partagées afin de multiplier les chances de détecter des erreurs, se tenir mutuellement au courant des

nouvelles évolutions, et collaborer à la rédaction d'articles scientifiques, se répand largement au sein de la communauté des chercheurs. Il trouve ensuite un écho dans la communauté des développeurs, notamment parmi les contributeurs de projets Open Source, dont le mouvement d'ouverture des données est philosophiquement proche.

Des développeurs, des chercheurs et des économistes de l'information créent ainsi l'Open Knowledge Foundation (OKFN) en 2003 et l'*Open Data* Foundation (ODaF)²⁵ en 2006 – organismes à but non lucratif qui cherchent à promouvoir l'adoption de normes mondiales de formats et de métadonnées (description des jeux de données), et le développement de solutions *open source* pour mutualiser les outils de mise en ligne de données.

Les promoteurs de l'*Open Data* ont ainsi développé des définitions formelles des critères qui caractérisent des données « ouvertes ». Un groupe de travail « Open Government Data » (OGD) d'experts du Web s'est fixé huit critères assez larges de définition des données publiques ouvertes²⁶. L'Open Knowledge Foundation (OKFN), pour sa part, propose onze critères²⁷, plus axés sur les conditions juridiques acceptables qui doivent se retrouver dans la licence de réutilisation.

Les deux ensembles de critères, qui établissent une première définition des données ouvertes, sont détaillés ci-dessous.

Données complètes : toutes les données sont mises à disposition, sauf exceptions de bon sens.

Données brutes : les données sont telles que collectées à la source. Elles ont la plus fine granularité possible et ne se présentent pas sous des formes agrégées ou modifiées.

Données récentes : les données sont mises en ligne aussi rapidement que possible.

²⁴

<http://www.nap.edu/readingroom.php?book=exch&page=summary.html>

²⁵ <http://www.opendatafoundation.org/>

²⁶ <http://www.opengovdata.org/home/8principles>

²⁷ <http://www.opendefinition.org/okd/>

Données accessibles : les données sont accessibles au plus grand nombre d'utilisateurs possible, et réutilisables pour des usages les plus variés.

Données libres de droits : les données ne sont pas soumises au droit d'auteur, à brevet, au droit des marques, ou au secret commercial.

Formats exploitables : les données sont structurées pour permettre le traitement automatisé, et disponibles sous une forme pratique et modifiable.

Formats non-propriétaires : les données sont accessibles dans un format sur lequel aucune entité n'a de contrôle exclusif.

Conditions non discriminatoires : les données sont accessibles à quiconque sans aucune obligation préalable ni inscription, et sans discrimination entre les domaines d'application. Les données sont notamment disponibles aussi bien pour des réutilisations commerciales que personnelles.

Paternité : la licence peut exiger, comme condition pour la redistribution et la réutilisation des données, d'identifier clairement les créateurs de l'œuvre ou les contributeurs initiaux. Si cette condition est imposée, elle ne doit pas être indûment complexe à remplir, et notamment la liste des contributeurs à citer doit être clairement indiquée.

Distribution de la licence : les droits attachés aux données s'appliquent à tous ceux à qui elles sont distribuées.

Non-transitivité de la licence : la licence ne peut exiger que toute donnée distribuée conjointement avec une donnée sous licence libre soit aussi elle-même sous licence libre, ce qui serait un frein à la réutilisation de ces données.

Droit d'extraction : si les données sont extraites d'un jeu de données, les utilisateurs de l'extrait doivent bénéficier des mêmes droits que ceux qui sont accordés au jeu de données dans son ensemble.

Bien que ce modèle des données ouvertes puisse s'appliquer à tous types de données, qu'elles soient produites par le secteur public ou non, l'importance démocratique de l'accès aux informations publiques a conduit à focaliser le débat sur l'Open Government Data, ou la réutilisation des données publiques.

Les évolutions technologiques, le cadre réglementaire (notamment les lois sur le droit à l'information et la Directive européenne sur la réutilisation des informations du secteur

public) et l'opportunité politique ont mené au lancement, aux États-Unis, en Grande-Bretagne et ailleurs en Europe notamment, d'initiatives majeures d'ouverture et de réutilisation des données publiques.

2. États-Unis : initiatives locales, puis démarche fédérale

L'ouverture des données s'est effectuée aux États-Unis du bas vers le haut. Les villes de Washington, New York, San Francisco, ont entrepris de recenser les collections de données qu'elles envisagent de mettre à disposition, puis d'ouvrir des « *data stores* » pour mettre ces jeux de données à la disposition du public des entreprises, des développeurs et des citoyens.²⁸

Ces démarches de « libération » des données publiques sont très souvent accompagnées de concours d'applications pour encourager la réutilisation des données rendues publiques.

Ce mouvement, amorcé par les grandes métropoles, fait désormais l'objet d'une politique au niveau fédéral.

(i) La réutilisation des données émerge dans les grandes métropoles.

L'année 2005 est marquée par une rupture majeure : l'ouverture par Google des « interfaces de programmation » (API) du service Google Maps donne lieu à une floraison d'applications et de *mashups*²⁹. Un grand nombre de ces applications consistent à positionner sur les cartes Google Maps des données d'origine publique.

Ainsi, à Chicago en 2005, Adrian Holovaty, journaliste et programmeur³⁰ lance ChicagoCrime.org, un site qui renseigne sur une carte les données relatives aux crimes et délits commis dans la ville et récupérées auprès de la municipalité³¹. Suite au succès de l'application, et grâce à un don d'un million de dollars de la Knight Foundation³² en 2007, il

²⁸ Benhamou, B., 2010, « Les concours d'applications mobiles dans le monde », Délégation aux Usages de l'Internet, 21 octobre 2010.

²⁹ Applications web combinant différentes sources au sein d'une seule plateforme, l'une des facettes du design collaboratif sur le web.

³⁰ Auteur de la librairie de développement en langage Python à succès, Django.

³¹ Entretien le 18 octobre 2010 avec Sylvain Parasio, LATTS, Ecole des Ponts ParisTech.

³² Organisation caritative dédiée aux avancées dans le domaine du journalisme. Ce type d'acteurs dotés en fonds

développe une version commerciale du site, EveryBlock³³ (pour « chaque pâté de maison »), qui recense et agrège un grand nombre de données potentiellement utiles pour les citoyens dans un plusieurs municipalités.

Des services du même type, conçus à partir de données publiques des municipalités voient le jour dans les domaines de l'éducation, de l'environnement ou de la santé.

(ii) Succès médiatique des concours d'applications.

Vivek Kundra, responsable des systèmes d'information du District de Columbia (autorité administrative de la ville de Washington), s'empare en 2008 du sujet. Il lance *Apps for Democracy*, un concours de développement d'applications de services publics, ouvert à tous les développeurs, libres de s'approprier des jeux de données mis à leur disposition pour l'occasion³⁴.

Contre toute attente, le concours d'application, doté par la Mairie d'un prix de 50 000 dollars, connaît un succès considérable. Alors qu'ils espéraient une dizaine de réponses, les organisateurs reçurent 47 dossiers en 30 jours seulement. Le coût estimé du travail de conception et de programmation ainsi mis à disposition des habitants de Washington a été évalué à plus de 2 600 000 dollars, soit un retour sur investissement de 5 100 % et un effet de levier de 52 pour 1.

Apps for Democracy fait école : les concours d'applications essaient au sein des grandes métropoles américaines (notamment New York, San Francisco, et Vancouver au Canada)³⁵. En 2010, la Banque Mondiale elle-même lance son concours (*Apps for Development*), qui reçoit plus d'une centaine de contributions.

(iii) Le « gouvernement ouvert » du Président Obama

Le candidat Barack Obama avait fait du « Gouvernement ouvert » un des axes de son programme. Le 21 janvier 2009, au lendemain

de sa cérémonie d'investiture, le Président Obama signe son premier acte politique : un mémorandum à destination des ministres et directeurs d'agences intitulé « *Transparency and Open Government* » (Transparence et Gouvernement Ouvert). Il assigne aux services fédéraux américains trois objectifs : transparence, implication des citoyens, et collaboration.

Adrian Fenty, le maire de Washington, D.C. à l'époque, recommande Vivek Kundra au Président Obama : le 5 mars 2009, ce dernier est nommé *Chief Information Officer* du gouvernement fédéral américain (Directeur des Systèmes d'Information), un poste nouveau et symbolique créé sur mesure, avec pour mission de mettre en place les technologies pour concrétiser la vision politique du chef de l'exécutif.

(iv) Le portail fédéral Data.gov

Dès le 11 mars, Vivek Kundra demande aux DSI de toutes les agences fédérales américaines de désigner des jeux de données (« *datasets* ») susceptibles d'être mis en ligne. Se basant sur son expérience à Washington, il se concentre sur les données brutes. Data.gov sera un point focal faisant autorité pour diffuser les données publiques brutes, sous des formats réexploitables.

Son équipe développe une première version beta du site en trois mois, grâce à l'utilisation de techniques de développement dites « agiles » qui découpent la réalisation du produit en cycles rapides (de vingt-quatre heures à quelques jours).

Cette méthodologie aboutit à un processus de développement rapide et itératif, qui assure une rapidité de développement et la possibilité d'intégrer continuellement les évolutions des besoins qui apparaissent très fréquemment dans le cadre de projets de développement logiciel : « *commencer petit à petit, réfléchir vite, faire un premier jet, le construire de façon incrémentale, prendre du recul. Trouver un moyen pour avancer* »³⁶.

Le site est mis en ligne le 21 mai 2009, 120 jours seulement après la signature du Mémorandum par le Président avec seulement 47 ensembles de données. Le portail connaît dans les mois suivants une évolution très

et organisés pour soutenir des programmes d'action ciblés, reste rare en France mais joue un rôle important dans le débat social aux États-Unis.

³³ EveryBlock.com est désormais une filiale de MSNBC.

<http://www.everyblock.com/>

³⁴ Lakhani, R., Austin, R. and Yi, Y. *Data.gov Case Study*. Harvard Business School.

http://www.data.gov/documents/hbs_datagov_case_study.pdf

³⁵ Bernard Benhamou. 2010. « Les Concours d'Applications Mobiles dans le Monde », note de la Délégation aux Usages de l'Internet.

³⁶ Lakhani, R., Austin, R. and Yi, Y. *Data.gov Case Study*. Harvard Business School (voir plus haut).

rapide. Lors du premier anniversaire, en mai 2010, l'équipe de Data.gov annonçait³⁷ :

- 250 000 jeux de données uniques en ligne ;
- Plus de 97 millions de requêtes en 1 an, soit près de 2 millions de pages vues ;
- Plus d'un millier d'applications listées construites à partir des « ensembles de données » mis à disposition des développeurs.

Quelques enseignements peuvent être tirés de cette expérience :

- La demande pour des données publiques brutes a un potentiel élevé ;
- L'intérêt du public est variable selon les types de données, mais difficile à prévoir *a priori* ;
- La rapidité de mise à disposition des données, sous forme brute et réexploitable, a créé les conditions du succès ;
- Les fonctionnalités d'abord limitées de la plateforme n'ont pas empêché une expansion rapide par la suite ;
- Les résistances à l'ouverture des données ont pu être dépassées par un appui politique de premier ordre, et le rôle de démonstrateur joué par la première version beta de la plateforme.

3. Grande Bretagne : une impulsion transpartisane du gouvernement

L'initiative britannique a pris corps sous le gouvernement de Gordon Brown : elle a été reconduite et même amplifiée par le gouvernement Cameron. L'annonce le 12 janvier 2011 de la création prochaine de la *UK Public Data Corporation* (agence britannique pour les données publiques) est une première.

(i) Une genèse qui doit au hasard

Sir Tim Berners-Lee, inventeur du Web en 1989 et artisan de l'émergence du « Web sémantique » a joué un rôle essentiel. Son appel à mettre en ligne les données publiques brutes lors de la conférence TED en 2009 avait reçu un écho important dans le monde

numérique, notamment d'un des principaux journaux britanniques, le *Guardian*, qui avait fait campagne pour l'ouverture des données publiques (« *Free Our Data* »).

À Gordon Brown qui l'interrogeait début 2009 sur les mesures phares qu'il pourrait prendre pour faire progresser le numérique britannique, Tim Berners-Lee avait répondu que la mise en ligne des données publiques britanniques serait à elle seule une avancée majeure³⁸.

« L'aspect le plus fascinant du Web c'est que, quand vous mettez quelque chose en ligne, vous ne savez pas qui va l'utiliser – mais vous savez que quelqu'un va l'utiliser. »

Si ces données – performances scolaires, localisation des adresses postales, météo, et surtout les cartes – étaient mises en ligne, explique-t-il, chacun y trouverait une utilité.³⁹

Avec le Professeur Nigel Shadbolt, chercheur reconnu dans les technologies du Web sémantique, Tim Berners-Lee se voit alors chargé d'accompagner les efforts de mise en ligne d'un portail d'informations publiques sur le modèle du Data.gov américain, Data.Gov.uk. Une première version du portail est ouverte le 30 septembre 2009⁴⁰, préfigurant le concours d'idées civiques *Show Us A Better Way*⁴¹. En parallèle, s'amorce un mouvement de mise en ligne de données géographiques, dont une grande partie a été rendue gratuite, une révolution pour l'Ordnance Survey, l'institut géographique national britannique⁴².

³⁸ Tom Chatfield and James Crabtree. 27 janvier 2010. « Mash the state : opening up public sector data is an old geek hobbyhorse, but could the man who invented the web reinvent the British government ? ». Prospect Magazine.

<http://www.prospectmagazine.co.uk/2010/01/mash-the-state/>

³⁹ Texte original, tiré de l'article de Chatfield et Crabtree cité plus haut :

As Berners-Lee explains: « *The thing people are amazed about with the web is that, when you put something online, you don't know who is going to use it—but it does get used.* » So his pitch to Gordon Brown was simple: if this data—exam results, postbox locations, weather reports, and most crucially, maps—was put online, people would find a use for it.

⁴⁰ « Sneak Peek Site Open ». 30 septembre 2009.

Data.gov.uk Blog. <http://data.gov.uk/blog/sneak-peek-site-open>

⁴¹ « What would you create with public information ? » 20 janvier 2010. Show Us A Better Way Blog.

<http://webarchive.nationalarchives.gov.uk/20100807004350/http://showusabetterway.co.uk>

⁴² Great Brittan's Ordnance Survey Annual report 2009.

<http://www.ordnancesurvey.co.uk/oswebsite/aboutus/reports/annualreport/09-os-annual-report-and-accounts-2010.pdf>

³⁷ The White House. 21 mai 2010. « Data.gov : Pretty Advanced for a One Year Old ». <http://www.whitehouse.gov/blog/2010/05/21/datagov-pretty-advanced-a-one-year-old>

(ii) Un mieux-disant politique autour de la mise en ligne des données publiques

Le thème acquiert une visibilité politique à l'occasion des élections parlementaires de mai 2010. En réaction au retentissement médiatique du scandale des notes de frais des parlementaires à l'été 2009, le thème de la transparence et celui de l'ouverture des données publiques figurent dans le programme de tous les candidats. Une sorte de surenchère à la mise en ligne de données publiques se met en place à l'occasion de la campagne.

Après l'alternance politique, David Cameron reprend à son compte l'initiative data.gov.uk, et réoriente ses priorités, de l'innovation et du soutien à la création de nouveaux services aux citoyens vers la mise en place d'outils de transparence de l'action de l'État. Il nomme un *Public Sector Transparency Board* (qui compte dans ses membres Tim Berners-Lee et Nigel Shadbolt) et affiche sa volonté de construire « *le gouvernement le plus transparent au monde* », en s'appuyant sur une mise en ligne généralisée des données de fonctionnement des ministères et agences britanniques, vu comme un élément de contrôle du « *Big State* » par la « *Big Society* » (son thème de campagne).

(iii) Fédérer un écosystème performant

Ce consensus pour l'ouverture des données de l'État au sein de la classe politique doit beaucoup aux efforts de conviction de Tim Berners-Lee et aux travaux de l'Open Knowledge Foundation⁴³.

Un rédacteur de ce rapport a pu assister à une conférence de presse le 19 novembre 2010 à Londres, au cours de laquelle fut annoncée la mise en ligne de l'ensemble des dépenses de l'administration centrale britannique, détaillées à partir de 25 000 livres et remises à jour tous les mois.

Cette conférence de presse, qui réunit Sir Tim Berners-Lee, Francis Maude, *Minister for the Cabinet Office, Paymaster General* et le Premier Ministre David Cameron (via une allocution vidéo)⁴⁴ a été largement relayée par les médias.

Un travail de collaboration entre le gouvernement, le journal *The Guardian* et l'*Open Knowledge Foundation* (qui a développé la technologie CKAN⁴⁵ utilisée par

⁴³ Open Knowledge Foundation. <http://okfn.org>

⁴⁴ Allocution par le Premier Ministre David Cameron, 19 novembre 2010. <http://data.gov.uk/about>

⁴⁵ Comprehensive Knowledge Archive Network. <http://ckan.net/>

la plateforme Data.gov.uk) a permis de mettre au point des outils de visualisation et d'exploration de ces ensembles de données. L'outil de visualisation du budget de l'État britannique destiné au grand public réalisé en amont par OKFN, *Where Does My Money Go*⁴⁶ fut présenté lors de cette conférence de presse.

Furent aussi mis en avant, le même jour, les acteurs privés les plus innovants en matière de réutilisation des données publiques, comme Timetric⁴⁷, qui développe un « YouTube des graphes de données⁴⁸ » et finance son produit consommateur gratuit par des services de *business intelligence* aux entreprises.

À l'occasion de cette conférence de presse fut aussi consacré le travail de Rufus Pollock, économiste à l'Université de Cambridge, cofondateur de l'*Open Knowledge Foundation* (organisatrice de l'*Open Government Data Camp* concomitant), et auteur de l'étude économique majeure sur le sujet des données publiques.

(iv) Tensions entre pouvoirs locaux et administration centrale

Dans un contexte de coupes budgétaires drastiques, l'ouverture des données publiques cristallise cependant des tensions entre le gouvernement central et les pouvoirs politiques locaux.

Eric Pickles, le ministre en charge des Collectivités Locales, avait fixé aux collectivités la date butoir du 31 janvier 2011 pour mettre en ligne les données relatives aux dépenses budgétaires excédant 500 livres. À quelques semaines de l'échéance, seules 150 collectivités sur les 350 concernées avaient publié leurs données⁴⁹, même si la majorité s'est ensuite pliée à cette demande.

Le ministre Eric Pickles est revenu sur le périmètre des données que ces institutions seraient tenues de diffuser⁵⁰ sans délai, de

⁴⁶ WhereDoesMyMoneyGo a été développé par Jonathan Gray de l'*Open Knowledge Foundation* en 2007, et remis à jour au fur et à mesure. <http://wheredoesmymoneygo.org/>

⁴⁷ Timetric : Indexing the World Economy. <http://timetric.com/>

⁴⁸ TechCrunch. 20 avril 2009. « Four teams selected for Mini Seedcamp day in London ». <http://eu.techcrunch.com/2009/04/20/four-teams-selected-from-mini-seedcamp-day-in-london/>

⁴⁹ The Register. 6 janvier 2011. « Pickles Throws Down Gauntlet on Data. » http://www.theregister.co.uk/2011/01/06/councils_must_put_spending_data_online_or_else/

⁵⁰ TransportXtra. N°562, 14 janvier 2011. « Pickles scraps accessibility and school mode choice datasets. » http://www.transportxtra.com/magazines/local_transport_today/news/?id=25331

manière à apaiser les tensions apparues en raison d'une trop forte ambition initiale.

3.2. La « Public Data Corporation », une compagnie des données publiques britannique

Francis Maude annonce le 12 janvier 2011 la création de la *UK Public Data Corporation* (agence britannique pour les données publiques)⁵¹.

*« Les données publiques soutiennent une part croissante de l'économie. Les technologies actuelles permettent à chacun d'utiliser et de réutiliser cette information de maintes façons nouvelles et innovantes. Le rôle du gouvernement est d'aider à maximiser les bénéfices que la société retire de ces développements. Aujourd'hui, beaucoup d'agences font face à un conflit entre les revenus issus de la vente de données et la possibilité de rendre les données disponibles gratuitement, pour être exploitées à des fins économiques ou sociales. Créer la Public Data Corporation permettra au moins de gérer les conflits d'intérêt en visant à ouvrir l'accès aux données, et au mieux de les éliminer ».*⁵²

L'agence pourra décider la mise en ligne des données. Arbitre des conflits d'intérêt entre génération de revenus et bénéfices sociétaux, elle sera chargée d'harmoniser les politiques tarifaires afin de préserver leur cohérence et de faciliter la réutilisation et l'innovation. Elle devrait aussi jouer un rôle dans les investissements pour des données publiques de qualité.

Les réactions dans la communauté des réutilisateurs britanniques sont mitigées : ils redoutent un retour vers un modèle de tarification généralisée des données publiques,

⁵¹ The Cabinet Office. 12 janvier 2011. « *Public Data Corporation to free up public data and drive innovation.* » <http://www.cabinetoffice.gov.uk/news/public-data-corporation-free-public-data-and-drive-innovation>

⁵² Texte original de l'annonce telle que relatée par le site du Cabinet Office le 12 janvier 2011 :

Public sector information underpins a growing part of the economy. The technology that is around today allows people to use and re-use this information in new and different ways. The role of Government is to help maximise the benefits of these developments. At present many state agencies face a conflict between maximising revenues from the sale of data and making the data freely available to be exploited for social and economic gain. Creating the PDC will enable the conflicts at the least to be managed consistently with a view to opening up access, and at best to be eliminated.

qui pourraient mettre à mal les efforts réalisés jusqu'à présent pour soutenir la réutilisation.⁵³

4. Situations contrastées en Europe

(i) Contrastes forts entre États-membres

La transcription en droit national et la mise en œuvre de la Directive « PSI » (Public Sector Information, ou données du secteur public) de 2003 débouchent sur des situations très contrastées dans les États-membres.

Les pays scandinaves, avec une culture historiquement ancrée de la transparence, sont souvent en pointe en matière de réutilisation des données publiques. La Finlande a notamment annoncé avoir lancé un processus exploratoire pour étudier la mise en ligne de ses données.

Les institutions européennes, Eurostat et le site de droit communautaire <http://e-justice.europa.eu> en premier lieu, autorisent la réutilisation des données et informations publiques issues des institutions européennes.

Le degré d'avancement des autres pays est difficile à bien percevoir, notamment du fait du manque de visibilité des projets de portails nationaux ou de l'approche très décentralisée qui a souvent été retenue (comme par exemple avec les Répertoires d'Information Publique choisis par la France).

(ii) Réexamen de la Directive PSI

La Commission européenne a lancé en septembre 2010 une consultation publique sur la directive « Informations du secteur public » : elle portera sur le champ d'application, les tarifs et licences relatives aux données ainsi que sur les formats électroniques disponibles. Selon Neelie Kroes, vice-présidente de la Commission européenne chargée de la stratégie numérique, « *une grande partie de ces informations, en Europe, sont sous-exploitées. Nous ne pouvons nous permettre de négliger ce potentiel. Nous devons examiner l'opportunité de modifier les règles de l'UE en matière de réutilisation des informations du*

⁵³ Chris Taggart (membre du comité du gouvernement britannique sur les données publiques locales), 14 janvier 2011. « What's that coming over the hill, is it... the Public Data Corporation? ». Publié sur : <http://countculture.wordpress.com/2011/01/14/whats-that-coming-over-the-hill-is-it-the-public-data-corporation/>

secteur public pour libérer pleinement leur potentiel économique »⁵⁴.

La Direction générale pour la Société de l'information et médias de la Commission Européenne a financé une plateforme électronique d'échange et de veille autour de ces questions : ePSIplatform.⁵⁵

La Commission a par ailleurs lancé une étude visant à évaluer l'opportunité d'une refonte de la Directive PSI, notamment vers un champ plus large, des règles de tarification plus contraignantes ou des préconisations plus contraignantes sur les licences de réutilisation.

Le travail de réflexion sur le sujet et d'animation de la communauté des réutilisateurs, notamment via un concours européen de développement d'applications, l'*Open Data Challenge*, va déboucher sur une forte visibilité de la question des données au niveau européen. L'*Open Data* a été l'un des sujets majeurs du *Digital Assembly* de juin 2011, premier rassemblement annuel de suivi de la stratégie européenne pour le numérique, ce qui en souligne l'actualité brûlante.

⁵⁴ Stratégie numérique: la Commission lance une consultation sur la réutilisation des informations du secteur public.

<http://europa.eu/rapid/pressReleasesAction.do?reference=IP/10/1103&format=HTML&aged=0&language=FR&guiLanguage=en>

⁵⁵ <http://www.epsiplus.net/>

Chapitre III. Les enjeux de la réutilisation des données publiques

Portée par des évolutions technologiques, économiques et politiques récentes, la réutilisation des données publiques est devenue une question sociétale.

Elle peut contribuer à rapprocher la puissance publique des citoyens et améliorer le fonctionnement interne de l'administration.

Elle constitue un élément de soutien à l'économie numérique, et ses applications peuvent encourager le développement durable ainsi que les secteurs essentiels pour l'avenir de notre économie, de la recherche et de l'innovation des PME.

Les données publiques constituent les prémisses d'une « infrastructure informationnelle » : au-delà des gains économiques liés à la transparence, elles permettent l'avènement d'un nouveau mode d'action publique, novateur et encore traversé par des tensions qu'il appartient à la puissance publique de réguler.

1. Démocratie, transparence et débat public

Le droit d'accès aux données publiques et, son corollaire de l'ère numérique, le droit à leur réutilisation non discriminatoire et sans distinction d'usage, sont des éléments à part entière du fonctionnement démocratique de notre société.

1.1. Contribuer à la transparence de l'action de l'État

L'obligation pour la puissance publique de rendre compte à ses administrés (« *accountability* » pour les Anglo-Saxons, littéralement « être comptable de ») est une valeur centrale des démocraties modernes.

Cette responsabilité du gouvernement se concrétise notamment par un droit d'information des citoyens, et par une nécessité de transparence de l'État et des collectivités quant à leur fonctionnement vis-à-vis du public.

En annonçant, en novembre 2010, la mise en ligne de l'ensemble des dépenses de

l'administration centrale britannique dépassant de 25 000 livres⁵⁶, le Premier Ministre David Cameron revendiquait le titre de « gouvernement le plus transparent et responsable de la planète ».

L'exemple le plus radical de cette transparence est probablement le site Police.uk, lancé par le même gouvernement, qui détaille l'ensemble des crimes, délits et incivilités, rue par rue, à travers tout le territoire britannique. Il a connu un succès extraordinaire, avec 35 millions de requêtes durant ses premières 24 heures de fonctionnement.

La transparence de l'action de l'État ne se limite pas aux domaines strictement régaliens. La transparence sur la qualité de l'environnement, des produits que nous consommons et de l'air que nous respirons est un ferment crucial de la confiance des citoyens envers leurs gouvernants.

Par exemple, lors de la crise nucléaire de Fukushima, l'Institut de Radioprotection et de Sécurité Nucléaire français, l'IRSN, a ainsi mis en ligne les mesures de radioactivité ambiante qu'il effectuait en temps réel à travers tout le territoire.

Enfin, la transparence publique est un enjeu démocratique qui doit évidemment être mis en place dans le respect de la vie personnelle et des secrets légaux (secret statistique, secrets liés à la sécurité, secret lié à la protection l'environnement...).

1.2. Éclairer le débat public

La mise à disposition de données brutes sur le fonctionnement de l'État va dans le sens de la généralisation de ce mouvement de transparence.

⁵⁶ Allocution de David Cameron, 19 novembre 2010, durant la conférence de presse à l'occasion de la mise en ligne des entrées de 25,000 livres et plus du budget de fonctionnement de l'État britannique.
<http://www.guardian.co.uk/politics/blog/2010/nov/19/government-spending-files-live-blog?intcmp=239>
Voir aussi l'allocution de Francis Maude, Ministre du Cabinet Office :
<http://www.guardian.co.uk/commentisfree/2010/nov/19/francis-maude-government-data-published?intcmp=239>

Une réutilisation libre et ouverte des données publiques transforme les pratiques traditionnelles du journalisme. L'accès direct à ces données confère aux journalistes une plus grande latitude pour la conduite d'investigations basées sur les faits et la construction d'argumentaires étayés. Ils peuvent y accéder à tout moment, les recouper à titre de vérification, les croiser avec d'autres sources pour mettre en perspective les phénomènes qui sont l'objet de leurs analyses.

Le journal britannique *The Guardian*⁵⁷ publie ainsi des analyses infographiques et des visualisations interactives qui mettent en lumière les éléments factuels qui sous-tendent le débat public, comme par exemple celle portant sur l'analyse du budget britannique⁵⁸.

Ce phénomène, le « *datajournalism* » ou journalisme des données, permet à la presse de mettre en œuvre une véritable pédagogie quant aux enjeux de société, sous une forme interactive et que le public peut facilement s'approprier, à partir des données qui s'y rapportent.

L'accès aux données brutes et le droit de réutilisation permettent donc d'éclairer le débat public, et de le recentrer sur des faits avérés plutôt que des positions idéologiques formées *a priori*.

1.3. Encourager l'implication des citoyens dans la vie publique

L'émergence de mécanismes participatifs, qui permettent d'associer les citoyens à la délibération politique menée par les institutions de la démocratie représentative, a précédé l'émergence d'Internet.

En France, ce principe de participation a été créé par la loi « Barnier » du 2 février 1995 relative à la protection de l'environnement et instituant la commission nationale du débat public (CNDP), et la loi « Vaillant » du 27 février 2002 relative à la démocratie de proximité instaurant les conseils de quartier.

Dans nos sociétés basées sur la communication et la fluidité des échanges, toute rétention d'information est perçue comme suspecte. En offrant des outils qui permettent aux citoyens de se faire une opinion indépendamment des pouvoirs publics ou du prisme d'analyse des médias, la mise à disposition des données peut permettre

l'émergence d'une réelle expertise citoyenne impliquée dans la vie publique.

Le site NosDéputés.fr est un cas de réutilisation par un collectif citoyen de données publiques (dans ce cas des données publiques produites par la Représentation nationale). Créé par l'association Regards citoyens, le site⁵⁹, lancé le 14 septembre 2009, ambitionne « de donner aux citoyens de nouveaux outils pour comprendre et analyser le travail de leurs représentants ». Pour cela, il synthétise les différentes activités, législatives et de contrôle du gouvernement. En réutilisant les résultats des élections législatives de 2007, l'association a aussi étudié l'impact du redécoupage de la carte électorale sur la répartition des sièges à l'Assemblée nationale⁶⁰.

L'ouverture la plus large des données publiques à la réutilisation permettrait de multiplier ces exemples, dans l'ensemble des domaines d'activité du monde associatif et de la société civile, de l'environnement à l'assistance aux personnes en passant par l'initiative économique, et encouragerait l'implication des citoyens dans la vie civique.

2. Modernisation de l'État

2.1. Renforcer l'efficacité de l'action publique et de l'administration

L'État est, dans les faits, son plus gros client sur le marché de la vente de données. Des études démontrent que jusqu'à 70 %⁶¹ du volume de vente de données publiques concerne des transactions contractées entre deux institutions publiques différentes dépendant du même gouvernement. La tarification pour réutilisation de données, plutôt qu'une source de revenus pour l'administration, est donc majoritairement un transfert budgétaire entre entités publiques, d'autant plus inefficace qu'il grève le budget de ces institutions des coûts importants de transactions inhérents aux finances publiques.

Mettre en ligne par défaut les données publiques plutôt qu'attendre une requête en demandant l'accès facilite donc aussi le travail de l'administration. Ouvrir ses silos de

⁵⁷ <http://www.guardian.co.uk/data>

⁵⁸ <http://www.guardian.co.uk/news/datablog/2010/oct/18/government-spending-department-2009-10>

⁵⁹ <http://www.regardscitoyens.org/nosdeputes-fr/>

⁶⁰ <http://www.regardscitoyens.org/redecoupage/>

⁶¹ Communication lors de l'Open Government Data Camp à Londres le 18 novembre 2010 : dans le cas des données publiques locales en Grande Bretagne, une étude a montré que 70% des achats de données (en volume monétaire) provenaient d'autres gouvernements ou institutions gouvernementales (locales ou nationales).

données fluidifierait les échanges d'informations en son sein, notamment sur des sujets transverses qui touchent aux prérogatives de différents ministères - par exemple, l'hypertrophie des algues en milieu marin (mesures réalisées par le Ministère de l'Écologie), fortement corrélée à l'utilisation de fertilisants phosphatés dans le cadre de la production agricole (données du Ministère de l'Agriculture). L'ouverture des données publiques permettrait donc aux services de l'État et des collectivités un accès plus rapide à l'information, sans avoir besoin de passer par la lourdeur de la procédure d'achat des données auprès d'une autre entité publique, ni la nécessité de disposer d'un budget pour les obtenir.

Le processus de normalisation inhérent à la mise en ligne des données publiques pourrait de plus être intégré directement dans le travail quotidien des services producteurs de données. S'assurer de la présence de la légende, des métadonnées, ou de l'interopérabilité du format utilisé permettrait à l'État de s'engager dans une démarche d'assurance qualité concernant les données publiques.

Enfin, la transparence accrue que pourrait permettre une large ouverture des données de l'État est un élément de contrôle des citoyens sur les performances des politiques publiques et de ses opérateurs. Selon le principe des théories économiques de l'agence⁶² qui veulent que ce que l'on mesure précisément concentre les efforts, une meilleure transparence quant au travail de l'État et aux effets de son action sur la société, l'économie et les services publics entraînera mécaniquement une amélioration de son efficacité et de la qualité des services rendus.

2.2. Proposer de nouveaux services publics

La réutilisation des données par des développeurs, des entreprises ou des associations, permet par ailleurs à ces derniers d'offrir aux citoyens de nouveaux services publics en ligne, souvent novateurs et que l'État n'aurait parfois pas su ou pu imaginer par manque de familiarité avec les nouveaux usages.

En Grande-Bretagne par exemple, c'est avec le slogan « *Releasing Innovation* » (libérer l'innovation) qu'a été mise en ligne la plateforme data.gov.uk. L'objectif affiché était de permettre à des tiers de créer, à partir des

⁶²

http://fr.wikipedia.org/wiki/Th%C3%A9orie_de_l'agence

données de l'État, des services en ligne plus innovants et de meilleure qualité.

En France, c'est aussi la possibilité pour des développeurs d'offrir de nouveaux services aux citoyens à partir des données mises à disposition par l'État qui avait poussé le groupe d'experts du numérique, réunis autour du député Frank Riester pour travailler sur l'amélioration du lien numérique entre l'administration et l'utilisateur, à proposer la mise en ligne d'un portail des données de l'État⁶³.

Le discours du Président de la République prononcé le 27 avril 2011 à l'occasion de l'installation du Conseil National du Numérique⁶⁴ abonde dans ce sens : « nous faisons appel aux entrepreneurs d'Internet pour imaginer les services que l'État n'a pas encore créés ».

La perception de l'administration par le citoyen en sortira modifiée, dans un sens qui pourrait être celui d'une meilleure appréciation du travail fourni et des enjeux forts des métiers de l'administration.

Cette démarche d'encouragement au développement de nouveaux services publics en partenariat avec les acteurs de l'innovation, comme celle d'ouverture et de transparence, participe à la réforme de l'État, de la modernisation et de l'amélioration de son rapport à l'utilisateur.

3. Innovation et soutien à l'économie numérique

En parallèle aux enjeux démocratiques et d'action publique que représente l'accès aux données publiques, leur réutilisation par des développeurs, des entrepreneurs et des *startups* est un enjeu économique fort, à court terme pour le développement du numérique, et à long terme pour la croissance, l'emploi et le développement des secteurs stratégiques de l'économie.

La réutilisation des données publiques pourrait stimuler l'innovation dans les industries de l'information, notamment dans les services contextuels.

⁶³ Proposition 22 du rapport du groupe d'experts du numérique sur l'amélioration du rapport numérique de l'administration à l'utilisateur :

http://www.modernisation.gouv.fr/fileadmin/Mes_fichiers/pdf/RapportGroupeExpertsNumeriques.pdf

⁶⁴ M. le Président de la République, 2011, « Discours à l'occasion de l'installation du Conseil National du Numérique le 27 avril 2011 »

<http://www.elysee.fr/president/les-actualites/discours/2011/discours-sur-internet-et-l-economie-numerique.11256.html>

3.1. La donnée publique : matière première des industries de l'information

Une étude réalisée pour le compte de la Commission Européenne estime la valeur potentielle du seul marché de la donnée publique entre 10 et 48 milliards d'euros, avec une projection médiane à 27 milliards d'euros⁶⁵. Le potentiel de croissance qu'offrirait une plus large ouverture des données est clair, alors que l'on estime que l'Union Européenne dans son ensemble perd près de 300 millions d'euros par an en recettes fiscales du fait du régime d'accès encore trop restreint à la donnée publique⁶⁶.

Les industries de l'information considèrent la réutilisation des données publiques comme une opportunité majeure⁶⁷. Le groupement professionnel Groupement Français de l'Industrie de l'Information (GFII) organise en particulier une veille fréquente et des colloques pour ses membres sur ce sujet considéré comme crucial⁶⁸ au développement du secteur. De nombreuses réussites industrielles sont basées en tout ou partie sur l'accès à l'information publique.

Dans le secteur des transports, une société comme NAVX a pu intégrer à son offre de services aux particuliers des informations sur le prix des carburants en pompe en temps réel. La *start-up* KelQuartier.fr propose une offre d'information compilée à partir de l'agrégation d'un grand nombre de données publiques, quartier par quartier⁶⁹. Des sociétés de services aux personnes affectées d'allergies ont pu intégrer les données de volume du bâti de la ville de Paris, afin de mieux prévoir le déplacement des allergènes dans l'air à travers la ville.

⁶⁵ Dekkers, M. et al. 2006. Measuring Public Sector Information Resources : Final Report of Study on Exploitation of public sector information. http://ec.europa.eu/information_society/policy/psi/docs/pdf/mepsir/final_report.pdf

⁶⁶ Pettifer, R. 25 août 2009. « Europe loses 300 million Euro in taxation per year. » ePSI Platform. http://www.epsiplatform.eu/guest_blogs/europe_looses_300_million_euro_per_year_in_taxation

⁶⁷ Groupement Français des Industries de l'Information. 20 mai 2010. « L'ouverture des données publiques : un enjeu pour le développement de l'économie numérique. » Colloque tenu à la Maison de l'Europe. http://www.gfii.asso.fr/article.php3?id_article=3226

⁶⁸ Groupement Français des Industries de l'Information. Juillet 2010. *PSI Reuse in France : Overview and Recent Developments*. http://www.gfii.asso.fr/article.php3?id_article=3315

⁶⁹ Billaut, J-M. 26 novembre 2010. « Connaissez-vous Eduardo Larrain from Paris ? KelQuartier, premier utilisateur d'open data ? » <http://billaut.typepad.com/jm/2010/11/connaissez-vous-eduardo-larrain-from-paris-.html>

Dans un domaine traditionnellement proche des techniques documentalistes et qui déploie aujourd'hui des solutions de référencement basées sur les technologies sémantiques décrites plus haut, LexisNexis⁷⁰ ou Caselex⁷¹ sont des sources d'informations importantes (notamment jurisprudentielles) pour les juristes professionnels.

3.2. L'Open Data favorise la croissance de l'économie numérique

L'importance de la question des données publiques se révèle à la lumière des développements technologiques récents.

Au-delà des secteurs traditionnels de l'industrie de l'information, c'est l'innovation et la création de nouveaux produits et services en ligne qui bénéficieraient de l'ouverture massive des données publiques.

Permettre aux jeunes pousses numériques d'une économie d'accéder facilement aux données publiques pour expérimenter, c'est faciliter l'émergence de tout un tissu potentiel de sociétés de services mobiles et en ligne, et donc contribuer au développement de l'économie numérique.

(i) Les applications créées à partir de données publiques connaissent une très forte demande

Un exemple parmi les applications présentes sur le portail de services publics sur mobile « Proxima Mobile », créé par la Délégation aux Usages de l'Internet : *AllBikesNow* permet de repérer les stations de Vélib' proches de soi, et interroge à distance les serveurs de l'opérateur de la flotte pour connaître, en temps réel, l'état de remplissage d'une station en vue d'y retirer un vélo ou d'en garer un. Il est à noter par contre que les données brutes sous-jacentes, produites par la société JC Decaux dans le cadre de la concession d'exploitation de la flotte Vélib' concédée par la Mairie de Paris, ne sont pas encore disponibles à la réutilisation par tous, même si cela est à l'étude.

Un autre exemple dans le secteur des transports en commun est l'application mobile *Locomote*, de la société Isokron, qui permet de connaître l'itinéraire le plus rapide en fonction de sa localisation et de l'état des réseaux de bus, de train et de vélos en partage dans les villes qui ont commencé à ouvrir l'accès à ces données, comme la ville de Rennes avec la société Kéolis.

⁷⁰ LexisNexis, Inc. <http://lexisnexis.com>

⁷¹ Caselex, Inc. <http://caselex.com>

Les données qui sous-tendent ces deux services ne sont pas toujours accessibles aux développeurs indépendants dans des conditions facilitant l'expérimentation de nouvelles idées, et, par exemple, la SNCF et la RATP n'ont pas encore ouvert l'accès aux données, théoriques (horaires) comme temps réel (position des bus et trains), issues de l'exploitation de leur réseaux.

Les nouvelles plateformes d'applications mobiles ont vu apparaître des catégories entières de services nouveaux, au succès émergent. Par exemple, *Foursquare*, le service le plus connu sur le marché naissant de la géolocalisation sociale, et qui permet aux utilisateurs de signaler (*check-in*) à leurs réseaux sociaux leur localisation en un point d'intérêt (bâtiment, musée, restaurant, bar, voire ministère), a déjà plus de 5 millions d'utilisateurs⁷² après 2 ans d'existence. DisMoiOù, un service français similaire, a ouvert son API en juin 2010⁷³.

Enfin, l'information pratique aux citoyens connaît un succès renouvelé sur les nouvelles plateformes mobiles, et pourrait bénéficier substantiellement de l'ouverture des données publiques. La société MyCityWay⁷⁴ propose par exemple, dans chaque métropole américaine, une application qui regroupe l'ensemble des services publics, des informations de transports, de loisirs et de culture utiles au citoyen. Ce modèle sera probablement amené à se développer en Europe.

(ii) Les données publiques pourraient être l'opportunité d'avancées technologiques

Google, opérateur de la plateforme et du système d'opération mobile à très forte croissance Android, a annoncé à Paris lors de la conférence LeWeb 2010, sa prochaine avancée technologique majeure : la découverte contextuelle (*'contextual discovery'*⁷⁵).

L'idée est de recouper l'ensemble des informations géolocalisées ou géocodées que Google a accumulées et ordonnées dans ses serveurs, d'en extraire les informations les plus à même d'intéresser un utilisateur d'Android se trouvant dans un contexte donné (géographique, temporel, mais aussi social ou professionnel, ce qui peut être déterminé grâce aux données personnelles confiées à Google), et de les présenter de façon non intrusive à l'utilisateur avant même qu'il n'ait pensé à initier une recherche Google. Ce service préfigure les publicités mobiles géolocalisées et personnalisées qui sont annoncées pour un futur proche.

Le défi technologique et ergonomique est majeur : ce type d'application pourrait devenir aussi incontournable dans le monde physique que le sont aujourd'hui les moteurs de recherche sur le Web. Indépendamment ou en parallèle aux services contextuels de Google, d'autres applications contextuelles pourront apparaître, plus spécialisées, ou dans des contextes géographiques spécifiques où leur producteur disposera d'un avantage compétitif. Le développement de ces nouveaux usages bénéficierait largement de l'ouverture à la réutilisation des données publiques.

3.3. Un secteur crucial pour la croissance et l'emploi

Le développement d'applications innovantes à partir des données publiques ouvertes permettra d'offrir de nouveaux services utiles aux consommateurs et aux citoyens. Ce sera aussi l'occasion pour des entreprises de créer ou de développer des activités commerciales en ligne, génératrices de revenus et d'activité économique.

L'économie numérique est un secteur crucial pour la France. Une étude⁷⁶ montre que cette industrie en forte croissance (3,7 % du PIB en 2010, et jusqu'à 5.5 % en prévision pour 2015) a représenté à elle seule 25 % de la croissance de l'économie depuis 2009, et est à l'origine d'un quart de la création nette d'emplois en France depuis 1995 (soit une contribution nette d'environ 700 000 emplois). Son développement est un des quatre leviers de croissance de l'économie française identifiés par le Conseil d'Analyse Stratégique⁷⁷, avec la

⁷² TechCrunch, 1^{er} décembre 2010, « Boom ! 5 Million Users Check-In to Foursquare » .
<http://techcrunch.com/2010/12/01/foursquare-hits-5-million-users/>

⁷³ DisMoiOù. 11 juin 2010. « Développeurs, découvrez l'API DisMoiOù ! » .
<http://blog.dismoiou.fr/2010/06/developpeurs-decouvrez-lapi-dismoiou.html>

⁷⁴ GigaOm. 19 novembre 2010. « MyCityWay Dives Into Data Comes Up With Gold » .
<http://gigaom.com/2010/11/19/mycityway-dives-into-data-comes-up-with-gold/>

⁷⁵ TechCrunch, 8 décembre 2010, « Marissa Mayer's Next Big Thing : 'Contextual Discovery' – Google Results Without Search » .
<http://techcrunch.com/2010/12/08/googles-next-big-thing/>

⁷⁶ McKinsey & Company, 2011, « Impact d'Internet sur l'économie française : Comment Internet transforme notre pays » .

⁷⁷ Conseil d'Analyse Stratégique, « Rapport thématique France 2025 : Production et emploi » . Gilbert Cette, Président de groupe.

réforme de l'enseignement supérieur et la baisse des rigidités du marché des biens et du travail : une diffusion plus large des TIC pourrait contribuer jusqu'à 0.3 % de croissance additionnelle par an.

Les données publiques pourraient plus particulièrement devenir un ingrédient important d'encouragement au développement d'applications mobiles. Le marché global de ces applications a atteint 4,4 milliards de dollars de revenus annualisés dans la première moitié de l'année 2010⁷⁸, et est estimé pouvoir atteindre près de 15 milliards de dollars d'ici à 2013⁷⁹. Parmi elles, le secteur des services contextuels et géolocalisés est en pleine croissance⁸⁰ et constitue plus qu'une évolution de l'informatique, une nouvelle opportunité pour les services numériques.

En France, une étude du Syntec Numérique⁸¹ publiée en fin d'année 2010 dénombre 5 400 éditeurs et sociétés de service en ingénierie informatique spécialisés dans la mobilité. 540 d'entre elles opèrent exclusivement dans la création d'applications mobiles, dont la majorité sont des entreprises de moins de 5 personnes. L'étude projette la création de 15 000 emplois dans ce secteur au cours des 5 prochaines années.

L'ouverture des données publiques pour permettre leur réutilisation la plus large est donc un élément d'encouragement au développement de l'économie numérique. Cette dernière est un secteur crucial pour l'innovation et l'activité économique, pour la croissance, et pour la création d'emplois.

4. Des données stratégiques pour l'économie

Au-delà de leur impact direct sur le développement de l'économie numérique, les

http://www.strategie.gouv.fr/IMG/pdf/4Productionetemplo_i.pdf

⁷⁸ TechCrunch. 20 août 2010. « Smartphone App Market Reached More Than \$2.2 Billion in the First Half of 2010 », <http://techcrunch.com/2010/08/20/smartphone-applications-market-size/>

⁷⁹ TechCrunch. 5 mars 2010. « Global Smartphone App Download Market Could Reach \$15 Billion by 2013 : Report », <http://techcrunch.com/2010/03/05/global-smartphone-app-download-market-could-reach-15-billion-by-2013-report/>

⁸⁰ GigaOm. 28 mai 2010. « GeoLocal : The Rise of Consumer Location-Based Services », <http://gigaom.com/mobile/geolocal-the-rise-of-consumer-location-based-services/>

⁸¹ Chambre professionnelle des Sociétés de Conseil et de Services Informatiques, des Editeurs de Logiciels et des Sociétés de Conseil en Technologies.

données publiques peuvent aussi jouer un rôle stratégique pour l'économie au sens large.

La transparence accrue initiée par le mouvement de mise en ligne des données publiques pourrait entraîner des gains économiques substantiels.

La réutilisation de ces données, pour créer de nouveaux services numériques ou pour améliorer les processus productifs internes aux entreprises, pourra soutenir les grands axes stratégiques pour l'avenir de l'économie française, de la recherche, du développement durable ou encore de la croissance des Petites et Moyennes Entreprises (PME).

4.1. Les gains économiques de la transparence

L'*Open Data* et ses enjeux sont souvent présentés comme séparés en deux problématiques majeures : la transparence de l'État et le débat public, d'une part, et la contribution à la croissance économique, de l'autre. La transparence peut cependant être elle aussi à l'origine de la création de valeur, tant au niveau microéconomique que macroéconomique.

La théorie de l'agence économique évoquée plus haut consacre le principe selon lequel ce qui est mesuré et visible concentre les efforts. En mettant au vu et au su de tous, et notamment des citoyens, des données sur l'action des différentes institutions publiques, l'*Open Data* contribuera mécaniquement à faire progresser la qualité et l'efficacité des services qu'ils rendent.

Des études ont montré par exemple que le simple fait de mesurer et de rendre public le taux de mortalité dans les hôpitaux (ajusté pour le risque) permettait d'atteindre des résultats presque similaires à la mise en place de systèmes d'incitations financières⁸².

Du point de vue de la concurrence, la mise en ligne de données produites ou collectées par le régulateur ou l'autorité de la concurrence pourrait permettre aux consommateurs de comparer plus efficacement les offres du marché, ce qui contribuerait à l'amélioration du pouvoir d'achat et à un marché des biens et services plus compétitif.

En parallèle à la publication des données publiques, accessibles à tous selon le modèle

⁸² New England Journal of Medicine. 2007, Public Reporting and Pay for Performance in Hospital Quality Improvement. January 26, 2007. 356:486-496 <http://www.nejm.org/doi/full/10.1056/NEJMs064964#t=articleBackground>

de l'*Open Data*, la mise à disposition par les entreprises privées à leurs clients des données d'usage privées qui les concernent, et ce sous forme réexploitable, pourrait permettre plus de compétition entre fournisseurs de services, en ligne (réseaux sociaux, moteurs de recherche) ou non (opérateurs mobiles)⁸³. Ce mouvement de « *Data Portability* »⁸⁴, ou portage des données privées, pourrait donner lieu à une offre de meilleure qualité pour les consommateurs.

Enfin, une transparence accrue quant à l'État de l'économie et de la société française contribue à recréer un lien de confiance entre les agents économiques, lien qui a pu être éprouvé par la profonde crise économique récente.

En participant au rétablissement de la confiance des investisseurs, des ménages et des entreprises dans nos institutions, dans la capacité de l'économie française à croître et dans celle de l'État à financer ses programmes de politiques publiques, l'*Open Data* peut donc contribuer à soutenir la stabilité macroéconomique de notre pays.

4.2. Un facteur de soutien aux priorités stratégiques de l'économie française

La réutilisation des données publiques permet la création de services innovants, qui peuvent favoriser le développement d'activité d'entreprises, et notamment de celles issues de l'économie numérique.

Ces services peuvent aussi avoir un impact indirect fort à moyen et long terme sur l'économie, en contribuant aux grands axes stratégiques d'avenir identifiés pour son développement.

La commission Juppé-Rocard, dirigée par les deux anciens Premiers, a été chargée en 2009 par le président de la République de formuler des propositions quant aux investissements d'avenir que la nation pourrait mettre en œuvre afin de préparer son développement futur.

Ses conclusions⁸⁵ identifient sept grands thèmes prioritaires :

- Soutenir l'enseignement supérieur, la recherche, l'innovation.
- Favoriser le développement des PME innovantes.
- Accélérer le développement des sciences du vivant.
- Développer les énergies décarbonées et l'efficacité dans la gestion des ressources.
- Faire émerger la ville de demain.
- Inventer la mobilité du futur.
- Investir dans la société du numérique.

Les nouveaux services basés sur la réutilisation des données pourraient permettre d'encourager et d'accompagner la mise en œuvre de chacune de ces priorités.

Au-delà des marchés de l'information publique aujourd'hui bien identifiés (information légale, géographique, etc.) et de l'économie numérique, les nouvelles applications de la réutilisation des données publiques pourraient donner lieu au développement de nouveaux marchés et à la création d'entreprises innovantes. Faciliter l'accès aux données publiques et leur réutilisation par les startups est donc un enjeu du développement de l'entrepreneuriat de haute technologie, et de compétitivité des PME françaises.

Le développement durable et la résilience de nos sociétés peuvent aussi profiter largement d'une réutilisation plus large des données publiques. Par exemple, la disponibilité d'informations intermodales riches et mises à jour accroît sensiblement l'usage des transports publics. Le développement des énergies propres est conditionné à l'accès à des données géophysiques de qualité, comme par exemple des mesures d'ensoleillement précises. Le partage des données publiques entre services d'urgences peut aussi faciliter la coordination des réponses aux catastrophes naturelles.

Enfin, une collecte et un partage plus large des données peut favoriser la recherche scientifique et le développement de nouveaux produits innovants. La recherche en biologie synthétique ou sur l'impact médical des substances chimiques, par exemple, repose largement sur le partage de résultats d'expérience. L'investissement des puissances publiques pour améliorer la mutualisation et le partage de ces données est un enjeu de compétitivité.

Les deux prochaines sections explorent plus en détail les leviers d'action à travers lesquels la réutilisation des données publiques

⁸³ New York Times, 23 avril 2011. « Show Us The Data (It's Ours After All) ». Richard H. Thaler. <http://www.nytimes.com/2011/04/24/business/24view.htm>

⁸⁴ Data Portability Project. <http://portabilitypolicy.org/index.html>

⁸⁵ Alain Juppé, Michel Rocard, 2009, « Investir pour l'Avenir : Priorités stratégiques d'investissement et emprunt national ». http://www.commission-investissement.fr/IMG/pdf/Rapport_191109.pdf

pourrait contribuer au développement durable, et encourager la croissance des PME françaises.

4.3. Exemple : les données publiques au service du développement durable

Les applications construites à partir des données publiques pourraient contribuer de multiples manières à encourager le développement durable sous ses différentes facettes.

(i) Développement des énergies propres

La production d'énergie est soumise à la disponibilité de sources de données fiables permettant de calculer et justifier les investissements engagés. C'est particulièrement vrai dans le domaine des énergies propres, et en particulier de la production d'électricité d'origine renouvelable, dont la rentabilité face à l'électricité d'origine fossile est l'enjeu crucial⁸⁶.

Un chargé de projets solaires au sein d'un grand groupe énergétique Français (anonyme et non cité parmi les intervenants) nous a signalé les carences du système de données publiques françaises sur l'ensoleillement moyen du territoire, facteur principal de la rentabilité d'un projet de génération photovoltaïque.

Il nous a confié avoir été contraint d'utiliser des données issues de la recherche américaine⁸⁷, pourtant bien moins précises que les données françaises sur le territoire national⁸⁸. Ces dernières n'étaient tout simplement pas exploitables dans le cadre de leur algorithme d'estimation de rentabilité d'un projet, car diffusées sous formes d'images numériques plutôt que de données brutes, et donc rendues inutilement complexes à réutiliser pour une simple question d'ergonomie du format utilisé.

Ce cas illustre au passage une dimension essentielle de la réutilisation des données publiques : il ne suffit pas d'avoir des données de qualité pour qu'elles soient utiles, encore faut-il qu'elles soient facilement exploitables

par des machines et programmes informatiques, c'est-à-dire distribuées dans un format ouvert et structuré.

(ii) Offre de transports publics multimodale

Les opérateurs de transports publics consacrent des efforts importants à l'information disponible à quai, en ligne et sur mobile. Cependant, nombre d'opérateurs et d'Autorités Organisatrices de Transports conservent encore très largement une approche en silo de leurs données, ce qui n'est pas favorable à l'innovation ouverte que la réutilisation des données publiques permettrait de mettre en œuvre. Concrètement, cela signifie que les développeurs indépendants ou les startups qui pourraient se développer ne peuvent pas réutiliser les données de fonctionnement de réseaux de transports afin de proposer aux consommateurs des solutions alternatives aux sites et applications mobiles officiels, parfois difficiles d'accès ou affectés de carences ergonomiques ou technologiques.

La question de l'accès à ces données est compliquée par une certaine incertitude juridique concernant leur statut. La société CanalTP⁸⁹, filiale de Kéolis qui propose des solutions d'interconnexion logicielle (middleware) entre systèmes de gestion de flotte des opérateurs de transports et système de calcul d'itinéraires, édite le service en ligne NAViTiA qui dessert plus de 10 millions de voyageurs par mois. Elle explique que les différents acteurs publics expriment parfois une crainte dans la diffusion des données vers le marché individuel, car ils redoutent le piratage informatique et l'exploitation de données confidentielles qui pourraient fragiliser leur système d'information voyageurs.

Au-delà de ces craintes, les collectivités que nous avons rencontrées expliquent que le statut légal et l'accès technique aux données peut être particulièrement complexe dans le domaine du transport. Des droits de propriété croisés génèrent une incertitude sur le statut des données générées dans le cadre de l'exploitation des délégations de service public. Leur propriété est souvent revendiquée à la fois par les autorités organisatrices et les opérateurs.

Des droits de propriété clairs permettraient aux différents interlocuteurs de trouver des modalités d'accord, et *in fine* d'intégrer plus largement l'offre de transports publics à travers le territoire français aux grands sites

⁸⁶ Tester, J.W., Drake, E.M., Driscoll, M.J., Golay, M.W. and Peters, W.A. 2005. *Sustainable Energy : Choosing Among Options*. MIT Press, Cambridge, MA. <http://mitpress.mit.edu/catalog/item/default.asp?tid=10566&tttype=2>

⁸⁷ U.S. National Renewable Energy Lab. International Solar Resource Data. http://www.nrel.gov/csp/troughnet/solar_data.html

⁸⁸ European Union Joint Research Center. Photovoltaic Geographical Information System Interactive Maps. <http://re.jrc.ec.europa.eu/pvgis/apps4/pvest.php#>

⁸⁹ CanalTP. <http://www.canaltp.fr/>

d'information géographique en ligne. L'ouverture à la réutilisation et la clarification du statut des données produites dans ce cadre, notamment à travers les contrats d'objectifs et de moyens passés entre autorités organisatrices et opérateurs, entraîneraient une meilleure information multimodale et favoriseraient l'utilisation des transports en commun par le public, préférable à l'usage de la voiture.

L'information intermodale sur les transports en commun ouvre un champ nouveau pour le développement d'applications mobiles, tels le projet Locomote de la société Isokron évoqué plus haut. Du fait des limitations extrêmement strictes d'accès aux données de la RATP et de l'opérateur de la flotte Vélib, le projet devra cependant se limiter largement dans un premier temps à des temps de trajets moyens dans Paris. Cela réduit l'intérêt de l'application, et donc ses chances d'atteindre une échelle qui lui permettrait de faire bénéficier le grand public d'une consommation plus intelligente des transports publics.

Au-delà des seuls transports publics, le thème des mobilités douces⁹⁰ et des villes digitales⁹¹ est un sujet de recherche de pointe, notamment au sein des universités américaines (MIT Media Lab⁹²) et des grandes écoles françaises (École des Ponts ParisTech⁹³). Le pôle de compétitivité sur la ville durable Advancity⁹⁴ peut jouer un rôle important dans le développement et la commercialisation des produits de ces recherches, dont beaucoup bénéficieraient de la mise à disposition la plus large des données publiques.

(iii) Favoriser les consommations intelligentes et l'efficacité énergétique

À grande échelle, les consommations intelligentes passent aussi par la capacité de calculer les choix les plus efficaces pour les consommateurs ou les investissements des sociétés.

Dans le domaine de la consommation énergétique, des solutions de calcul de

l'efficacité énergétique d'une maison ont vu le jour, par exemple au Danemark⁹⁵, grâce à la mise à disposition de données environnementales.

Pour l'électricité en particulier, les réseaux intelligents (*smart grids*) reposent en grande partie sur une communication approfondie entre l'ensemble des nœuds du réseau. Aux États-Unis, l'équipe de Data.gov a par exemple permis la création d'un site d'information énergétique ouvert, Open Energy Information⁹⁶, qui propose des *mashups* didactiques⁹⁷ basés sur l'information énergétique en temps réel.

Le site de RTE (la société de transport d'électricité française) fournit aussi une visualisation⁹⁸ en temps réel de la consommation électrique en France, de la répartition de la production par source énergétique, ainsi que des émissions de CO₂ qui y sont liées. Ce type d'applications permettra à terme d'adapter en temps réel l'état du réseau électrique.

Dans le domaine du numérique, le mouvement du « *green IT* », ou technologies informatiques vertes, cherche à réduire la part considérable de la consommation énergétique qui revient aux centres de calcul, soulignant que les applications des données publiques peuvent elles-mêmes être hébergées et mises en œuvre d'une façon respectueuse de l'environnement.⁹⁹

(iv) Aménagement durable et surveillance de l'environnement

Le développement durable passe aussi par l'aménagement du territoire. En Australie, en application d'une loi de 2007 sur l'information sur l'eau, un sujet éminemment sensible dans un pays aride, le *Bureau of Meteorology* a annoncé l'ouverture d'un système intégré d'information sur l'eau au niveau national¹⁰⁰,

⁹⁰ Fioretti, M. 23 octobre 2009. « Walkability : Check It Before Choosing Your Next Home » <http://stop.zona-m.net/2009/10/walkability-check-it-before-choosing-your-next-home/>

⁹¹ Ratti, C. 2 octobre 2009. « Digital Cities : 'Sense-able' urban design. » *Wired UK Magazine*. <http://www.wired.co.uk/magazine/archive/2009/11/features/digital-cities-sense-able-urban-design>

⁹² Senseable City Lab, MIT Media Lab. <http://senseable.mit.edu/>

⁹³ Antoine Picon, chercheur associé au LATTs. <http://www.enpc.fr/enseignements/Picon/index.html>

⁹⁴ Pôle de compétitivité Advancity. <http://www.advancity.eu/>

⁹⁵ HusetsWeb Danemark. <http://www.husetseb.dk/about>

⁹⁶ Open Energy Information. http://en.openei.org/wiki/Main_Page

⁹⁷ Open Energy Information Mashathon 2010. <http://en.openei.org/apps/mashathon2010/>

⁹⁸ ÉCO2mix : Consommation, Production et Contenu CO₂ de l'Électricité Française.

<http://www.rte-france.com/fr/developpement-durable/maitriser-sa-consommation-electrique/eco2mix-consommation-production-et-contenu-co2-de-l-electricite-francaise>

⁹⁹ Pascal Faure, Vice Président du Conseil Général de l'Industrie, de l'Énergie et des Technologies, explique l'importance du « *green IT* » aux assises du numérique 2009. http://www.dailymotion.com/video/xaov8c_pascal-faure-assises-du-numerique-2_news

¹⁰⁰ <http://creativecommons.org.au/weblog/entry/269>

agrégeant des centaines de ressources issues des administrations nationales et locales.

Face aux déchaînements malheureux de la nature, un système robuste et complet de données peut jouer un rôle de prévention (limites de zones inondables, prévisions météorologiques) ou aider immensément l'effort de secours aux populations sinistrées, dans le contexte de pays développés comme dans des pays en voie de développement¹⁰¹.

L'accès aux données publiques et les nouveaux services que leur réutilisation pourrait permettre de créer sont donc un élément d'encouragement et de soutien substantiel au développement durable de notre économie.

4.4. Exemple : Un avantage comparatif pour le développement des PME

La France possède un avantage comparatif dans la production de données publiques. L'État ayant historiquement un rôle fort dans l'économie de la nation, il a développé une capacité de production de données de qualité dans de nombreux domaines de l'économie et de l'action publique.

De plus, certaines données françaises répondent à une demande commerciale forte à l'international. Dans le domaine de la culture, la reconnaissance dont bénéficient les œuvres françaises se traduit par des fréquentations en ligne très importantes. Par exemple, avec plus de 2 millions de téléchargements dont 60 % en provenance des États-Unis,¹⁰² l'application « Musée du Louvre » est la plus téléchargée parmi les lauréats de l'appel à projet Proxima Mobile.

Développer cet avantage comparatif est particulièrement intéressant pour deux raisons. Il permet d'une part, en soutenant le développement d'entreprises qui nécessitent une expertise locale forte, de créer des emplois difficiles à délocaliser¹⁰³, car tributaires d'une connaissance fine du contexte français, culturel en l'espèce.

D'autre part, il permet de soutenir un secteur, l'économie numérique, dont les entreprises à succès ont structurellement

tendance à travers le monde à croître au-delà de la catégorie des Petites et Moyennes Entreprises, pour devenir des Entreprises de Taille Intermédiaire.¹⁰⁴

L'économie française pêche tout particulièrement dans cette catégorie, et toute opportunité d'améliorer cette situation doit être agressivement mise en œuvre.

Comme d'autres secteurs stratégiques pour l'avenir de notre économie, le développement durable et l'encouragement à la croissance des PME sont donc une priorité de l'action publique, en faveur de laquelle l'ouverture des données publiques et leur réutilisation peuvent jouer un rôle décisif.

5. Un nouveau mode d'action publique

Les implications de l'ouverture des données publiques à la réutilisation s'apprécient mieux à la lumière de l'histoire économique des technologies de l'information.

Deux analogies techniques sont particulièrement riches d'enseignements pour mieux comprendre les nouveaux modes d'action publique à l'ère numérique : la notion d'« infrastructure informationnelle », transposée immatérielle de l'infrastructure bâtie du monde réel, d'une part ; et le paradigme de l'État repensé comme une plateforme logicielle (« *government as a platform* »), d'autre part.

5.1. Une infrastructure cognitive pour l'économie de l'information

La mise en ligne des données publiques n'est pas qu'un service fourni aux administrés parmi d'autres : elle constitue pour l'économie de l'information une forme d'infrastructure, « informationnelle » ou « cognitive¹⁰⁵ ».

Rufus Pollock, économiste à l'Université de Cambridge et membre fondateur de l'*Open*

¹⁰¹ OWNI.fr. 12 décembre 2010. « Partage des Données pour Mieux Réagir Face aux Catastrophes Naturelles ». <http://owni.fr/2010/12/12/partager-des-donnees-pour-mieux-reagir-face-aux-catastrophes-naturelles/>

¹⁰² Source Proxima Mobile et entretien avec Bernard Benhamou les 8 octobre, 16 novembre et 14 décembre 2010.

¹⁰³ Entretien avec Bernard Benhamou les 8 octobre, 16 novembre et 14 décembre 2010.

¹⁰⁴ Benhamou, B., 2010, « Eléments pour une stratégie numérique européenne », Délégation aux Usages de l'Internet, 15 mars 2010.

¹⁰⁵ Henri Verdier. 24 février 2009. « Stimulus package : les américains construisent une infrastructure cognitive ». <http://www.henriverdier.com/2009/11/stimulus-package-les-americains.html>

Knowledge Foundation, en propose une définition métaphorique¹⁰⁶ :

« L'importance potentielle des données publiques se mesure (...) à l'aune d'une analogie simple mais forte de sens : de même que la mise en place d'une infrastructure matérielle fondamentale (réseaux d'électricité, de transports, de télécommunications) est essentielle à l'économie traditionnelle, la mise en place d'une « infrastructure » informationnelle et cognitive fondamentale (ensembles de données dans les secteurs critiques : géographie, météorologie, transport, etc.) est essentielle à l'économie de l'« information »¹⁰⁷.

Dans ce sens, les données publiques sont cruciales comme socle d'un nouveau mode d'action publique, et comme infrastructure fondamentale pour le développement de la société numérique.

5.2. La régulation des monopoles informationnels

À la mission de fournisseur de données s'ajoute désormais pour l'État celle de régulateur d'un genre nouveau, en charge d'un champ de responsabilité analogue et symétrique à celui qu'il assure dans le domaine de la concurrence.

Il s'agit de s'assurer que les données, ressources essentielles au bon fonctionnement de l'économie numérique, ne sont pas accaparées, ni par le secteur privé (se reporter à ce sujet au scénario « Capture » du chapitre VIII), ni par la puissance publique dans l'hypothèse où la réutilisation des données se ferait dans des conditions légales et tarifaires discriminatoires.¹⁰⁸

La révolution industrielle avait conduit les États à développer des politiques et un droit de la concurrence, comme les lois *antitrust*

établies au début du XX^e siècle en réponse au développement des premiers monopoles naturels de réseaux : réseau de raffinage avec la *Standard Oil*, réseaux de transports avec les chemins de fer, mais aussi déjà réseaux d'information avec le télégramme et téléphone.

La révolution informatique doit conduire l'État à assurer une forme de régulation dans la sphère de l'information. La mise à disposition de données permet de donner corps à de nouvelles offres de services et de nouveaux marchés, et est donc un outil essentiel d'action de l'État en faveur de l'économie numérique.

Cette évolution, en réponse aux ruptures technologiques et sociologiques fortes issues du numérique, est finalement assez naturelle. Le rôle de garant de l'efficacité sociale est constitutif de l'action publique et de l'administration, qui doit se préoccuper certes d'efficacité interne, mais aussi et surtout d'efficacité sociale¹⁰⁹.

5.3. Le modèle de l'État comme plateforme numérique (« *Government as a Platform* »)

L'un des enjeux majeurs de la troisième révolution industrielle pour l'administration est de s'approprier les modèles économiques et les formes d'action caractéristiques du numérique.

« [Nous aimerions voir] des politiques publiques [adopter une certaine] sérénité dans leur rapport au numérique, ni fascinant ni menaçant, simplement un outil précieux, dont les logiques profondes recèlent de nouveaux mécanismes de pensée et d'action, et qu'il faut savoir utiliser à sa juste place¹¹⁰ » explique Henri Verdier, Président du pôle de compétitivité Cap Digital.

Pour ce faire, l'État peut s'inspirer des leçons tirées de l'émergence des grandes plateformes informatiques, caractéristiques de la révolution numérique.

On entend par plateforme tout système informatique qui agit comme un socle, permettant le développement d'extensions et d'applications tierces qui proposent de nouveaux services en son sein.

Internet, système décentralisé d'échange de documents sur lequel des suites logicielles entières ont été portées et même inventées, est

¹⁰⁶ Rufus Pollock, 2 décembre 2008. « The Economics of Public Sector Information. » University of Cambridge Department of Economics. http://www.rufuspollock.org/economics/papers/economics_of_psi.pdf

¹⁰⁷ Texte original (traduction des auteurs) :
“The potential importance of (public sector) information can also be gauged from a simple but significant analogy: just as the supply of basic physical infrastructure (power, transport, telecommunications) is essential to the traditional economy, so the supply of basic information ‘infrastructure’ (core datasets in the major areas of geography, weather, transport etc) is essential to the ‘information’ economy.”

¹⁰⁸ Rufus Pollock, 2008, « The Economics of Public Sector Information. »

¹⁰⁹ Jean-Claude Thoenig, 4 septembre 2010, conférence donnée aux élèves du corps des Ponts, Eaux et Forêts.

¹¹⁰ Henri Verdier, 24 février 2009, « Stimulus package : les américains construisent une infrastructure cognitive ». <http://www.henriverdier.com/2009/11/stimulus-package-les-americains.html>

une plateforme. Le système d'exploitation Windows, couplé au standard informatique IBM/PC, mais aussi plus récemment, les couples Android/Android Market de Google et iPhone/App Store d'Apple, à partir desquels des milliards d'applications mobiles ont été téléchargées, sont d'autres exemples.

Tim O'Reilly, philosophe, éditeur d'ouvrages techniques à succès, investisseur en capital-risque et expert reconnu de la société de l'information, tire¹¹¹ sept leçons de l'histoire de l'informatique, dont les politiques d'e-administration au niveau territorial et national pourraient s'inspirer.

(i) Les standards ouverts génèrent l'innovation et la croissance

Les standards ouverts (structuration des données selon des normes de format communes et non propriétaires) permettent l'interopérabilité. Ils lubrifient figurativement les rouages de l'échange numérique, de la même manière qu'à la Révolution l'instauration du système métrique (exemple éminent d'infrastructure cognitive) a permis de sécuriser et développer le commerce au sein de la République naissante.

(ii) Construire un système simple et le laisser évoluer

La propension des maîtres d'ouvrage de systèmes d'information dans le secteur public à consacrer des mois voire des années entières à spécifier l'ensemble des fonctionnalités de leurs projets (en partie due à la lourdeur des processus de marchés publics) retarde chroniquement le développement de nouvelles solutions, forçant les agents à fonctionner dans l'entre deux et le provisoire, avec des solutions techniques en retard de phase sur les attentes des citoyens.

Il serait bien plus judicieux d'adopter l'approche incrémentale popularisée par les méthodes de développement agile, prévalant aujourd'hui en Silicon Valley et dans les équipes d'ingénieurs travaillant à l'état de l'art. L'administration doit prendre en compte la valeur d'option¹¹² inhérente à la capacité d'évolution rapide que cette approche rend possible, et la possibilité qu'elle offre d'incorporer plus facilement les retours de futurs utilisateurs confrontés aux premières versions de test, afin d'améliorer

substantiellement le logiciel jusqu'à un stade avancé du développement.

(iii) Concevoir et planifier en se préoccupant de la participation

Une leçon majeure du Web 2.0 est que la notion de participation et d'interaction sociale autour et au sein d'un service en ligne ne saurait être ajoutée après coup en extension d'un projet préexistant, mais doit au contraire être intégrée dès sa conception.

Dans le domaine de la mise en ligne des données, il est important d'intégrer la capacité à commenter, référencer et échanger de l'information autour des jeux de données dès les premières phases des projets de portails *Open Data*, et de prendre en compte les attentes de la communauté des réutilisateurs.

(iv) Apprendre en observant les utilisateurs les plus en pointe

Eric von Hippel, professeur au MIT, parle d'utilisateurs « en pointe » (*lead users*) pour désigner les quelques utilisateurs les plus avancés d'un service en ligne, qui pointent souvent vers ses futures évolutions possibles et révèlent les idées de fonctionnalités dont l'intérêt n'est pas encore apparent à tous.

L'exemple de l'API Google Map, facteur majeur du succès de ce produit, mais qui n'a été exposée officiellement au public qu'après avoir observé un utilisateur qui avait développé son propre système de placement d'objets sur les cartes Google, est édifiant à cet égard.

Au lieu de lui intimer l'ordre de cesser cette utilisation imprévue et non conforme de l'outil technique, l'entreprise embaucha l'ingénieur en question pour développer une version commerciale de ce système de placement d'objets géographiques sur carte en ligne. Sa qualité et sa facilité d'utilisation sont à l'origine du succès de Google Maps, hégémonique dans le domaine de la cartographie en ligne.

L'administration doit de même s'inspirer des utilisations originales qui seront faites de ses données, les encourager, et même les reprendre à son compte.

(v) Collecter des données d'usage pour mieux profiter de l'ouverture des données

Il est important de mesurer la fréquentation et l'utilisation qui est faite des sites de l'administration (c'est par ailleurs un type de donnée publique qui est encore rarement accessible).

¹¹¹ Tim O'Reilly, 2010, « Government As A Platform. » Chap. 1 dans *Open Government*. O'Reilly éditions.

<http://opengovernment.labs.oreilly.com/>

¹¹² http://fr.wikipedia.org/wiki/Analyse_par_les_options_r%C3%A9elles

Le comportement des utilisateurs sur un site ou une application est de fait un type implicite de participation à l'évolution d'un système, qu'il faut utiliser à bon escient pour mieux répondre aux attentes implicites et aux besoins latents.

(vi) Abaisser les barrières à l'expérimentation

La capacité pour les acteurs de la société et de l'économie numérique d'expérimenter facilement est une condition nécessaire et d'importance cruciale pour l'émergence de nouvelles applications originales et utiles aux citoyens.

L'État et les producteurs de données publiques doivent donc s'attacher à abaisser le plus possible les barrières et les cloisons artificielles qui pourraient s'opposer au développement de services innovants à partir des données publiques.

Il lui faudra notamment clarifier les conditions légales de réutilisation, mettre en place les modèles économiques permettant la réutilisation la plus large possible, et encourager l'expérimentation par les entrepreneurs, les startups et les acteurs innovants de l'économie numérique.

(vii) Diriger en montrant l'exemple.

En conclusion, il s'agit pour l'État de s'approprier les leçons des évolutions technologiques et économiques de nos sociétés, et de devenir un acteur de la diffusion de ces nouveaux modes de pensée au sein de la société, en menant le mouvement par l'exemple.

Les données publiques sont un outil de la démocratie exemplaire et contribuent au renforcement de la transparence de l'action de l'État. Leur réutilisation soutient l'innovation, encourage la recherche et favorise l'émergence de l'Internet de demain. La communauté des développeurs et des entrepreneurs peut les réutiliser pour inventer de nouveaux usages et créer de nouveaux services aux citoyens.

Elles forment une infrastructure informationnelle fondamentale pour la société numérique. Le rôle de garant de l'efficacité sociale de leur réutilisation est une extension naturelle du rôle de la puissance publique, qu'elle doit saisir avec ambition et détermination.

Partie 2. État des lieux de la réutilisation des données publiques en France

Chapitre IV. Le cadre de la réutilisation des données publiques en France

L'*Open Data* est un mouvement global, qui trouve une résonance forte au sein des communautés de développeurs à travers le monde. Il s'est plus particulièrement développé dans les pays anglo-saxons, notamment à la suite des initiatives data.gov aux États-Unis, et data.gov.uk en Grande-Bretagne.

La loi française, en transcrivant en 2005 la directive européenne « PSI » de 2003 sur les informations du secteur public, a instauré une obligation d'ouverture et un droit à la réutilisation des données publiques. Ce chapitre dresse un état des lieux du cadre juridique et réglementaire, puis des acteurs institutionnels de l'ouverture et de la réutilisation des données publiques en France.

1. Le cadre juridique pose en principe la liberté de réutilisation

1.1. Principe de réutilisation

Le droit français établit le principe de la réutilisation libre et gratuite des données publiques. Les conditions d'accès à un document administratif, ou de réutilisation d'une information publique, aux sens définis par la loi, peuvent faire l'objet d'une requête auprès de la Commission d'Accès aux Documents Administratifs (CADA).

La possibilité reste offerte aux administrations de soumettre à une redevance le droit de réutilisation des jeux de données pour lesquels cela s'avérerait nécessaire. Toute nouvelle redevance sera cependant désormais arrêtée par décret, donc encadrée par le Premier ministre¹¹³.

1.2. Cadre juridique européen

Au niveau européen, la mise à disposition des données publiques et le droit à leur réutilisation sont régis par trois sources législatives.

- La directive européenne 90/313/CEE concernant l'accès aux données environnementales de 1990 qui a abouti sur la Convention d'Aarhus en 2002, transcrite en 2003 dans la directive européenne 2003/4/CE et en 2005 en droit français.
- La directive européenne PSI n° 2003/98/CE du 17 novembre 2003 concernant la réutilisation des informations du secteur public, transposée en droit français par l'ordonnance n° 2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques. Cette directive est en cours de réexamen pour la deuxième fois.
- La directive européenne Inspire n° 2007/2/CE du 14 mars 2007 établissant une infrastructure d'information géographique dans la Communauté européenne, transposée en droit français par l'ordonnance n° 2010-1232 du 21 octobre 2010 dans les articles L. 127-1 à L. 127-10 du code de l'environnement introduits.

Les principes posés par le droit européen sont que les données publiques doivent être accessibles et réutilisables, à des fins commerciales ou non, d'une manière non discriminatoire et non exclusive à des tarifs qui n'excèdent pas leur coût de mise à disposition et éventuellement de production.

¹¹³ Regards sur le Numérique, 17 mars 2011, « Data.gouv.fr : La France va se lancer en version bêta avant fin 2011 ». <http://www.rslnmag.fr/blog/2011/3/17/data-gouv-fr-la-france-va-se-lancer-en-version-beta-avant-fin-2011/>

Directive	Environnement	PSI	Inspire
Année	1990	2003	2007
Données concernées	Données environnementales	Toute donnée ou document public diffusé	Données spatiales
Données exclues	Documents non finis, statistiques ou délibération de l'État	Documents non finis, Propriété intellectuelle d'un tiers, statistiques ou délibération de l'État	Données personnelles, protection de l'environnement, document non fini ou couvert par un secret, propriété intellectuelle d'un tiers
Par qui	Tout organisme, public ou privé, ayant une mission de service public.	État hors services commerciaux et hors services culturels	Toute autorité publique détentrice de données géographiques ou prestataire mandaté par une autorité publique
Comment	Cataloguer les ressources mais sans métadonnées précises	Tarifs bornés par les coûts	Obligation de créer : - métadonnées, services de recherche et de visualisation, gratuitement. - des services de téléchargement et de transformation.
Limites	La réutilisation commerciale n'est pas traitée.	La réutilisation commerciale est laissée à la discrétion des États.	Le plan cadastral reste la référence du bâti et du parcellaire

Depuis, différentes modifications ont été proposées, comme par exemple « d'étendre le champ d'application de la directive de manière à obliger les organismes publics à autoriser la réutilisation, d'imposer la tarification fondée sur les coûts marginaux, d'établir des listes nationales des ressources et d'obliger les États membres à présenter des rapports annuels à la Commission »¹¹⁴. Des précisions ont également été demandées « sur certains termes de la directive [...] trop ambigus, tels que les définitions des documents, des missions de service public, de la tarification fondée sur les coûts marginaux ou du retour sur investissement raisonnable. »

La révision de la directive PSI apparaissait comme prématurée en 2008 lorsque la Commission en a étudié l'opportunité.

« La majorité des États membres ayant participé à la consultation estiment qu'il est prématuré de modifier la directive et préfèrent qu'un délai supplémentaire soit accordé pour permettre sa totale mise en œuvre. Les organismes du secteur public jugent que le cadre actuel fourni par la directive donne satisfaction. »

Cependant, la Commission procède à l'heure actuelle à un réexamen qui pourrait mener à une initiative législative d'ici la fin de l'année, et à une révision éventuelle de la Directive PSI.

¹¹⁴ Commission Européenne, 2009, « Réutilisation des informations du secteur public: Réexamen de la directive 2003/98/CE ». 7 mai 2009.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0212:FIN:FR:PDF>

1.3. Cadre juridique français

Le principe de réutilisation des données publiques peut se trouver un fondement constitutionnel dans les articles XIV et XV de

la déclaration des droits de l'homme et du citoyen de 1789¹¹⁵ :

Article XIV : Tous les Citoyens ont le droit de constater, par eux-mêmes ou par leurs Représentants, la nécessité de la contribution publique, de la consentir librement, d'en suivre l'emploi et d'en déterminer la quotité, l'assiette, le recouvrement et la durée.

Article XV : La Société a le droit de demander compte à tout Agent public de son administration.

La réutilisation des données publiques est désormais un droit régi par la loi et les dispositions réglementaires, et d'éventuelles conditions qui seraient spécifiques au jeu de données en question, précisées dans une licence détaillant le cadre de réutilisation.

(i) De la loi de 1978 sur l'accès aux documents administratifs à la transposition de la directive

La mise à disposition des données publiques est désormais régie par l'ordonnance du 6 juin 2005, transposant la directive 2003/98 du 17 novembre 2003, et modifiant la loi « CADA » n° 78-753 du 17 juillet 1978.

Le législateur a tenu à souligner l'importance du sujet puisque c'est le titre même de la loi de 1978 qui a été modifié, « la liberté d'accès aux documents administratifs » devenant désormais « la liberté d'accès aux documents administratifs et la réutilisation des informations publiques ».

Deux autres dispositions ont complété cet ensemble réglementaire. Un décret n° 2005-1755 du 30 décembre 2005 a été pris en application de l'ordonnance afin d'en préciser les modalités d'application, et notamment de permettre une réutilisation à des fins commerciales autant que non commerciales. Enfin, une circulaire du Premier Ministre n° 5156/SG du 29 mai 2006 a apporté un certain nombre de précisions.

Le principe général posé par ces textes est clair : c'est celui de la réutilisation libre et gratuite des informations publiques. Au sens de la législation française, les informations publiques sont les informations contenues dans les documents administratifs tels que définis par la loi « CADA », produites ou reçues par un acteur public ou privé dans l'exercice d'une mission de service public, hors service public industriel ou commercial.

Selon ces textes, les producteurs de données publiques sont donc l'État, les collectivités territoriales (Régions, Départements et Communes), et les autres personnes morales de droit public ou privé chargées d'une mission de service public qui ne soit pas à caractère industriel ou commercial.

Les établissements d'enseignement et de recherche et les établissements, organismes ou services culturels bénéficient par ailleurs d'un régime dérogatoire quant aux modalités et à l'application de ce droit de réutilisation des données publiques.

Les données publiques elles-mêmes sont définies de façon large puisqu'il s'agit des « informations figurant dans des documents produits ou reçus par les administrations mentionnées à l'article 1^{er}, quel que soit le support. » À l'origine, la loi de 1978 visait les documents administratifs tels que les rapports, notes, études, à l'exception des documents destinés à nourrir les réflexions des autorités gouvernementales avant que celles-ci n'arrêtent leur décision (secret du délibéré de l'État)¹¹⁶.

Elle couvre désormais les données géographiques, les statistiques, les mesures et relevés scientifiques, les budgets détaillés, etc. lorsque ces données ne sont pas couvertes par un secret déterminé.

(ii) Protection de la vie privée

Hormis les exceptions mentionnées ci-dessus, la principale limite posée à la réutilisation des données publiques concerne les informations personnelles, et celles dont la mise à disposition serait susceptible d'affecter la vie privée de citoyens.

Sur ce point, la primauté de protection des données personnelles est garantie, tout en préservant le principe de la liberté de réutilisation des données publiques. Leur réutilisation n'est possible que dans deux hypothèses : si la personne intéressée y a consenti, ou si le producteur des données est en mesure de les rendre anonymes. Si ces deux conditions étaient impossibles à réunir, la loi précise que le législateur a toujours la possibilité d'accorder une autorisation spécifique au cas par cas.

¹¹⁵ <http://www.assemblee-nationale.fr/histoire/dudh/1789.asp>

¹¹⁶ Pour plus d'informations, consulter le site de la CADA : <http://www.cada.fr/fr/guide/center4b1.htm>

(iii) Pas d'exclusivité

La dernière limite imposée par le législateur prend tout son sens avec le passage à l'ère numérique, notamment au regard des principes de neutralité d'Internet et du développement d'un marché concurrentiel dans le secteur des communications. Aucune administration ne peut accorder de contrat d'exclusivité pour la réutilisation de ses données, sauf si cette exclusivité est nécessaire à l'exercice de sa mission de service public. Dans tous les cas, cette exclusivité devra alors être révisée tous les trois ans.

(iv) Redevance et licence

Le législateur a également prévu l'hypothèse d'une administration qui souhaiterait soumettre ses données à redevance. Considérant cette modalité comme légitime, il ne l'a soumise qu'à la seule condition que cette mise à disposition soit alors soumise à une licence publique, proposée à tous, et non discriminatoire, et que le montant de la redevance ne dépasse pas un retour raisonnable sur investissement pour l'institution productrice.

Dans cette hypothèse de mise à disposition payante des données publiques, une licence est obligatoire afin de clarifier l'étendue des droits conférés aux réutilisateurs. Lorsqu'il n'y a pas de redevance, une licence n'est pas obligatoire et l'administration choisit souvent un simple rappel des conditions de réutilisation fixées par la loi.

Cependant, l'administration peut aussi diffuser ses données selon des licences qu'elle établirait elle-même afin de garantir une meilleure information aux réutilisateurs, et d'assurer leur sécurité juridique. En effet, le problème est souvent soulevé, notamment par les jeunes sociétés innovantes qui doivent se justifier sur ce point dans leurs dossiers de financement.

Dans la mesure où les données sont librement réutilisables, elles peuvent alors être intégrées dans tous les projets sous licences libres compatibles (type GPL, MIT, *Creative Commons* etc.).

Finalement, il est également possible pour l'administration de proposer une solution mixte, avec simultanément un accès libre et gratuit aux données et un accès payant mais qui impliquerait des services supplémentaires. C'est par exemple le cas de la licence fixée pour la réutilisation en masse des données légales et réglementaires via l'API du site

Legifrance, qui est gratuite en deçà d'un certain seuil de volume et payante au-delà.

(v) Le régime dérogatoire des données scientifiques et culturelles.

En ce qui concerne les données publiques d'enseignement, de recherche et culturelles, c'est-à-dire notamment toutes les données dépendant du ministère de la culture, dont les archives, il convient de rappeler que les dispositions de la loi n° 78-753 du 17 juillet 1978 leur concèdent un régime dérogatoire.

En effet, selon l'article 11 de cette loi, les conditions de réutilisation des informations contenues dans des documents élaborés ou détenus par des établissements, organismes ou services d'enseignement, de recherche ou du domaine de la culture peuvent être librement fixées par ces derniers¹¹⁷.

Une difficulté vient du fait que les données culturelles sont très fréquemment grevées de droits de propriété intellectuelle envers des tiers (artiste, photographe, etc.), ce qui a pour effet d'en compliquer le droit à réutilisation.

Ce régime dérogatoire découle des exceptions consenties dans les textes européens : la directive PSI exclue de son champ d'application les organismes culturels, d'enseignement et de recherche, ainsi que les radiodiffuseurs publics.

Néanmoins au cours de son réexamen en 2008, cette question était revenue à l'ordre du jour. Un « comité des sages », nommé par la Commission européenne sur la mise en ligne du patrimoine culturel européen, a remis début 2011 des conclusions¹¹⁸ qui vont fortement dans le sens d'une plus large ouverture à la réutilisation des données culturelles. Le sujet pourrait donc évoluer à l'occasion d'une éventuelle révision de la directive PSI.

¹¹⁷Une démarche exploratoire visant à généraliser la réutilisation des données a été mise en œuvre : Ministère de la Culture, 2008, « *Partager notre patrimoine culturel : Propositions pour une charte de la diffusion et de la réutilisation des données publiques culturelles numériques* ».

http://www.culture.gouv.fr/mcc/content/download/5331/35792/version/1/file/CHARTRE_BOL_21082009noirblancVAL-1.pdf

¹¹⁸Reflection group on bringing Europe's cultural heritage online, 2011, "The New Renaissance: Report of the 'Comité des sages'".

http://ec.europa.eu/information_society/activities/digital_libraries/doc/refgroup/final_report_cds.pdf

2. Les acteurs de la réutilisation de données publiques

Le processus de réutilisation des données publiques implique trois types d'acteurs : des producteurs, des réutilisateurs, et des instances publiques de contrôle, de pilotage et de mise en œuvre de cette politique publique.

2.1. Les producteurs de données

(i) Les acteurs publics

Les administrations (centrales, ou déconcentrées), les collectivités territoriales et les établissements publics à caractère administratif (EPA) produisent et reçoivent une quantité considérable de données publiques, puisque ces dernières représentent par définition l'ensemble de la « matière première » à partir de laquelle l'administration travaille.

Ces données sont produites dans un contexte de « monopole naturel » du fait de la position de l'administration pour réaliser et suivre l'exécution des politiques publiques ou pour la collecte d'indicateurs de pilotage. Elles sont uniques et ont souvent une forte valeur. Malgré les difficultés liées à la qualité de données inhérentes à tout système d'information de grande envergure, ces données sont généralement fiables, précises, complètes et de qualité.

Les fonctionnaires et agents de ces institutions, qui connaissent très bien les métiers de leur organisation, ont un rôle important à jouer dans l'identification, le recensement et la qualification des données publiques. En Grande Bretagne, « les fonctionnaires dans les administrations ont identifié les bases de données à ouvrir et les ont proposées aux coordinateurs de projets, parfois en contournant une hiérarchie peu au fait de l'intérêt de la démarche, voire hostile¹¹⁹ ».

(ii) Les entreprises exerçant une mission de service public

Dans les limites indiquées plus haut, les données produites par un établissement public ou une entreprise privée, dans le cadre d'une mission de service public autre qu'à caractère

industriel et commercial, entrent dans le périmètre des données publiques réutilisables.

La mise à disposition de ces données pour réutilisation est cependant un sujet complexe, car les limites du caractère commercial ou industriel d'un service public peuvent être floues et discutables. En mettant des données liées à leur fonctionnement à la disposition du public, les entreprises peuvent craindre de communiquer des informations qui devraient relever du secret commercial.

Au-delà de ces craintes, la démarche d'ouverture de données n'est pas spontanément naturelle, et les grandes entreprises les plus concernées (SNCF, RATP, La Poste) seront probablement amenées à évoluer graduellement sur le sujet, grâce aux exemples vertueux mis en avant par la communauté de réutilisateurs, et à mesure que les bénéfices potentiels d'une telle mise à disposition leur apparaîtront plus clairement.

(iii) La recherche scientifique

Les chercheurs sont par nature des collecteurs voire des producteurs de données, des plus élémentaires aux plus complexes. Le monde de la science est de plus régi par le principe de la publication : la reconnaissance d'un scientifique passe par les résultats qu'il publie et qu'il soumet à la communauté scientifique.

Pour donner de la visibilité à la production scientifique, et en allant dans le sens d'une dynamique globale¹²⁰, le Centre National de Recherche Scientifique (CNRS) s'est lancé à partir de 2000, dans le développement d'archives ouvertes via Internet. La base HAL¹²¹, comme « Hyper article en ligne » est un site sur lequel les chercheurs peuvent venir déposer directement leurs articles. « Ouverte à toutes les disciplines, en progression constante, la plateforme HAL reçoit actuellement environ 1 800 nouveaux articles chaque mois, soit près d'un quart de la production scientifique française. »¹²²

¹¹⁹ Valérie Peugeot : « Les enjeux publics, économiques et citoyens de l'ouverture des données : l'expérience britannique »
http://academia.edu.documents.s3.amazonaws.com/1850987/CommunicationDocSoc2010_VPeugeot_opendata.pdf

¹²⁰ D'après le site Internet du CNRS : « un mouvement international en faveur de la communication scientifique directe sur Internet initié dans les années 90 par le physicien Paul Ginsparg. Ce scientifique américain a été le premier à mettre au point une archive ouverte, « ArXiv », consacrée à l'origine aux articles de physique et devenue progressivement l'archive internationale de référence en physique, mathématiques, informatique et certaines spécialités de biologie. »

¹²¹ Entretien avec Stéphane Cottin, Secrétaire général du gouvernement.

¹²² <http://www.ambafrance-cn.org/L-archives-ouvertes-HAL-plus-de-100-000-articles-scientifiques-en-ligne.html>

Au-delà des articles scientifiques, le mouvement d'ouverture des données brutes touche de plus en plus le monde de la recherche. Les grands journaux scientifiques demandent notamment aujourd'hui aux auteurs de publier, en parallèle à leurs conclusions, les données brutes issues de leurs expériences.

2.2. Les institutions de la réutilisation des données publiques

(i) La Commission d'Accès aux Documents Administratifs (CADA)

La Commission d'Accès aux Documents Administratifs (CADA)¹²³ est une autorité administrative indépendante. Ses statuts ont été modifiés par le décret n° 2005-1755 du 30 décembre 2005 relatif à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques.

La CADA joue un rôle consultatif et peut être saisie de requêtes lorsqu'un citoyen juge non conformes à la loi les conditions d'accès ou de réutilisation qui lui sont imposées. Elle peut intervenir au profit des utilisateurs faisant face à un refus de réutilisation d'informations publiques, ou répondre aux sollicitations des administrations demandant des conseils sur le caractère communicable ou réutilisable des documents.

Bien qu'intervenant pour tous les documents détenus par un service de l'État, une collectivité territoriale, un établissement public ou un organisme chargé de la gestion d'un service public, qu'il soit public ou privé, son rôle reste limité aux documents administratifs et aux informations publiques.

En 2009, les problèmes ayant trait à la réutilisation des données ont représenté seulement 5 % du total des demandes adressées à la CADA¹²⁴. Dans ces cas-là, l'intégralité des réponses a été favorable aux réutilisateurs.

(ii) CNIL

Par ailleurs, la Commission Nationale de l'Informatique et des Libertés (CNIL)¹²⁵,

¹²³ Commission d'Accès aux Documents Administratifs
<http://www.cada.fr/>

¹²⁴ Martinez, 2010, « Topic report : PSI reuse in France, overview and recent developments ».
http://www.epsiplatform.eu/topic_reports/topic_report_no_10_psi_re_use_in_france_overview_and_recent_developments

¹²⁵ La CNIL a été créée par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, modifiée en 2004. <http://www.cnil.fr/la-cnil/>

chargée de veiller au respect de la préservation des droits de la vie privée et des libertés dans un monde numérique, peut être saisie, ou se saisir d'elle-même, de tout projet ou des démarches de réutilisation de données publiques, quand ces données présentent des implications en matière de vie privée et de données personnelles. Notamment, en parallèle à la Commission du Secret statistique de l'INSEE, elle veille à l'anonymisation des données ou des statistiques qui pourraient révéler des informations personnelles.

2.3. La communauté des réutilisateurs

(i) Les citoyens

Chaque citoyen peut récupérer des données publiques et les enrichir avec des données personnelles ou en les combinant avec d'autres données ouvertes. Il peut les utiliser à titre personnel ou en tant que producteur de données au profit de la collectivité.

Des développeurs informatiques passionnés appliquent leurs compétences à la réutilisation et à la visualisation des données publiques. Souvent issus des communautés organisées autour du développement de logiciels libres, ces développeurs sont aujourd'hui des parties prenantes essentielles au mouvement d'ouverture des données. Leurs réalisations illustrent les possibilités liées à l'*Open Data*.

Ainsi, le météorologue et prévisionniste Guillaume Séchet a pu créer à partir de relevés personnels son propre site de prévisions météorologiques pour Paris, MétéoParis. À partir d'une base de données historiques sur les événements climatiques en France, il diffuse en ligne lui-même (service Web et mobile) ses prévisions météorologiques personnelles.

(ii) Les chercheurs

Comme mentionné plus haut, le monde de la recherche a joué un rôle essentiel dans l'éclosion du mouvement de l'*Open Data*.

La réutilisation de données produites ou mises en commun par d'autres chercheurs a été historiquement, et plus encore aujourd'hui, un élément important du succès de grands projets scientifiques, du séquençage du génome humain au développement du répertoire standards de « pièces » de génome et protéome utilisé en biologie synthétique¹²⁶.

Aujourd'hui, avec des capacités de traitement toujours plus importantes, les

¹²⁶ BioBricks (« repository of parts for synthetic biology »)
<http://biobricks.org>

possibilités offertes par l'ouverture des données offrent un champ de recherche considérable. Pour cette raison, l'INSEE, dans sa politique de mise à disposition des données statistiques, propose un service spécifique dédié à la recherche via le centre QUETELET lié au CNRS¹²⁷. L'IGN fait de même.

(iii) Les acteurs associatifs

Plusieurs associations ou collectifs se sont donnés pour objectif d'encourager à la mise en ligne des données publiques au niveau national comme au niveau local, comme Regards Citoyens, LiberTIC ou la FING (Fondation Internet Nouvelle Génération).

Dans une démarche de transparence démocratique, certaines associations, par exemple dans les domaines de l'environnement ou des droits des consommateurs, souhaitent pouvoir collationner des données brutes en vue de documenter leurs analyses ou de dresser un bilan ou une évaluation d'une situation.

Les acteurs associatifs peuvent aussi participer de manière collaborative à des actions de collecte de données publiques présentes en ligne (« *wiki scraping* »), voire à la génération des données elles-mêmes.

Par exemple suite au séisme qui a frappé Haïti le 12 janvier 2010, la communauté d'utilisateurs d'OpenStreetMap a réalisé très rapidement une cartographie de meilleure qualité que l'existant, enrichie de nombreuses données actualisées comme l'état des axes de circulation et l'emplacement des campements de secours. Ce type de productions non institutionnelles peut s'avérer d'une grande utilité pour la gestion des crises.

(iv) Les acteurs économiques

La réutilisation commerciale des données peut permettre l'éclosion de nouvelles activités économiques, ou le développement et l'amélioration d'initiatives déjà existantes.

En France, le Groupement Français de l'Industrie de l'Information (GFII) a créé un

groupe de travail spécifique et très actif sur l'ouverture des données publiques¹²⁸.

La réutilisation des données s'adresse bien sûr aux acteurs les plus établis de l'économie, et notamment les grandes entreprises des secteurs de l'information, qui peuvent utiliser la donnée publique pour améliorer leurs processus productifs.

Parce qu'elle représente un fort potentiel d'innovation, la donnée publique est un ingrédient particulièrement important du développement de startups et de jeunes entreprises innovantes de l'économie numérique.

Le développement de ce secteur est un enjeu majeur pour l'emploi : Internet a permis plus d'un quart des créations nettes d'emploi depuis 1995 en France. Il est donc crucial que les conditions de réutilisation permettent aux startups et aux entrepreneurs de s'approprier les données publiques et d'expérimenter de nouvelles activités à partir de leur réutilisation.

Des conditions véritablement non discriminatoires pour des acteurs économiques à très faibles moyens, comme des startups en phase préliminaire à une levée de fonds, ne devraient pas imposer de redevances excessives. La facturation immédiate de plusieurs dizaines de milliers d'euros (e.g. 123 000 euros dans le cas des données de fiscalité locale¹²⁹) pour réutilisation commerciale est à ce titre un exemple de ce qui peut fortement décourager l'innovation.

(v) La presse et les médias

Le journalisme de données (« *data journalism* ») est un exemple marquant de l'importance potentielle de la réutilisation des données publiques pour le débat démocratique. The Guardian¹³⁰ a ainsi joué un rôle essentiel au Royaume-Uni, et le New York Times a décidé en 2009 de mettre à disposition sous forme de données ouvertes sa base de données de personnes, d'institutions et d'événements qu'il tient à jour depuis 1913.

En France, le site d'information OWNI¹³¹ se présente comme un « *digital think tank* » avec pour ambition d'être le « laboratoire du

¹²⁷ Réseau Quételet : <http://www.reseau-quetelet.cnrs.fr/>. Il s'agit d'un réseau constitué par le Centre de données socio-politiques (CDSP) de Sciences-Po, le Centre Maurice Halbwachs – Archives de données issues de la statistique publique (CMH-ADISP, UMR 8097), et l'Institut national d'études démographiques (INED). Il permet aux chercheurs français et étrangers d'obtenir des bases de données nécessaires à leurs traitements dans les domaines des grandes enquêtes, recensements et autres bases de données issues de la statistique publique française, grandes enquêtes françaises provenant de la recherche, et un accès privilégié à des enquêtes internationales.

¹²⁸ http://www.gfii.asso.fr/rubrique.php3?id_rubrique=60

¹²⁹ Décret fixant la redevance pour le fichier « REI » <http://www.legifrance.gouv.fr/affichTexte.do;jsessionid=?cidTexte=JORFTEXT000022807097&dateTexte=&oldAction=rechJO&categorieLien=id>

¹³⁰ Guardian Open Platform

<http://www.guardian.co.uk/open-platform>

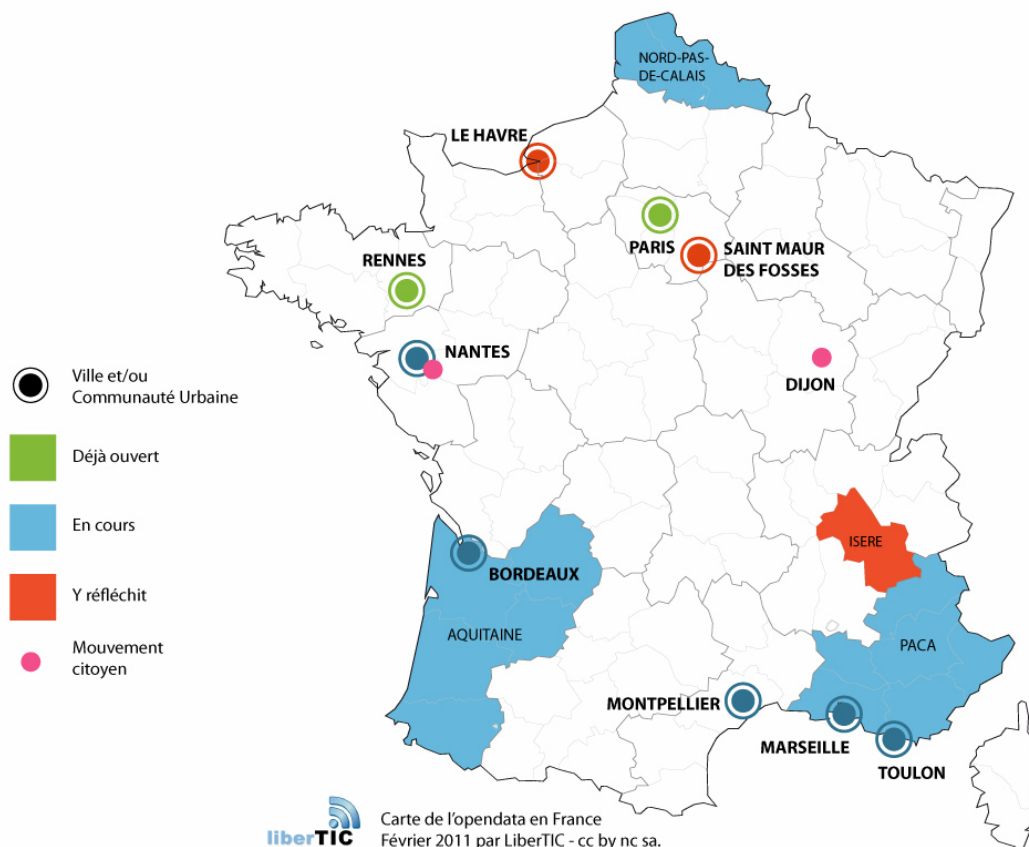
¹³¹ OWNI (Objet Web Non Identifié)

<http://owni.fr>

journalisme innovant », « à la pointe de l'expérimentation en matière de *datajournalism* ». Il a été lauréat d'un prix aux

Online Journalism Awards en octobre 2010 dans la catégorie « meilleur site en langue non-anglaise ».

Chapitre V. Premières expériences d'ouverture des données publiques en France



Ce chapitre propose un état des lieux des principales initiatives de réutilisation des données publiques en France¹³², au niveau des villes, des régions, et national.

1. La réutilisation des données dans les collectivités territoriales

1.1. Rennes

La ville de Rennes a été pionnière avec l'ouverture de données publiques¹³³ sur le transport à Rennes Métropole : 11 applications mobiles ont ainsi vu le jour en 2 mois.

¹³² Gallon, C., 2011, « Carte contributive sur l'ouverture des données en France », Libertic, <http://libertic.wordpress.com/2011/02/11/carte-contributive-sur-louverture-des-donnees-publiques-en-france/>

¹³³ <http://www.data.rennes-metropole.fr/fr/lentrepot.html>

Rennes Métropole, la Ville de Rennes et Kéolis Rennes organisent, en partenariat avec la Caisse des dépôts, un concours pour favoriser le développement d'applications autour des données disponibles : transports, guide du vivre à Rennes, données géographiques. Le concours a permis de récompenser plusieurs applications utiles parmi les 117 participants.

Les données mises à disposition par la Ville de Rennes et Rennes Métropole :

- Un accès direct vers les données de transports publics grâce au portail data.keolis-rennes.com. Ces données concernent les bus et le métro, les lignes, les horaires, la géolocalisation des stations, leurs équipements et le service Vélo STAR. Elles sont publiées par Kéolis Rennes, exploitant du réseau STAR (Service des Transports de l'Agglomération Rennaise) et des Vélo STAR.

- Les données de la base « Guide Vivre à Rennes » : toutes les informations pratiques de 1 500 organismes publics et parapublics, géolocalisés sur Rennes. Elles sont accessibles grâce à une API, interface permettant programmation et interrogation, développée spécialement par InCité Solutions pour le compte de la Ville de Rennes.
- Des données issues du Système d'Information Géographique de Rennes Métropole et de la Ville de Rennes : voies et adresses, stationnements pour handicapés, points d'apports volontaires, et bientôt l'orthophotographie de l'ensemble de la Métropole. Ces données sont fournies sous forme de fichiers à télécharger.

1.2. Ville de Paris

Le principe de l'ouverture des données publiques de la Ville de Paris a été arrêté suite à un débat du Conseil de Paris en juin 2010¹³⁴. Les services de la direction des affaires juridiques de la ville ont alors engagé une concertation avec des acteurs promouvant les démarches d'ouverture des données publiques, notamment Regards Citoyens, *Creative Commons* France et Veni, Vidi, Libri, afin de déterminer la licence la mieux adaptée. La plateforme privée *Data Publica*¹³⁵ a aussi repris un certain nombre de ces données, afin de compléter son offre de données publiques.

La licence ODbL (*Open Database Licence*) développée par l'*Open Knowledge Foundation* a été retenue. Adoptée par le Conseil de Paris de décembre 2010¹³⁶, elle autorise les entreprises qui le souhaitent à reprendre les données pour une utilisation commerciale, tant qu'elles acceptent ensuite de partager leur création dans les mêmes conditions (clause « *share alike* »), de manière similaire aux dispositions de la licence *Creative Commons* CC-BY-SA¹³⁷.

Par ailleurs, sur la demande d'entreprises qui ne souhaiteraient pas se soumettre à cette condition de partage à l'identique, la ville de

Paris offre aux acteurs qui le souhaitent de souscrire une licence payante non exclusive.

Fin décembre 2011, la Ville de Paris recense les données susceptibles d'être diffusées lors d'un premier « lâcher de données » en janvier 2011 sur le portail opendata.paris.fr¹³⁸. Les données personnelles sont transformées en informations statistiques agrégées et anonymisées le cas échéant.

Pour aller plus loin, une animation de la communauté est prévue avec la mise en place d'un site d'échanges entre les utilisateurs des données publiques et d'un concours d'applications pendant un séminaire informel qui a eu lieu en février 2011 (« *BarCamp* »).

Quelques exemples des premiers jeux de données ouverts par la Ville de Paris :

- Citoyenneté : résultats d'élections, cartes des bureaux de vote, cartes électorales, décès, mariages, naissances, répertoire des prénoms déclarés.
- Urbanisme : fonds de cartes des plans de voirie, tracés des routes, description des trottoirs, volume du bâti.
- Déplacements : nomenclature des voies.
- Services : liste des établissements scolaires, des écoles maternelles et élémentaires, des crèches, des haltes-garderies, des piscines.
- Environnement : liste des parcs et jardins, carte des arbres d'alignement, bâtiments, référentiels de la flore.
- Culture : liste d'ouvrages disponibles, notices des œuvres de fonds des musées, statistiques des prêts dans les bibliothèques en 2010.

1.3. L'expérience de l'Atelier Parisien d'Urbanisme : diffuser la donnée brute plutôt que mise en forme

À la demande du Conseil de Paris, l'Atelier Parisien d'Urbanisme (APUR) s'est engagé dans l'ouverture de ses données dans le cadre de la directive « Inspire ». Le Conseil d'administration de l'APUR a initié en 2008 une réflexion, en y associant l'ensemble des partenaires de l'APUR (INSEE, Préfecture de Police, Pompiers, CAF, etc.) afin d'en étudier les modalités.

¹³⁴ Un des auteurs de ce rapport est élu parisien et administrateur de l'APUR, l'Atelier Parisien d'Urbanisme.

¹³⁵ *Data Publica* : <http://www.data-publica.com>

¹³⁶ Licence consultable sur : <http://www.paris.fr/portail/viewmultimediacomment?multimediacomment-id=94658>

¹³⁷ Licence *Creative Commons* BY-NC-SA 2.0 France : <http://creativecommons.org/licenses/by-nc-sa/2.0/fr/>

¹³⁸ Open Data Paris : <http://opendata.paris.fr/>

Les débats ont porté d'emblée sur la définition de la donnée publique, en particulier lorsqu'elle provient de l'agrégation de sources issues de divers partenariats, et sur l'utilité d'une diffusion des données brutes plutôt qu'une diffusion de données mises en forme.

Afin d'explorer la faisabilité de diffuser des données mises en forme, une maquette a été programmée à partir de l'étude réalisée annuellement pour l'observation statistique des familles parisiennes. Le produit réalisé permettait de consulter et télécharger les tableaux de données brutes liés à chaque statistique ou graphique présenté dans l'étude. Ces études elles-mêmes sont par ailleurs désormais diffusées de façon gratuite.

Cette réalisation a nécessité un travail important de développement informatique et de mise en forme des données, allongeant ainsi le temps nécessaire pour fournir un produit pouvant être diffusé. Les surcoûts générés par une telle démarche ont conduit à l'abandonner pour s'orienter vers une diffusion des bases de données, sans remise en forme préalable – ou du moins avec une remise en forme limitée aux questions de formats de données.

Aujourd'hui, l'APUR a commencé à déclarer ses métadonnées sur le géocatalogue de l'IGN¹³⁹. Les données elles-mêmes sont très détaillées, donc volumineuses. Leur diffusion soulève des problèmes de capacités techniques que l'APUR n'est pas nécessairement à même de résoudre en propre¹⁴⁰.

Cette expérience montre cependant l'intérêt pour le producteur de diffuser la donnée brute plutôt que mise en forme ou enrichie, car cela lui permet de mettre directement à disposition les fichiers que ses agents utilisent pour leur propre travail, économisant donc l'effort et les moyens nécessaires à une éventuelle phase de remise en forme.

1.4. Étude de cas : partage de l'information géographique par les régions

Certains départements et certaines régions ont aussi commencé à mettre en ligne leurs données publiques. Des collectivités ont par exemple mis à disposition une partie de leurs

archives en ligne. C'est le cas du département de la Loire¹⁴¹, et de Paris¹⁴².

Un grand nombre de collectivités territoriales, principalement des régions, mettent à disposition sur leurs sites des données géographiques. Cette section repose sur une enquête conduite par les auteurs en novembre et décembre 2010 sur les portails régionaux d'information géographique. Les résultats n'ont pas la prétention de l'exhaustivité.

(i) Disparités dans la mise en œuvre

Les données mises à disposition sont en général des données dites « métier », principalement de type cartographique, qui ne sont généralement pas ouvertes à d'autres tiers que les entités publiques.

Le besoin communément exprimé est celui de partager des données entre entités publiques dans l'espoir de réaliser des économies d'échelle. La Bourgogne a ainsi pu économiser près d'un million d'euros grâce à l'achat de données IGN en licence étendue.

Du fait de cette finalité d'économies, ces sites n'ont pas été conçus ou adaptés pour le grand public. Une exception est le site de la Basse-Normandie, développé spécifiquement pour le grand public par le Pôle Géomatique Normand avec un atlas interactif.

Ces différents services ont des statuts hétérogènes. Certains de ces sites ont été développés à l'initiative des services de l'État, d'autres des régions ou des départements. On recense la coexistence de statuts associatifs, d'établissements publics ou parapublics (notamment EPIC ou GIP), et de services de l'État ou de collectivités territoriales. Il existe aussi un cas de statut privé, celui du PRIGE (Pas-de-Calais).

Enfin, certains sites ne sont pas constitués en personne morale, comme Géopal, Géobourgogne, Géobretagne, et PGN. Ils sont alors intégrés dans les sites des organismes partenaires (conseil régional, DREAL, etc.).

Certaines plateformes ont été développées en régie quand d'autres ont fait l'objet d'un marché public. La Bourgogne réalise la sienne dans le cadre d'un partenariat public-privé.

¹³⁹ <http://www.geocatalogue.fr/> Le géocatalogue centralise en France les métadonnées concernées par « Inspire ».

¹⁴⁰ D'autre part, le manque à gagner lié à une diffusion gratuite est évalué au montant des données vendues, soit environ 50 à 100.000 euros par an, prestations à la demande comprises.

¹⁴¹ Archives de la Loire
http://www.loire.fr/display.jsp?id=c_828986

¹⁴² <http://canadp-archivesenligne.paris.fr/>

(ii) Avancées dans l'interopérabilité

Pour certaines régions – telles que la Haute-Normandie, la Lorraine, la Franche-Comté, la Corse, la Champagne-Ardenne ou la région Midi-Pyrénées – il existe plusieurs plates-formes d'information géographique auxquelles sont associés les services de l'État, des Conseils Régionaux et des Conseils Généraux. Ces plateformes ne sont pas toutes interopérables en raison de l'existence de plusieurs standards promus par différentes entités publiques (régions, État, ministère de l'Écologie ou ministère de l'Intérieur).

Des démarches de mutualisation ont donc été mises en place. Prodiges, « Plate-forme Régionale pour Organiser et Diffuser l'Information Géographique », est ainsi née en 2003 d'une initiative des services déconcentrés de l'État en région Rhône Alpes. Une circulaire du 24 octobre 2007 des ministères de l'Économie et des Finances et de l'Intérieur, incite les préfets à favoriser les démarches de mutualisation de type Prodiges en respectant les référentiels nationaux et en s'inscrivant dans une démarche d'interopérabilité.

Prodiges a été depuis mise en service également en Pays de la Loire, et de nombreuses régions sont en train de développer des projets similaires : Poitou-Charentes, Picardie, Alsace, Corse, Franche-Comté, Basse et Haute Normandie, Réunion et Guadeloupe. Cela permet aux services de l'État de structurer et d'organiser l'offre régionale de données à référence spatiale, et de faciliter les échanges de données entre services mais aussi avec les partenaires institutionnels.

En parallèle, le Ministère de l'Écologie a développé CARMEN, la « Carte du Ministère de l'Environnement », répondant aux mêmes objectifs de partage d'informations géographiques, avec des acteurs et des pratiques différents. Existente également l'outil Cartélie, développé par l'ex-Équipement, et Géoweb, développé par le ministère de l'Agriculture.

Une convergence de Carmen et Prodiges a été amorcée avec l'objectif premier de ne pas perturber les habitudes des communautés d'utilisateurs. Cela a conduit en septembre 2010 à une version de Prodiges intégrant deux modules de Carmen 1.6 : la composition et la visualisation. De son côté, Géoweb est en cours de remplacement par Carmen.

L'un des sites interrogés a souligné le manque de préconisations nationales qui répondraient aux besoins locaux. Il semble

donc important de rechercher la convergence de ces différents outils et la généralisation de l'interopérabilité, qui seraient facilitées par un pilotage interministériel et un travail de coordination avec les collectivités locales.

Certaines régions signalent toutefois une difficulté de mise en place d'une véritable culture du partage des données entre acteurs publics et a fortiori avec le grand public.

En ce qui concerne l'ouverture des données au grand public, les responsables des sites signalent aussi le besoin de rassurer les partenaires sur les aspects juridiques de la diffusion des données afin de lever les craintes qu'ils ont à ce sujet.

(iii) Des dynamiques régionales d'ouverture des données

D'après les réponses obtenues, les régions déjà bien organisées dans ce domaine (Nord-Pas-de-Calais, Auvergne, Provence-Alpes-Côte-d'Azur, Aquitaine, Basse-Normandie, Languedoc-Roussillon, Rhône-Alpes, Pays de la Loire, Picardie, Poitou-Charentes, Alsace) poursuivent leurs activités en développant leurs partenariats et en renforçant les projets thématiques. Les Départements et Collectivités d'Outre-Mer ne sont pas en reste. La Guyane, la Martinique, la Guadeloupe, la Réunion et la Nouvelle-Calédonie innovent elles aussi dans ce domaine.

La création de ces sites régionaux s'est accélérée du fait de deux dynamiques différentes. Tout d'abord, nombre de ces projets, développés en partenariat entre les services de l'État et le conseil Régional, ont bénéficié de la signature de Contrats de Plans État Région pour la période 2007 – 2013.

D'autre part, les obligations créées par la directive Inspire ont accéléré le processus. Bien qu'elle soit rarement mentionnée par les responsables que nous avons interrogés, la mise en œuvre de la directive est un moteur prédominant dans la création de sites plus récents et de la refonte des sites plus anciens.

Un nouvel écosystème de l'information géographique est ainsi en train de prendre forme. Alors que les régions pionnières ont développé seules leurs sites Internet, le réseau des Centres Régionaux d'Information Géographique (CRIGE) s'est mis en place. Il permet un accompagnement des nouveaux entrants par les plus expérimentés, et offre parfois la possibilité d'utiliser les mêmes codes sources sous licence libre.

(iv) *Hétérogénéité des moyens économiques*

Afin d'avoir une vision plus précise des moyens engagés par les régions, l'enquête menée auprès des sites régionaux a été complétée par la consultation des données du CRIGE. Les résultats sont résumés dans le tableau suivant.

		Investissement (k€)	Fonctionnement (k€)	Personnel (ETP)
Auvergne	<i>CRAIG</i> *	Appel d'offres en cours	35	1
Midi – Pyrénées	<i>GéoMIP</i>	28	9	1
Nord – Pas-de-Calais	<i>PPIGE</i> *	500	200	2
Basse Normandie	<i>PGN</i> *	58	20	1
Pays-de-Loire	<i>GéoPAL</i> *	100	37,5	2
	<i>SIGloire</i>	350	65	8
PACA	<i>CRIGE PACA</i> *	200	35	0,7
Rhône – Alpes	<i>PRODIGE</i>	150	19	15
Moyenne		198	52,6	3,8

Tableau de synthèse des moyens mis en œuvre par quelques régions.

* indique que les informations proviennent directement des plateformes.

Dans les autres cas, les informations proviennent du catalogue des infrastructures de données géographiques françaises 2010 de l'AFIGEO et de leurs sites Internet.

2. L'ouverture des données au niveau national : trois études de cas

2.1. Étude de cas : Le Ministère de l'Écologie

Le ministère de l'Écologie, du Développement Durable, des Transports et du Logement (MEDDTL) a lancé une démarche de mise en ligne de ses données.

Un portail¹⁴³ rassemblant les sites comportant des informations ou des données environnementales est disponible depuis le 18 juillet 2009¹⁴⁴. Ce portail propose un accès à quatre-vingt sites adhérents et reçoit 22 000 visites par mois. Le ministère souhaite développer et soutenir la volonté affichée d'animer une communauté d'adhérents et d'utilisateurs : une conférence des adhérents devrait être prochainement organisée.

La fusion des ministères de l'Environnement et de l'Équipement, aux traditions de diffusion assez différentes, a rendu la constitution du portail plus complexe. Certains services comme le CERTU¹⁴⁵ et le SETRA vendent historiquement certaines de leurs publications, ce qui en limite la diffusion.

On observe une structuration progressive à partir de certains services, et notamment le SOeS¹⁴⁶, les DREAL et les services de l'Écologie. Ces initiatives tendent à mettre en cohérence l'organisation de la diffusion des données environnementales.

La création du portail Environnement a nécessité un investissement d'environ 350 000 euros, et un effort de maintenance annuelle de 150 000 euros par an. À cela s'ajoute l'emploi de deux fonctionnaires de catégorie B à temps plein pour l'animation (relation avec les adhérents et l'équipe technique du portail) et l'équivalent d'un tiers d'équivalent temps plein d'un fonctionnaire de catégorie A.

2.2. Étude de cas : l'INSEE

(i) L'INSEE diffuse ses données en ligne depuis 2003

L'Institut National de la Statistique et des Études Économiques (INSEE)¹⁴⁷ a pour mission de produire et diffuser des statistiques et des études économiques. La seule contrainte majeure à la diffusion est celle de la confidentialité, assurée par le Comité du secret statistique et des procédures d'anonymisation. L'INSEE publie ainsi un très grand nombre de jeux de données (plus de 52 000 fichiers Excel seraient disponibles sur son site en janvier 2011¹⁴⁸).

L'institut décide dès 2003 de diffuser gratuitement sur Internet une partie de ses données, motivé à la fois par la forte demande pour ses données et par la facilité de leur mise à disposition au grand public via Internet.

L'INSEE a privilégié la rapidité de mise à disposition des données, ce qui a eu une incidence sur l'ergonomie du site et donc sur l'accessibilité des données. La quantité de données ouvertes étant considérable, les utilisateurs expriment le sentiment d'être noyés dans un site « *confusant et décevant* », selon une enquête de satisfaction (par ailleurs positive) conduite par l'INSEE. La visibilité de la démarche en a aussi été amoindrie malgré son importance.

(ii) 24 millions de visites annuelles

Le vecteur de cette diffusion est le site de l'INSEE, INSEE.fr, qui reçoit 24 millions de visites annuelles avec 12 millions de visiteurs uniques par an, soit un trafic exceptionnel pour un site public.

Le public peut être séparé en plusieurs catégories d'utilisateurs : la presse, les collectivités territoriales, les particuliers, les entreprises, le public des enseignants et celui des chercheurs.

Ces derniers représentent un public aux besoins spécifiques. Pour leurs travaux, ils peuvent obtenir des données rendues anonymes et produites à la demande via le réseau « QUETELET¹⁴⁹ » lié au CNRS. Pour ces travaux spécifiques, la tarification est en cours d'élaboration.

¹⁴³ <http://www.toutsurlenvironnement.fr/>

¹⁴⁴ Les éléments présentés dans cette section proviennent des réponses apportées par mail par Laurent Coudercy, responsable du portail environnemental jusqu'au 30 novembre 2010.

¹⁴⁵ CERTU. <http://www.certu.fr>

¹⁴⁶ Service de l'Observation et des Statistiques du Commissariat général au développement durable (CGDD). <http://www.stats.environnement.developpement-durable.fr>

¹⁴⁷ Les éléments présentés dans cette section proviennent d'un entretien mené le 26 novembre 2010 avec M. Pierre Audibert (directeur de la diffusion à l'INSEE) et M. Thierry Aouizerate (chef du département INSEE Info Service).

¹⁴⁸ Bancelhon, F. et Gans, B., 2011, « Size and Structure of the French Public Sector Information » http://gov.opendata.at/site/sites/default/files/Size%20and%20Structure%20of%20French%20PSI_0.pdf

¹⁴⁹ Ibid.

(iii) Mission de service public et diffusion à titre commercial

Les activités de diffusion l'INSEE se répartissent en deux ensembles distincts, la diffusion de service public et la diffusion commerciale.

L'activité de service public de l'INSEE étant financée par le budget de l'État, les produits standards, notamment l'ensemble des jeux de données disponibles en ligne, sont diffusés et réutilisables gratuitement (sauf en version papier du fait des frais d'impression).

L'indépendance de la statistique publique étant un élément important pour le bon fonctionnement de notre démocratie, l'INSEE doit statutairement générer une partie de ses revenus par des activités commerciales lui garantissant une certaine indépendance des choix budgétaires de l'État.

Les produits réalisés sur mesure, tels des extractions de base de données spécifiques ou des recoupements réalisés à la demande, sont soumis à une redevance pour service rendu, qui n'est pas liée au droit sur les données publiques.

La réutilisation d'un certain nombre d'informations publiques à titre commercial, dans le cadre de la loi de 1978 modifiée en 2005, est aussi soumise à une redevance.

Ainsi, la base de donnée des prénoms les plus usités par les Français, ou encore la base SIRENE, qui regroupe les données relatives aux entreprises et à leurs établissements, sont soumises à redevance. Ce choix est contesté par certains usagers qui souhaiteraient que la base SIRENE, financée par l'argent public et créée à partir des informations saisies par les entreprises elles-mêmes, soit diffusée gratuitement¹⁵⁰.

Le choix de tarification de la base SIRENE a donné lieu à un contentieux juridique. Le Conseil d'État l'a tranché par une décision¹⁵¹ en date du 24 juillet 2006 qui a rejeté un recours intenté par un utilisateur. Ce dernier contestait la politique tarifaire de diffusion arguant notamment « que le ministre de l'économie, des finances et de l'industrie ne

pouvait légalement prendre en compte, pour déterminer le montant de la redevance acquittée par les rediffuseurs, le coût des opérations d'amélioration de la base SIRENE et le moyen tiré de ce que cette redevance devrait, en tout état de cause, baisser en raison de l'augmentation du nombre des clients de l'INSEE ».

La décision de rejet a été notamment motivée pour les raisons suivantes : « la redevance correspondant à la mise à disposition de la base de données SIRENE, qui doit être regardée comme une œuvre collective, peut légalement inclure des droits relevant de la propriété intellectuelle au profit de l'INSEE [...]. L'INSEE peut, par suite, prendre en compte, pour déterminer le montant de la redevance demandée aux rediffuseurs, le coût des opérations d'amélioration de la base SIRENE [...] » tout en veillant à ce que « les droits relevant de la propriété intellectuelle et artistique ne doivent pas conduire l'opérateur public à réaliser des profits d'un montant manifestement disproportionné. »

(iv) Fournir de nouveaux produits

Conscient des besoins nouveaux des usagers, l'INSEE va rénover son portail et souhaite développer ces données sur un maillage plus fin que l'IRIS utilisé jusqu'ici.

Son ambition est de disposer de bases de données qui permettraient à l'utilisateur d'obtenir l'information selon une échelle glissante, du local jusqu'au niveau national.

Des restrictions liées aux règles de confidentialité devraient toutefois s'appliquer en fonction de l'échelle souhaitée car l'anonymat est fortement lié à la précision spatiale.

Après avoir ouvert très tôt une grande partie de ses données, l'INSEE prévoit aujourd'hui de faire évoluer l'architecture de son site et d'améliorer son ergonomie.

2.3. Étude de cas : L'IGN

(i) L'IGN finance ses opérations par une subvention pour service public et des redevances pour réutilisation

L'Institut Géographique National (IGN)¹⁵² est investi d'une mission de service public définie dans son décret fondateur (décret

¹⁵⁰ Table-ronde entre développeurs et collectivités locales lors de la conférence « Ouverture des Données Publiques Locales et Réutilisations : Focus sur les Initiatives Locales en Europe » le 29 novembre 2010 à Rennes.

¹⁵¹ Décision du Conseil d'État du 24 juillet 2006 : <http://arianeInternet.conseil-État.fr/arianeInternet/ViewRoot.asp?View=Html&DMode=Html&PushDirectUrl=1&Item=2&fond=DCE&texte=247769&Page=1&querytype=simple&NbEltPerPage=4&Puriels=True>

¹⁵² Les éléments présentés dans cette section proviennent d'un entretien mené le 17 décembre 2010 avec Jean-Yves Bréard, directeur par intérim de la maîtrise d'ouvrage déléguée du service public.

n° 81-505 du 12 mai 1981 modifié). Cette mission consiste à décrire, d'un point de vue géométrique et physique, la surface du territoire national et l'occupation de son sol, en faire toutes les représentations appropriées, diffuser les informations correspondantes.

L'institut reçoit une subvention pour charge de service public, établie de manière à ce que l'argent public finance la part d'activité purement publique et l'utilisateur rémunère la part d'activité de nature concurrentielle et commerciale.

Ainsi, la subvention finance la recherche, l'enseignement, la couverture photographique patrimoniale, ainsi que la constitution et l'entretien du réseau géodésique. Ces travaux stratégiques conditionnent la connaissance homogène du territoire national par la puissance publique.

La subvention participe aussi à la mise à disposition des données publiques, notamment la diffusion des données sur le Géoportail. Certaines de ces données sont réutilisables gratuitement, y compris à titre commercial.

En revanche, une partie du périmètre d'activité de l'IGN relève d'activités commerciales. La production, la rédaction, l'impression et la commercialisation des fonds de carte, ou leur diffusion électronique, sont laissées à la charge de l'utilisateur. Il en va de même pour les prestations réalisées à titre de service rendu, soumises à redevance.

Enfin, une partie de l'activité de l'IGN répond à des besoins qui concernent autant l'État que d'autres utilisateurs, dont les collectivités territoriales : les coûts afférant sont alors partagés entre la subvention et l'utilisateur.

C'est le cas notamment de la production de données à moyenne ou à petite échelle (notamment les 1:100.000, 1:250.000 et 1:1.000.000), la production des fonds cartographiques principaux et la production du RGE (Référentiel à Grande Échelle)¹⁵³. Pour ce dernier, la subvention assure environ deux tiers des coûts quand l'utilisateur en supporte un tiers. Le choix de cette clé de répartition est essentiel pour comprendre la formation du tarif des données.

¹⁵³ Le RGE est composé de quatre composantes qui couvrent l'étendue du territoire : les orthophotographies, la base de données topographiques, la base de données parcellaires et la base de données adresses.

(ii) Le périmètre des données publiques : une question complexe

La loi CADA concerne l'ensemble des documents administratifs et des informations publiques. L'information géographique présente, elle, la particularité d'une grande technicité et d'une certaine difficulté de réutilisation des données brutes servant à sa production.

Il est difficile d'établir une distinction nette entre les données produites dans le cadre de la mission de service public (et donc légitimement diffusables gratuitement) et les données de nature commerciale en raison de la complexité de la chaîne de traitement.

Ainsi, même la photographie aérienne ne constitue pas une donnée brute diffusable. À la sortie de l'avion, il s'agit encore de fichiers numériques bruts, inexploitable en pratique hors de la chaîne de production de l'IGN. Pour devenir un produit de type « donné brute diffusable », ces photographies doivent subir des traitements importants dans une chaîne longue et complexe – qui échoue parfois – et dont l'issue est la production de l'orthophotographie.

On pourrait donc argumenter que l'orthophotographie est une donnée brute. Il existe pourtant un marché concurrentiel de production d'orthophotographies¹⁵⁴. Il n'est donc pas possible de la diffuser à tous pour une réutilisation gratuite, à moins d'annihiler ce marché et d'établir de graves distorsions de concurrence.

Cet exemple souligne la difficulté qu'il peut y avoir à définir le périmètre des données gratuites car relevant de la mission de service public, et celui des données tarifées car relevant d'une activité commerciale.

(iii) Virage tarifaire autour du RGE

La complexité des règles de formation des tarifs des données de l'IGN, leur montant relativement élevé¹⁵⁵ et leurs licences parfois restrictives pour le partage des données ont nourri de nombreuses critiques, notamment de la part des collectivités locales. Ces critiques sont alimentées aussi par le raisonnement général voulant que les données publiques

¹⁵⁴ On souligne toutefois que les photographies numériques de l'IGN sont diffusées plusieurs mois après le vol après un premier traitement (aérotriangulation). Les données de vol qui permettent de produire les orthophotographies sont considérées comme des produits intermédiaires au sens de la CADA, et ne sont donc pas diffusées.

¹⁵⁵ À titre d'exemple tiré du catalogue des prix publics : une zone de 100 km x 100 km de la base de données topographique est vendue 55 550 euros (janvier 2011).

devraient être réutilisables gratuitement par tous car financées par de l'argent public.

La réflexion menée à l'occasion du nouveau contrat d'objectif de performances (COP 2010-2013) a débouché sur la décision de rendre gratuites les données de base (RGE) auprès des institutions chargées d'une mission de service public sans caractère industriel ou commercial¹⁵⁶. Sont donc concernés l'État, les collectivités publiques et leurs établissements publics, et les prestataires privés de services publics non industriels ou commerciaux. Les instances de l'Union Européenne ainsi que les autorités publiques des États membres sont aussi concernées.

Ce virage conduit à une perte substantielle de recettes commerciales pour l'IGN, perte compensée par l'État par une augmentation à due concurrence de la subvention pour charge de service public, soit un montant annuel de l'ordre de 10 millions d'euros.

La gratuité du RGE pour les acteurs publics évite naturellement une duplication des dépenses de constitution de référentiels locaux ou régionaux, et permettra une mutualisation de l'infrastructure d'information géographique à travers le pays. L'ensemble des organismes éligibles à la gratuité peut librement s'échanger toutes les données concernées¹⁵⁷, ce qui facilite leur enrichissement collaboratif.

Même si cette gratuité ne concerne pas pour l'instant les opérateurs privés, et notamment les entrepreneurs ou les startups, c'est une première étape, et un mouvement dans le sens de la gratuité.

3. La mise en place d'une politique nationale des données publiques

Les derniers mois ont vu le sujet de la réutilisation des données publiques gagner en importance dans le débat public en France. Plus de 120 personnes ont pris part au premier *Open Data Camp* français à Paris fin 2009, le sujet figure au programme de la conférence

¹⁵⁶ La livraison par téléchargement sur le site Internet <http://professionnels.ign.fr/2/telechargement.htm> ou via le flux émis par le Géoportail est gratuite car financée par la subvention (alors que les coûts de diffusion de la carte papier ne sont pas subventionnés). Les coûts de reproduction et diffusion d'une emprise départementale sont de 329 euros hors taxe pour un DVD et de 514 euros hors taxes pour un disque dur.

¹⁵⁷ Pour aller plus loin : consulter la rubrique « Espace professionnel » du site ign.fr, et en particulier le document « Présentation des nouvelles modalités de diffusion du RGE ».

LeWeb 2010¹⁵⁸ et fait l'objet d'un colloque Regards sur le Numérique en mars 2011. Des collectivités se saisissent du sujet et ouvrent des portails de mise à disposition de données... Depuis plusieurs mois, le gouvernement met en place les éléments d'une politique nationale d'ouverture et de réutilisation des données publiques.

(i) La mission « Etalab » coordonne les administrations en matière de données publiques

En 2008, Éric Besson, secrétaire d'État chargé de la Prospective, de l'Évaluation des politiques publiques et du Développement de l'économie numérique auprès du Premier ministre, propose dans le cadre du plan « France Numérique 2012 » la création d'un portail unique des données publiques¹⁵⁹.

En février 2010, le député Frank Riester réunit un groupe d'experts du numérique chargé de faire des propositions à Nathalie Kosciusko-Morizet, secrétaire d'État auprès du Premier ministre, chargée de la Prospective et du Développement de l'économie numérique, en vue de l'amélioration de l'administration numérique et de son lien avec les usagers internautes.

Ils font de la création de ce portail l'une de leurs principales propositions¹⁶⁰, dans le but d'encourager la création de services innovants en libérant les données publiques et en facilitant leur réutilisation par la communauté des entrepreneurs et des startups.

Le principe de sa création est validé dans le cadre de la Révision Générale des Politiques Publiques en juin 2010, et fait l'objet d'une communication en Conseil des ministres le 24 novembre 2010¹⁶¹.

¹⁵⁸ Lacombe, R., 2010, « Tocqueville 2.0 and Open Government Data ». LeWeb 10 Ignite Session.

<http://www.youtube.com/watch?v=6qj8jS2Xsfk>

¹⁵⁹ Besson, E., 2008, « France numérique 2012 : Plan de développement de l'économie numérique », octobre 2008. <http://lesrapports.ladocumentationfrancaise.fr/BRP/084000664/0000.pdf>

Action n° 39 « Favoriser le développement de nouveaux produits et services en créant un portail unique d'accès aux données publiques ».

¹⁶⁰ Riester, F. et al., 2010, « Améliorer la Relation Numérique à l'Usager : Rapport Issu des Travaux du Groupe Experts Numériques ».

http://www.modernisation.gouv.fr/fileadmin/Mes_fichiers/pdf/RapportGroupeExpertsNumeriques.pdf

Proposition n° 22 : « Créer une plateforme d'innovation de services 'État Lab' permettant aux acteurs tiers de développer des services innovants à partir de données publiques ».

¹⁶¹ Conseil des ministres du 24 novembre 2010 : communication relative à l'administration électronique, par le ministre du budget, des comptes publics, de la

Le Premier ministre François Fillon crée par décret la mission Etalab, rattachée sous son autorité au Secrétariat général du Gouvernement, le 21 février 2011¹⁶².

En plus de « coordonner l'action des administrations de l'État et apporter son appui à ses établissements publics administratifs pour faciliter la réutilisation la plus large possible de leurs informations publiques », la mission Etalab est chargée de la création d'un portail unique interministériel, la future plateforme *Open Data* de l'État « data.gouv.fr ».

(ii) Politique nationale de réutilisation des données publiques

La mise en place de la plateforme *Open Data* de l'État dessine les contours d'une politique nationale de réutilisation des données publiques.

Le discours du Président de la République à l'occasion de l'installation du Conseil National du Numérique le 27 avril 2011¹⁶³ en définit les objectifs : « c'est pour un État plus transparent, une démocratie exemplaire, que j'ai demandé au gouvernement de créer un portail des données publiques en ligne [...]. La question de l'accès aux données de base est une question absolument centrale ».

Le premier objectif est celui de la transparence. « Elle n'est pas sans poser des difficultés [...]. L'accès aux données [des] entreprises est un élément de la démocratie économique, [et] ce n'est pas parce que cela pose des problèmes qu'il ne faut pas le faire. Il faut le faire, c'est un choix et c'est un choix, de mon point de vue, sans retour. »

Le deuxième objectif est l'encouragement à l'innovation. « L'ouverture des données publiques devra permettre à l'écosystème des développeurs et des startups d'inventer de nouveaux services pour les citoyens. »

Un communiqué de presse du Premier ministre à l'occasion de la création de la

mission Etalab¹⁶⁴ précise le cadre de mise en œuvre de cette politique nationale. « Etalab mettra à disposition librement des données brutes dans un format exploitable et développera de nouveaux services en ligne au profit des citoyens, à partir de certaines de ces données. La réutilisation de ces informations offre de larges opportunités à la communauté des développeurs qui pourront les utiliser afin de proposer des services applicatifs innovants. »

(iii) La future plateforme Open Data nationale data.gouv.fr

Le portail « data.gouv.fr » sera la plateforme de mise à disposition des données publiques de l'État, qui devra « permettre un accès simple et identifié pour la consultation et la réutilisation »¹⁶⁵. Le périmètre inclut les administrations centrales et déconcentrées, les établissements publics, et les institutions chargées d'une mission de service public à caractère autre qu'industriel et commercial.

Prémisses d'une stratégie nationale des données ouvertes, la plateforme pourra « rendre accessibles les données de chaque collectivité qui le souhaite ». Cela va dans le sens de la mutualisation d'outils techniques pour faciliter la mise en ligne des données des collectivités en vue de leur réutilisation.

L'objectif de la plateforme data.gouv.fr est triple : « permettre la réutilisation la plus facile et la plus large possible, ensuite encourager l'innovation par toute la communauté des développeurs et des entrepreneurs pour soutenir le développement de l'économie numérique, et enfin contribuer à renforcer la transparence de l'action de l'État en mettant en valeur le travail des administrations ».

La plateforme « mettra à disposition, librement, facilement et gratuitement, le plus grand nombre de données quelle que soit la réutilisation », le tout basé sur « une licence gratuite ».

Toute tarification éventuelle sera strictement encadrée, assurant la gratuité la plus large possible de la réutilisation des données. « Dans certains cas particuliers, la

fonction publique et de la réforme de l'État, porte-parole du Gouvernement :

<http://www.gouvernement.fr/gouvernement/l-administration-electronique>

¹⁶² Décret n° 2011-194 du 21 février 2011 portant création d'une mission « Etalab » chargée de la création d'un portail unique interministériel des données publiques

<http://legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000023619063&categorieLien=id>

¹⁶³ M. le Président de la République, 2011, « Discours à l'occasion de l'installation du Conseil National du Numérique le 27 avril 2011 »

<http://www.elysee.fr/president/les-actualites/discours/2011/discours-sur-internet-et-l-economie-numerique.11256.html>

¹⁶⁴ M. le Premier ministre, 2011, « Création de la mission Etalab, chargée de la mise en ligne de data.gouv.fr », communiqué de presse du 22 février 2011.

http://www.gouvernement.fr/sites/default/files/communiqués/02.22_Communique_de_presse_-_Etalab.pdf

¹⁶⁵ L'Informaticien, 2011, « Interview de Séverin Naudet, Directeur d'Etalab : Encourager l'innovation par toute la communauté des développeurs et des entrepreneurs pour soutenir le développement de l'économie numérique ». Numéro 91, mai 2011.

réutilisation peut faire l'objet d'une redevance, notamment lorsque l'administration doit supporter un investissement exceptionnel lié à cette réutilisation, mais le principe de création d'une redevance sera décidé par décret, c'est-à-dire par le Premier ministre ». Le Premier ministre a précisé la mise en œuvre de cette direction politique à travers une circulaire adressée à l'ensemble des ministres, et un décret encadrant les redevances pour

réutilisation de données publiques, parus le 26 mai 2011.

Data.gouv.fr sera mis en ligne fin 2011, avec « une version beta en ligne en décembre ». Elle sera construite sur « une gouvernance technique ouverte afin que le portail et ses développements soient le plus conformes aux attentes des utilisateurs ».

Partie 3. Comment favoriser la réutilisation des données publiques ?

Chapitre VI. Abaisser les barrières à la réutilisation

Afin d'encourager la réutilisation des données publiques et de permettre la concrétisation des conséquences positives que l'on peut en attendre, il est nécessaire d'abaisser autant que possible les barrières à la réutilisation.

Nous détaillons dans ce chapitre les défis et les difficultés qui peuvent se poser, ainsi que des pistes de solutions. Il s'agit essentiellement de répondre aux contraintes techniques de l'ouverture des données, d'en gérer la complexité juridique, d'anticiper son impact éventuel sur les institutions publiques, et de mettre en place les modèles économiques les plus propices à encourager l'innovation.

1. Répondre aux contraintes techniques

(i) Modèles et formats de données peuvent rendre la donnée plus utile

L'un des objectifs de l'ouverture des données est de stimuler les activités construites à partir sur ces données. Il est crucial de diffuser les données à l'état brut, dans des formats ouverts et réexploitables¹⁶⁶.

La finalité et l'importance des questions de format doivent donc être clairement expliquées, et les processus de collecte et de recensement doivent être établis et assimilés par l'ensemble des opérateurs, ce qui nécessite un effort pédagogique soutenu.

(ii) Ergonomie de l'accès aux données

Une autre contrainte technique tient au besoin d'ergonomie et de simplicité qu'expriment les utilisateurs des outils de diffusion de données.

Le volume de données aujourd'hui hébergé sur certains sites, comme celui de l'INSEE, est tel que de nouvelles problématiques de recherche interactive sont soulevées.

La simplicité est le maître mot en la matière. Ed Parsons¹⁶⁷ de Google, en fait une exigence centrale :

« Les gens veulent accéder à l'information à travers des outils faciles d'emploi. Il faut que ces outils parlent leur langage, ne les oblige pas à utiliser une manière de procéder qui leur soit étrangère, ils veulent juste accéder à l'information dont ils ont besoin. L'infrastructure support est pour l'essentiel invisible pour les utilisateurs, tout comme celle qui est derrière une prise électrique ou derrière un robinet : elle est invisible derrière la page d'accueil de Google. »

Rendre simple et efficace un service en ligne exige d'anticiper les difficultés éventuelles d'ergonomie et de design d'interaction dès la phase de conception, en soumettant le projet des utilisateurs test. Cela implique aussi de gérer la complexité durant les phases ultérieures, pendant le développement comme lorsque l'outil fonctionne en production.

(iii) La complexité des données géographiques au cœur des évolutions futures du secteur

Les technologies géomatiques¹⁶⁸ permettent de superposer sur des fonds de carte des données cartographiques diverses, d'ajouter des couches d'informations variées (physiques, économiques, environnementales), et surtout de visualiser ces phénomènes et leur évolution en représentation spatiale, sur lesquelles le cerveau humain est capable d'inférences rapides et complexes.

Globalement, l'information géographique numérique est devenu un outil puissant d'analyse de données. Elle constitue aussi une donnée structurante pour la réutilisation de nombreuses autres données publiques ou pour la production de nouveaux services.

¹⁶⁶ Description formelle de la représentation des données dans la base de données. Voir <http://www.ac-grenoble.fr/ecogest/pedago/administration/prodpeda/mcd.htm>.

¹⁶⁷ Idib.

¹⁶⁸ Voir la plaquette « L'information géographique : l'expérience française » publiée le 21 mai 2008 par le Ministère de l'écologie, de l'énergie, du développement durable et de l'aménagement du territoire.

Traditionnellement, les cartes et les informations sur ces cartes étaient rangées séparément dans des cartothèques. La même organisation conceptuelle est reproduite sur les données géographiques numériques avec des efforts importants sur la structuration des données et des métadonnées.

Inspire, le portail environnement, le Géoportail couplé au Géocatalogue, les normes AFNOR, CEN et ISO sur l'information géographique et sur les métadonnées en sont autant d'illustrations¹⁶⁹. Construire une infrastructure capable de gérer la complexité de l'information géographique tout en préservant la simplicité de l'interface est crucial pour faciliter la réutilisation des données.

(iv) Les nouveaux outils du Web des données

Face à la croissance toujours plus rapide des capacités de calcul et de stockage des centres de calcul des grandes entreprises du Web, le monde fait aujourd'hui l'expérience d'un véritable « déluge des données¹⁷⁰ ».

On estime que l'humanité, qui aurait produit 150 Exaoctets (150 milliards de Gigaoctets, soit à peu près 40 milliards de DVD) en 2005, créera 1,200 Exaoctets de données numériques nouvelles en 2011. Cette échelle de quantité de données à stocker et analyser crée de nouveaux défis et de nouvelles opportunités technologiques.

L'apparition du *cloud computing*¹⁷¹ et le développement d'outils de traitement algorithmique des données à très grande échelle (par exemple Map/Reduce, Hadoop, et les outils *Big Data*¹⁷²) vont permettre le stockage, le traitement et l'analyse de pans de données numériques d'une taille inimaginable jusqu'alors¹⁷³. « Tout ce que nous connaissions du Web va changer à nouveau », et ces outils permettront d'amener des réponses aux difficultés liées au volume des données.

¹⁶⁹ Le premier portail de Yahoo recensait les sites les plus populaires, organisés selon leurs contenus. Il procédait ainsi à une séparation analogue entre donnée et métadonnée.

¹⁷⁰ The Economist. 25 février 2010. « The Data Deluge : Businesses, Governments and Society are Only Starting to Tap Its Vast Potential »

¹⁷¹ http://fr.wikipedia.org/wiki/Cloud_computing

¹⁷² ReadWriteWeb. 2010. *The Age of Exabytes : Tools & Approaches for Managing Big Data*. <http://www.readwriteweb.com/reports/big-data/>

¹⁷³ Henri Verdier. 14 décembre 2010. « Big Data : Making sense at scale. » <http://www.henriverdier.com/2010/12/big-data-making-sense-at-scale.html>

(v) Web Sémantique : l'avenir du Web ?

En représentant les données sous forme de graphe de relations entre entités plutôt que de tables structurées, les technologies du Web sémantique et des « données ouvertes liées » (*Linked Open Data*) permettent aux machines de donner un « sens » à l'information disponible en ligne¹⁷⁴. On parle d'ontologie pour décrire la structure des différents types d'entités représentées et les concepts de relations qui les relient.

Malgré l'enthousiasme d'une partie de la communauté pour ces technologies sémantiques, elles soulèvent des questions techniques et financières difficiles à ignorer pour les administrations publiques.

Une première étape dans l'intégration des données issues du secteur public à un hypothétique Web des données consiste à mettre en ligne les données sous forme brute, et à permettre aux membres de la communauté qui souhaiteraient s'impliquer de réaliser eux-mêmes la sémantisation des jeux de données.

Au cours d'une présentation qu'il a effectuée à la conférence TED¹⁷⁵ en 2009, Tim Berners Lee a lancé un vibrant appel à la mise en ligne des données brutes au plus tôt de par le monde : « *We want raw data now !* »

La mise en ligne de données brutes est une approche pragmatique qui permet d'accélérer l'ouverture des données publiques, et une première étape vers la construction d'un Web des données.

2. Gérer la complexité juridique

(i) Les producteurs de données ont développé plusieurs licences

Selon la loi de 1978, le principe général est celui de la réutilisation libre, gratuite et sans conditions des données publiques. Les licences ne sont obligatoires que dans les cas où l'administration souhaite imposer une réutilisation payante de ses données.

Néanmoins, de nombreux acteurs considèrent que ce régime ne garantit pas une réelle sécurité juridique, et attendent des pouvoirs publics qu'ils explicitent les usages

¹⁷⁴ Segaran, T. 2009. *Programming the Semantic Web*. O'Reilly éditions.

¹⁷⁵ TED (Technology Entertainment and Design) est un ensemble de conférences lancées en 1984 par le secteur privé à but non lucratif américain pour diffuser des "idées qui méritent d'être répandues ».

autorisés sous forme de licences, y compris dans le cas d'une réutilisation gratuite.

Schématiquement il existe cinq possibilités sur les conditions d'accès aux données publiques :

- Sur demande : l'accès peut être accordé dans certains cas et sous conditions.
- Licence payante : l'accès aux données est assujéti à un paiement.
- Accord exclusif d'accès avec une seule entreprise déléguée pour diffuser les données: en cas d'investissements importants réalisés par le privé mais pour une durée limitée seulement, peut être le cas dans le domaine culturel.
- Licence ouverte sous conditions : l'accès est ouvert mais la diffusion est soumise à conditions.
- Licence ouverte sans restriction : l'accès et la réutilisation sont libres, c'est le cas des données entrées dans le domaine public ou, par défaut, des données publiques ouvertes sans licence.

De nombreuses licences ont déjà été élaborées par des services publics, certaines pouvant être contractualisées par un simple clic.

L'Agence pour le Patrimoine Immatériel de l'État (APIE), une agence de conseil du ministère de la Réforme de l'État et des Comptes Publics en a développé deux pour livraison de données soumises à redevance (une pour téléchargement simple et l'autre pour livraisons régulières). La Direction de l'Information Légale et Administrative a mis en place des licences payantes pour la réutilisation de certaines de ses données par ailleurs accessibles gratuitement, comme les débats « questions-réponses » de l'Assemblée nationale, ou les téléchargements de masse depuis Légifrance (avec livraison au format XML sur un site dédié).

Dans le monde des contenus culturels, un certain nombre de licences standards ont émergé des travaux sur l'évolution du copyright face aux nouveaux usages de partage, de *mashups* ou de *remix*¹⁷⁶. Les différentes licences *Creative Commons*, qui définissent des conditions différentes (avec ou sans usage commercial, avec ou sans partage à l'identique, attribution, etc.) sont par exemple

très utilisées sur Internet, notamment par Wikipedia.

Les licences pour réutilisation de données comportent en général des clauses plus techniques liées à la nature d'une base de données, qui peut être différente d'un contenu média. Un certain nombre de licences « *Open Data* » standard pour réutilisation de données s'inspire cependant largement de *Creative Commons*. La licence ODbL¹⁷⁷ (*Open Database License*) reprend notamment des notions de partage à l'identique et d'attribution, et permet l'accès, l'utilisation, le téléchargement, la copie, le partage et la distribution des données publiques ouvertes.

Le ministère de la Justice a élaboré une licence « information publique librement réutilisable » appelée LIP¹⁷⁸. Elle précise les droits d'adaptation des informations publiques en vue d'une activité commerciale ou non. L'adaptation et la modification sont autorisées à condition qu'il y ait enrichissement documentaire, technique ou éditorial et que le sens ne soit pas dénaturé. La rediffusion et la vente sont possibles si les données ont subi un nouveau traitement et qu'elles sont intégrées dans un produit ou un service nouveau auprès des tiers.

(ii) Une complexité liée à l'accumulation de licences

Toutes ces licences, bien qu'elles soient fondées sur les mêmes principes généraux, ne sont pas nécessairement compatibles les unes avec les autres. Or, les données n'ont de valeur pour les usagers finaux des services qui les réutilisent que si elles sont partagées, recoupées et réutilisées.

Certaines données sont disponibles à titre gratuit et d'autres à titre payant, certaines sont soumises à des droits, d'autres sont en libre accès au public, d'autres encore portent des clauses d'attribution parfois lourdes. Dans ce contexte parfois contradictoire, connaître précisément les conditions de réutilisation de données produites par recombinaisons successives peut être complexe.

L'accumulation de licences différentes pour des jeux de données distincts crée une complexité et un risque juridique potentiellement coûteux. Il apparaît important d'éviter la prolifération de différentes licences

¹⁷⁷ Open Database License

<http://www.opendatacommons.org/licenses/odbl/>

¹⁷⁸ Licence Information Publique

http://www.rip.justice.fr/information_publicque_librement_reutilisable

¹⁷⁶ Lessig, 2009. Idib.

proposant des dispositions ne variant que de façon marginale, et de privilégier l'utilisation la plus large de licences standards, idéalement aux standards de l'*Open Data*.

(iii) Les différences de droit entre pays, source de confusion

Le droit à la réutilisation, ses limites et son interaction avec les droits de propriété intellectuelle n'est pas uniforme entre les pays. Des différences marquées apparaissent entre pays de droit civil et pays de *common law*.

Ainsi, reprendre une œuvre française mise sous une licence *creative commons* donnée n'offre pas les mêmes avantages que si cette œuvre était américaine. Le droit de réutilisation des données, et les droits couvrant les bases de données elles-mêmes sont d'autant plus complexes que les données proviennent de pays différents.

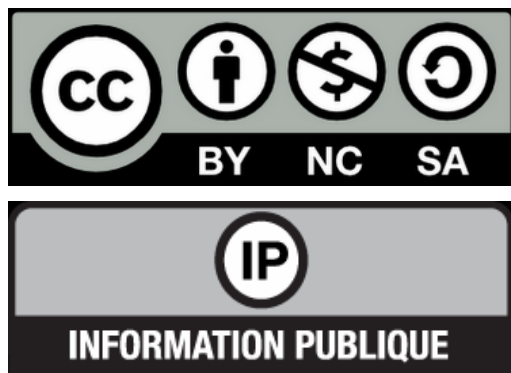
Ces confusions peuvent nuire à la facilité de réutilisation et créent un risque juridique. L'émergence de standards internationaux de licences compatibles doit devenir une priorité.

(iv) Des pictogrammes pour des conditions de réutilisation plus explicites

Il est primordial, pour assurer la clarté juridique des conditions de réutilisation, de permettre aux réutilisateurs d'en prendre connaissance simplement.

La Fondation *Creative Commons* a mis au point à cet effet un certain nombre de pictogrammes permettant d'identifier facilement les conditions spécifiques de réutilisation d'un contenu couvert par une telle licence.

C'est aussi dans cette perspective de lisibilité que Thomas Saint-Aubin, au sein du ministère de la Justice, a conçu le pictogramme « Information Publique », signalant un jeu de données disponible en licence LIP.



(v) Des conditions de licences inspirées de « Creative Commons » et du « share alike »

Lorsque les données produites sont commercialisées, il peut arriver que leur mise à disposition à titre gratuit pour une réutilisation commerciale ne soit pas possible, notamment pour des raisons de distorsion de concurrence sur des marchés existants.

Il est par contre souvent possible de mettre à disposition les données pour réutilisation gratuite même dans un but commercial, mais avec une restriction de taille, qui est celle du « *share alike* » – le partage des données dans des conditions identiques.

Ce type de licences gratuites spécifiques permettent d'adapter, de modifier, de transformer et de construire d'autres données à partir de celles mises à disposition, à condition de reverser au profit de l'ensemble de la collectivité les données produites, dans des conditions identiques c'est-à-dire ouvertes (redistribution imposée sous la même licence).

Mettre en place une licence gratuite de type « *share alike* » n'exclut absolument pas de maintenir la tarification de la réutilisation des données hors de ce cadre. Les autres réutilisateurs peuvent toujours accéder à une licence payante non exclusive.

Certains interlocuteurs au sein de la communauté des réutilisateurs considèrent de telles clauses comme contraires à l'esprit de l'*Open Data*. D'autres y voient une approche pragmatique permettant de concilier des finalités différentes.

3. Anticiper l'impact organisationnel

(i) Les redondances locales révèlent le besoin d'une infrastructure nationale

L'étude de cas présentée au chapitre V sur le partage de l'information géographique dans les régions montre que des recoupements au moins partiel des initiatives locales de partage des données géographiques publiques existent. L'État, qui donne l'impulsion d'une infrastructure nationale pour l'information géographique et se trouve au centre des initiatives ou au moins associé à celles-ci, pourrait rechercher plus systématiquement des synergies entre elles.

(ii) Privilégier la rapidité d'exécution

Dans le développement de projets de systèmes d'information du secteur public, on constate souvent une recherche de qualité des données et d'exhaustivité du service offert. Cela se fait au détriment de la rapidité de mise en œuvre.

Même si l'intention est louable et compréhensible, elle aboutit en pratique à des délais souvent si longs que le service, dépassé par la courbe technologique, ne répond plus à la demande avant même sa sortie.

Il est largement préférable de privilégier la rapidité d'exécution et d'adopter une démarche de développement agile. La mise en ligne rapide d'une version bêta avec un petit nombre de jeux de données permet d'illustrer rapidement les enjeux et d'enclencher le cercle vertueux qui créera les conditions favorables à un développement rapide du projet.

(iii) Les freins classiques au changement restent à l'œuvre

Le savoir étant souvent associé au pouvoir, des freins à l'ouverture des données apparaissent naturellement dans les organisations, et s'ajoutent aux résistances qu'un changement aussi profond que l'*Open Data* ne peut manquer de créer.

L'apparition de nouveaux processus au sein de l'administration du fait de la collecte et du recensement des données nécessite des compétences propres et des formations adéquates. C'est un point essentiel pour la réussite de la construction d'une nouvelle relation entre le producteur et les réutilisateurs éventuels des données.

L'animation conjointe de la communauté de producteurs et de réutilisateurs, que ce soit par l'organisation de concours de développement, d'ateliers de travail ou de *Bar Camps*, doit permettre de rapprocher ces communautés et de catalyser les efforts de mise en ligne et de réutilisation.

Cela soulève la question de la place de l'innovation au sein des organisations. Par exemple l'INSEE, considérant qu'il est trop tôt pour restructurer l'organisation dans son ensemble autour de la question du web sémantique¹⁷⁹, s'est investi dans le projet *Data Lift*¹⁸⁰ de l'INRIA pour adapter au web sémantique certaines de ses données.

(iv) Le coût de la mise à disposition est compensé par les bénéfices induits

Un élément récurrent du débat sur l'ouverture des données publiques est celui du coût supposé démesuré qu'elle impliquerait pour la puissance publique.

Notre étude sur la mise en commun des données géographiques dans les régions permet de faire émerger quelques ordres de grandeur.

Les moyens en investissement des initiatives de mise en ligne de données correspondent essentiellement aux dépenses liées aux serveurs, soit pour les portails géomatiques des régions une moyenne de 200 000 euros par initiative. Les coûts de fonctionnement moyens sont de 50 000 euros par an et de 4 équivalents temps plein.

Pour le plus grand projet de mise à disposition de données géographiques en France, le Géoportail de l'IGN, qui répond aux obligations de la directive INSPIRE, 22 agents sont dédiés à son activité spécifique de service public à l'exclusion de toute activité commerciale. Les coûts en investissement s'élèvent à 6 millions d'euros, et le coût annuel de fonctionnement est de 2 millions d'euros.

Ces coûts nous semblent largement compensés par les avantages de l'ouverture des données. L'État a d'ailleurs montré qu'il était prêt à les couvrir, voire à compenser le manque à gagner commercial qu'un passage à la gratuité pouvait engendrer lorsque les bénéfices le justifiaient, comme ce fut le cas avec le passage du RGE de l'IGN à la gratuité pour le secteur public en 2010 – et ce d'autant plus que dans ce cas, le produit de la commercialisation provenait du secteur public lui-même.

La mutualisation des plateformes par les collectivités, voire la mise en ligne de leurs données sur data.gouv.fr, devrait permettre de dépasser ce débat.

¹⁷⁹ Entretien du 26 novembre 2010, *ibid.*

¹⁸⁰ Projet de recherche sur l'interconnexion des jeux de données sémantiques. <http://datalift.org/>

4. Établir des modèles économiques propices à l'innovation

Cette section aborde et décrit les caractéristiques particulières de la donnée publique en tant qu'objet économique, les prescriptions normatives de la théorie, et les conséquences que l'on peut en tirer quant au modèle économique le plus à même d'encourager la réutilisation des données et de permettre de réaliser les bénéfices que la société peut en attendre.

4.1. Les données publiques comme objet économique

(i) Particularités

Du fait de leur caractère immatériel, les données publiques présentent des particularités par rapport aux biens classiques, même informationnels. Leur non-rivalité revêt une importance technique, au sens où leur coût de reproduction est quasi nul. Leur coût de distribution est donc très faible lui aussi, car limité au coût de bande passante, très largement négligeable, et de stockage, qui reste très modeste des bienfaits issus de la réutilisation des données.

En sens inverse par contre, les coûts fixes liés à la production de données sont significatifs à chaque étape de la chaîne de valeur.

Si le coût marginal de production et de distribution est quasi nul, se pose en revanche la question de la couverture des coûts fixes. C'est là que repose toute la complexité de la question de la tarification.

Un autre caractère particulier de la donnée publique tient au fait qu'elle présente un très fort potentiel de réutilisation au sein de biens et services fortement hétérogènes. Par opposition à un bien physique, souvent destiné à un usage spécifique une fois produit, une donnée publique peut servir à un éventail virtuellement infini d'applications.

Cela pose un problème concret à l'analyse économique des évolutions du marché pour une donnée particulière, car rien ne permet de savoir si elle constitue un marché mature ou si son potentiel de réutilisation innovante est encore latent.

Surtout, les données publiques ont la caractéristique d'être avant tout des facteurs de production d'autres biens ou (majoritairement)

services. Ce n'est que très rarement la donnée en tant que telle qui a une valeur directe pour le consommateur ou l'entreprise, mais l'information qu'elle contient, l'analyse qu'elle permet d'étayer, l'objet qu'elle permet de construire, l'application qu'elle permet de développer.

L'évolutivité des techniques rend impossible la prévision, même dans les grandes lignes, des usages tirés de l'ouverture des données qui rencontreront un succès commercial. Il ne faut donc pas restreindre les données susceptibles d'être ouvertes en fonction d'une analyse *a priori* de leur intérêt pour les réutilisateurs.

Ne mesurer que la demande directe des données par les réutilisateurs, c'est ignorer l'effet multiplicateur de la demande de services exprimée ou latente chez les utilisateurs.

L'analyse du modèle économique des données doit prendre en compte les externalités positives découlant de la création de nouveaux services et de l'incitation à l'innovation portée par l'ouverture la plus large des données publiques.

(ii) Modèles économiques des données publiques

Les exercices de modélisation économique de la production et de la réutilisation des données publiques se limitent en général à trois types de coûts : les coûts de production de la donnée brute, les coûts de transformation (de données « *upstream* » ou « brute » en donnée « *downstream* » ou « à valeur ajoutée »), et les coûts de diffusions.

Dans la réalité, il faut aussi bien sûr tenir compte des coûts de transaction, qu'ils soient fixes (mise en place d'un moyen de paiement) ou variables (frais de transactions financières à chaque paiement de redevance, et temps de travail nécessaire à leur réalisation). L'expérience empirique sur la tarification des données montre que le coût de mise en œuvre d'une infrastructure de paiement est souvent à peine compensé par les revenus générés¹⁸¹, ce qui doit pousser à restreindre le plus possible l'établissement de redevances pour réutilisation.

Le point crucial à rappeler est que la donnée, avant d'être un bien dont les consommateurs peuvent bénéficier directement ou indirectement, est l'outil de travail fondamental de l'administration. Il faut éviter

¹⁸¹ Table ronde avec des collectivités locales à l'Open Government Data Camp à Londres le 18 novembre 2010.

la multiplication au sein des administrations d'un travail de raffinement des données à destination d'une activité de vente, pour deux raisons : d'abord cela crée des risques de distorsions majeures sur les marchés, comme noté par Joseph Stiglitz (prix Nobel d'économie) et Peter Orszag (ancien ministre du Budget américain d'Obama) dans le rapport de référence sur le rôle du gouvernement dans l'ère numérique¹⁸² ; et si des données sont de qualité suffisante pour le travail de l'administration, il ne doit pas y avoir lieu de les transformer en profondeur pour les rendre utiles à des réutilisateurs.

Il faut donc bien définir les coûts dont la couverture nous amène à rechercher un modèle économique optimal : il ne s'agit non pas de l'ensemble des coûts de production, mais de ceux générés *du fait et seulement du fait* de la mise à disposition d'un ensemble de données.

(iii) Doctrine des ressources essentielles en droit de la concurrence

Un point important qui élargit l'enjeu de la tarification à celui de l'accès aux données est son incidence sur des problèmes de concurrence imparfaite. Les conflits concurrentiels qui peuvent apparaître dans un monde où l'administration a le monopole structurel de la production d'un certain nombre de données importantes ont été notés dès les prémices des réflexions sur l'accès aux données publiques dans les années 1990^{183 184}.

L'importance du sujet a conduit le Premier ministre Lionel Jospin¹⁸⁵ à répéter alors, lors du discours d'Hourtin de 1997, la doctrine des ressources essentielles et son application aux données publiques.

Selon cette doctrine de droit de la concurrence, toute ressource dont la réutilisation est essentielle à l'établissement d'un marché efficace et concurrentiel, mais qui peut en empêcher la formation si elle n'est pas accessible, doit être régulée de manière à

permettre l'émergence de ce marché et son bon fonctionnement.¹⁸⁶

En particulier, il est important de noter que cette doctrine a vocation à s'appliquer aussi au secteur privé, et l'on peut imaginer qu'elle fasse un jour jurisprudence pour permettre l'accès à des données certes de droit privé, mais constituant une ressource essentielle pour des intérêts de service public.^{187 188}

4.2. Choix optimal de tarification et hypothèses économiques

(i) Trois types de tarification permettent de traiter le problème des coûts fixes

Les études économiques retiennent en général trois modèles de tarification pour couvrir les coûts fixes découlant de la mise à disposition de données publiques.

Le modèle monopolistique correspond à un modèle où l'objectif est la maximisation des revenus directs pour l'institution productrice. Il participe souvent de l'exercice d'une position dominante, car seule la puissance publique a ce à et peut produire un certain nombre de données. Les distorsions et inefficacités que ce comportement de monopole peut créer sont bien connues et particulièrement fortes selon la théorie.

Le modèle de coût moyen suppose la possibilité pour un producteur d'observer la courbe de demande de ses réutilisateurs, et de fixer un prix qui équilibre exactement ses revenus directs issus des redevances avec ses coûts d'ouverture et de diffusion des données.

Enfin, le modèle de coût marginal, dans les faits celui de la gratuité, recherche l'optimum d'efficacité économique, que la science économique détermine comme atteint lorsque le prix d'un bien (ici le montant de la redevance) est égal au coût marginal de fourniture du bien, c'est-à-dire d'une unité supplémentaire (ici très proche de zéro car la donnée est immatérielle). Il fait par contre porter les coûts fixes de production et diffusion des données sur le budget de l'institution productrice, c'est-à-dire indirectement sur le réutilisateur via l'impôt (avec les incidences en terme de redistribution que cela peut impliquer).

¹⁸² Joseph Stiglitz et Peter Orszag. Octobre 2000. *The Role of Government in a Digital Age*.

http://archive.epinet.org/real_media/010111/materials/stiglitz.pdf

¹⁸³ Entretien avec Maurice Ronai.

¹⁸⁴ Maurice Ronai. 1996. *Données publiques : accès, diffusion, commercialisation*. Revue Problèmes Politiques et Sociaux : Dossiers d'Actualité Mondiale. Nos 773-774, 1^{er} novembre 1996. La Documentation Française.

¹⁸⁵ Lionel Jospin. Discours à Hourtin le 25 août 1997.

Préparer la France pour la Société de l'Information.
<http://www.admiroutes.asso.fr/action/theme/politic/lionel.htm>

¹⁸⁶ Entretien avec Pierre-Jean Benghozi le 12 octobre 2010.

¹⁸⁷ Entretien avec Maurice Ronai, auteur du rapport à la documentation française de 1996 sur le sujet.

¹⁸⁸ Entretien le 12 novembre 2010 avec Guillaume Crouigneau, directeur de CanalTP, filiale de la SNCF.

La littérature sur le sujet est encore réduite, mais trois études en particulier comparent systématiquement les trois modèles, leurs avantages et leurs inconvénients :

« *Models of provision of Public Sector Information by trust funds* » de l'Université de Cambridge¹⁸⁹ qui fait autorité auprès du gouvernement britannique (Newbery et al.), l'étude Bureau d'Économie Théorique et Appliquée de l'Université de Strasbourg sur la valorisation des informations du secteur public réalisée en collaboration avec l'Agence pour le Patrimoine Immatériel de l'État¹⁹⁰, et « *Economics of Public Sector Information*¹⁹¹ », l'étude complémentaire à l'étude de Cambridge, menée fin 2008 par Rufus Pollock, mathématicien et économiste à l'Université de Cambridge et nommé par le Premier ministre David Cameron au Public Sector Transparency Board.

(ii) Une référence : "Economics of Public Sector Information" de Rufus Pollock

L'étude de Rufus Pollock est la référence la plus largement citée et la plus reconnue. Elle aborde l'ensemble des problèmes théoriques liés à l'économie des données publiques, rassemble les résultats empiriques connus, et propose des solutions de financement alternatives tels que le paiement à l'enregistrement des données.

Le cœur de l'étude détaille quatre types de problèmes majeurs qui apparaissent avec les modèles de tarification cités précédemment¹⁹² : les questions de crédibilité du producteur quand à la pérennité de sa tarification (le risque étant la capture de valeur injustifiée par la puissance publique une fois de nouveaux modèles commerciaux découverts), les incitations à la production et à la réutilisation, les distorsions de compétition, et les problèmes d'asymétrie d'information.

Le modèle de coût moyen pose des problèmes insolubles d'asymétrie d'information, car il est irréaliste de penser qu'un producteur puisse observer exactement

le point sur une courbe de demande qui lui permette d'équilibrer son budget.

Au contraire, le modèle de coût marginal, ou la gratuité, qui aurait cours sur un marché en concurrence pure et parfaite, s'abstrait de cette contrainte irréaliste.

Mais c'est sur le sujet des distorsions de concurrence que les modèles de coût moyen et monopolistique présentent des risques d'inefficacités extrêmes. Ces risques proviennent d'abord des positions dominantes par nature des acteurs publics dans la production de données publiques, mais aussi de la sous-optimalité de ces deux modèles qui mettent en place une structure distordue de signaux d'incitations.

Au contraire, le modèle de tarification marginale aligne exactement le prix de la licence de réutilisation des données sur le coût marginal d'ouverture et de diffusion. Selon le modèle de la Pareto-efficience, c'est donc la solution optimale pour la société.

Pour Pollock, le modèle de la gratuité s'impose donc et doit être généralisé. La seule question qui peut rester ouverte elle celle de savoir, soit dans quels cas théoriques le modèle de coûts moyens devrait prévaloir sur le modèle en coûts marginaux, soit dans quelles exceptions ces coûts marginaux deviennent non négligeables et imposent une tarification modeste autre que la gratuité.

(iii) L'« étude de Strasbourg » révèle les difficultés liées aux hypothèses économiques

L'étude récente menée par Julien Pénin et quatre coauteurs au Bureau d'Économie Théorique et Appliquée de l'Université de Strasbourg, en partenariat avec l'Agence pour le Patrimoine Immatériel de l'État (APIE), se propose d'aborder le sujet sous l'angle de la valorisation des informations du secteur public¹⁹³.

Ses conclusions s'approchent dans ses grandes lignes de celles de la littérature existante, à savoir que la tarification devrait être gratuite pour tous les usages personnels, que le modèle de tarification au coût marginal est optimal pour le plus grand nombre de réutilisations commerciales. Elle affirme

¹⁸⁹ David Newbery, Lionel Bently, Rufus Pollock. 26 février 2008. « Models of Public Sector Information Provision via Trading Funds. » Cambridge University. <http://www.berr.gov.uk/files/file45136.pdf>

¹⁹⁰ APIE. 12 janvier 2011. « Etude économique ». https://www.apiefrance.fr/sections/acces_thematique/reutilisation-des-informations-publiques/etude-economique/view

¹⁹¹ Rufus Pollock. 2 décembre 2008. « The Economics of Public Sector Information ». University of Cambridge. http://www.rufuspollock.org/economics/papers/economics_of_psi.pdf

¹⁹² Tableau page 16 dans Pollock, 2008.

¹⁹³ Julien Pénin et al. 12 janvier 2011. « Rapport Final : La valorisation des informations du secteur public (ISP) : un modèle économique de tarification optimale ». Université de Strasbourg. https://www.apiefrance.fr/sections/acces_thematique/reutilisation-des-informations-publiques/etude-economique/downloadFile/attachedFile/Rapport_final_BE_TA.pdf?nocache=1294846257.31

cependant qu'il existe des cas dans les marchés matures avec forte contrainte budgétaire où la tarification au coût moyen peut aussi être optimale.

L'étude révèle par contre des difficultés liées à la compréhension des enjeux concrets de la réutilisation des données.

Les auteurs considèrent tout d'abord que le but d'une politique de réutilisation est la diffusion large d'informations et de savoirs (c'est-à-dire, selon leurs termes, « d'informations enrichies »).

Dans la mesure où c'est l'application des données plutôt que les données elles-mêmes qui ont une valeur pour les citoyens et les consommateurs, c'est au contraire à la diffusion la plus large des *bénéfices* issus des données qu'il faut viser. Le but d'une politique de réutilisation devrait donc de préférence porter sur la diffusion large *auprès du public des réutilisateurs* des données brutes, dont « l'état brut sans explications (...) ne permet (certes) pas la rediffusion au plus grand nombre » mais est en revanche exactement la qualité recherchée par les développeurs pour encourager à la réutilisation¹⁹⁴.

Cela indique que la demande usuellement supposée serait sous-estimée face à la demande actuelle du public, encore mal comblée, et sa demande latente pour des besoins que les entrepreneurs chercheront à découvrir. L'absence de prise en compte d'un multiplicateur sur les fonctions de demande pourrait remettre en cause les conclusions de l'étude sur l'intérêt du modèle de coût moyen.

L'autre mécompréhension manifeste est la question de savoir quels coûts sont engagés *en plus, du fait et du seul fait* de la mise en ligne de données à fins de réutilisations.

Si l'administration, dans le cadre de ses activités, éprouve le besoin pour elle-même de raffiner les données en leur ajoutant une structure ou de l'information et des métadonnées, c'est alors cette « information enrichie » qui constitue la donnée brute au sens de la réutilisation. Prendre en compte ce coût de mise en forme dans l'établissement d'une redevance revient à faire porter sur les réutilisateurs une partie des coûts de production des données nécessaires au fonctionnement normal de l'administration, alors qu'ils sont déjà financés par le budget de l'État. Cela serait injuste, inefficace, et distorsif sur le marché.

Enfin, l'étude n'aborde qu'en conclusion la question de la mise à disposition des données brutes nécessaires à l'élaboration des informations du secteur public tarifées. Si l'État propose une activité de type concurrentiel à partir des données publiques sans mettre à disposition du marché les données brutes qui sous-tendent son activité, il risque d'interdire l'accès d'entreprises à un marché concurrentiel, ce qui contredit directement la doctrine des biens essentiels. Il en résulterait une position de monopole injustifiée et inefficace.

En ce sens, il est possible d'interpréter les résultats de cette étude comme un soutien à la généralisation la plus large de la gratuité de réutilisation des données publiques.

(iv) Un consensus sur la tarification optimale apparaît en précisant les hypothèses

Les différentes études font émerger un consensus des sciences économiques sur les mécanismes à l'œuvre, si l'on précise bien les hypothèses sur lesquelles se basent les modèles mathématiques utilisés.

Le choix d'une hypothèse de données de type brut (*upstream*) ou à valeur ajoutée (*downstream*) est important. Pour la communauté technique, ce sont ces presque exclusivement ces données *upstream* sous formats bruts qui intéressent les créateurs de services numériques, et ce sont celles-là qu'il convient de rendre gratuites.

La prise en compte ou non de la contrainte budgétaire dans les modèles mathématiques peut aussi faire pencher vers un modèle de coûts moyens dans certaines exceptions. C'est une vraie contrainte qui peut se manifester concrètement. On peut cependant remettre en cause, comme le font Tirole et Laffont¹⁹⁵, la crédibilité de l'hypothèse d'un régulateur qui se justifierait par des contraintes créées de son propre fait pour ne pas appliquer le modèle économique le plus efficace pour tous.

Enfin, les modèles doivent faire l'hypothèse de coûts de transactions élevés, vérifiée empiriquement, qui rend irréaliste une tarification pour la grande majorité des ensembles de données. Pour la grande majorité des jeux de données, toute tarification serait rendue non rentable par les coûts de transaction. La gratuité s'impose alors.

Dans la pratique, les évolutions rapides du contexte technologique et les incertitudes

¹⁹⁴ Table ronde entre collectivités et développeurs à la conférence ePSI de Rennes, 29 novembre 2010.

¹⁹⁵ Laffont, J. et Tirole, J. 1993. « A Theory of Incentives in Procurement ». MIT Press.

inhérentes à l'innovation poussent à généraliser le plus possible la gratuité afin d'encourager la création de nouvelles activités à partir des données publiques.

Plusieurs des acteurs à qui nous avons pu parler estiment que la majorité des données doit être réutilisable gratuitement même à titre commercial.

Il s'agit alors de délimiter clairement les cas dans lesquels il est légitime d'adopter une tarification. L'exercice d'analyse économique permet de dessiner les contours d'une politique efficace de tarification de la réutilisation des données publiques.

4.3. Un modèle légitime de tarification efficace

(i) La réutilisation des données doit être gratuite par défaut

Par défaut, la réutilisation d'une donnée publique ne doit pas être tarifée, ni lorsque la réutilisation se fait à titre personnel, ni lorsqu'elle se fait à titre commercial.

C'est particulièrement crucial pour les marchés les moins matures où l'innovation liée aux données publiques est émergente. Le choix de la gratuité est raisonnable dans ce cas, et notamment du point de vue du développement des entreprises innovantes.

Pour produire les effets les plus positifs sur les acteurs et la dynamique économique, la politique de réutilisation des données publiques privilégiera la gratuité par défaut.

(ii) Des cas légitimes de redevance existent

Pour des données dont le coût marginal d'ouverture et de distribution est très important (très forte taille de fichiers par exemple), pour des données dont la diffusion implique de très forts coûts fixes par rapport à une demande faible des réutilisateurs, ou des données à forte valeur ajoutée dans des marchés matures et établis, il peut être légitime d'adopter une tarification, qui selon le droit et la jurisprudence ne saurait être excessive en regard de ces coûts.

Nous préconiserons donc de limiter strictement les redevances pour réutilisation à ces cas bien identifiés, au risque sinon de brider l'innovation émergente dans le secteur des données publiques, et les bénéfices que la société tout entière peut en attendre.

(iii) Donner accès aux données brutes sous-jacentes

Le rôle de l'État dans l'ère numérique est de mettre en place les structures qui permettent le fonctionnement efficace et juste des marchés, mais en aucun cas de préjuger des préférences des acteurs de ce marché.

Pour éviter que les réutilisateurs soient contraints d'utiliser les données à valeur ajoutée tarifées même lorsqu'ils ne sont intéressés que par la donnée brute sous-jacente, il faut assurer la possibilité d'accéder, gratuitement, à ces données brutes à chaque fois qu'une redevance est établie pour réutilisation de données à valeur ajoutée.

L'application de ce principe, outre la création de valeur qu'il permet sur le marché, ne peut qu'inciter l'acteur public concerné à améliorer le service et la valeur ajoutée qu'il apporte dans ses données *downstream*. Une telle émulation provoquée par la concurrence sera positive à la condition impérative que l'acteur public joue pleinement le jeu de l'ouverture des données *upstream*. À défaut, l'intervention d'un régulateur s'imposerait.

4.4. L'accès aux données publiques encourage le développement des jeunes entreprises

(i) L'entrepreneuriat comme processus d'exploration

Les développements récents menés par Éric Ries, entrepreneur, auteur et intervenant à la Harvard Business School et la Stanford Graduate School of Business, ont amené à repenser l'entrepreneuriat comme une activité d'exploration¹⁹⁶.

En particulier, une jeune pousse, plutôt qu'une version de petite taille d'une grande entreprise, est définie comme une organisation temporaire chargée de développer la solution commerciale et technologique qui répond le mieux à un problème supposé, mais qu'il faut découvrir, comprendre et confirmer, ou infirmer et explorer à nouveau. C'est presque littéralement d'une expérience scientifique qu'il s'agit, et le développement des startups technologiques dépend de la réussite d'un processus d'exploration.

Dans ce contexte, tarifier l'accès à un grand nombre de données publiques pour des réutilisations commerciales risquerait de réduire fortement l'opportunité de développer

¹⁹⁶ Concept de jeune pousse maigre ou « *Lean Startup* ». <http://www.startuplessonslearned.com/>

un nouveau tissu de jeunes entreprises innovantes autour de ce domaine.

(ii) Les coûts liés à la tarification freinent l'expérimentation

La nécessité d'entrer dans des négociations légales et la perspective de devoir s'acquitter de droits d'exploitations conséquents avant d'avoir expérimenté la viabilité commerciale d'un produit est une barrière dirimante pour nombre de jeunes sociétés.

Elle peut être insurmontable pour un développeur indépendant ou employé par ailleurs et qui songerait à lancer une activité nouvelle sans disposer de capitaux importants.

(iii) Il est souhaitable de permettre l'expérimentation rapide

Le milieu des nouvelles technologies progresse à une vitesse impressionnante. Si la France ne facilite pas l'émergence à très courte échéance (1 an à 2 ans) de nouvelles créations d'entreprises à partir de la réutilisation de données publiques, le risque est grand de voir le marché français des services *Open Data* capturé par d'autres firmes dominantes, ou les startups que d'autres économies auront su aider à émerger et se développer.

Il faut donc faciliter le plus possible l'accès aux ensembles de données intéressants par les entrepreneurs, en particulier à travers une tarification qui laisse la place à l'expérimentation.

Les entreprises à succès devront légitimement participer au financement d'une infrastructure de qualité, mais les innovateurs doivent pouvoir explorer de nouveaux modèles d'affaires dans les premières années de leur projet, ce que des redevances élevées et discriminatoires ne permettraient pas¹⁹⁷.

(iv) Une organisation dédiée pour arbitrer les conflits d'intérêts

Les deux études que nous venons d'étudier mettent en valeur des risques de conflits d'intérêt clairs entre d'une part les activités de service de type concurrentiel d'un vendeur de données à valeur ajoutée et d'autre part ses activités de service public.

On garde à l'esprit l'exemple de l'Institut Météorologique Finlandais, reconnu coupable de détériorer volontairement la qualité de ses

images radars destinées au service public, pour avantager ses opérations concurrentielles. Il a finalement été obligé de se scinder en deux sous la pression de l'autorité de la concurrence.

Les théories de l'organisation indiquent qu'en matière de coordination, la présence d'un coordinateur mandaté (et doté des leviers légaux nécessaires) est souhaitable pour faire respecter la cohérence d'une politique de tarification à laquelle les institutions n'ont qu'un intérêt commun et pas individuel.

Son rôle serait de veiller en permanence pour éviter qu'un acteur se place dans une posture du type « tragédie des communs » décrite par Hardin.

L'annonce faite par la Grande Bretagne le 12 janvier 2011 de la création d'une entité en charge des données publiques (se reporter au 1.2.) est à ce titre importante, car elle reconnaît de fait les risques forts de conflits d'intérêts, et propose une solution pragmatique, qui aura de plus la capacité à faciliter l'accès à tous types de données et le mandat d'investir dans la production de données afin de pérenniser leur qualité.

En conclusion, des solutions techniques, juridiques, organisationnelles et financières existent pour encourager l'ouverture et la réutilisation des données publiques et permettre d'en tirer toutes les conséquences positives que la société et l'économie peuvent en attendre.

Sans remettre en cause les exceptions pour lesquelles une tarification à coût marginal est nécessaire, il est crucial d'abaisser autant que possible les barrières à la réutilisation par la généralisation du modèle économique de la gratuité de réutilisation.

¹⁹⁷ Henri Verdier, 8 juillet 2009. « Quelques Remarques sur les Données Publiques (#opendata). » <http://www.henriversier.com/2010/07/quelques-remarques-sur-les-donnees.html>

Chapitre VII. Prospective : Trois scénarios pour l'avenir

Nous proposons dans ce chapitre trois scénarios schématiques, correspondant à trois directions d'évolution possibles de la réutilisation des données publiques.

Le premier scénario dessine un monde dans lequel le développement de l'offre de données publiques se développe sans accélération notable de sa dynamique historique, en retrait des évolutions technologiques rapides actuelles.

Le second scénario verrait l'offre privée accélérer et dépasser l'offre publique de données. Les monopoles informationnels et les grandes entreprises du numérique s'accapameraient alors largement l'espace occupé aujourd'hui par la puissance publique.

Le troisième scénario correspond à une action concertée de la part de la puissance publique pour catalyser l'émergence d'un écosystème dans lequel producteurs, réutilisateurs et médiateurs collaboreraient efficacement dans un cercle vertueux.

1. « Inertie » : Développement de l'offre de données publiques à son rythme historique

Le passé est un miroir tendu à l'avenir ; un premier scénario logique pour le développement futur de la mise en ligne des données publiques se construit en extrapolant son rythme historique.

Les grandes directions historiques de cette politique, jusqu'à la mise en place récente de la mission Etalab, étaient basées sur un remplissage graduel et mal coordonné des Répertoires d'Information Publique plutôt que sur impulsion centrale cohérente, et sur la recherche de revenus pour l'État et ses différentes institutions à partir des données publiques.

Ces orientations permettent de dessiner les contours et les risques d'un futur où cette tendance historique passée se poursuivrait.

1.1. Mise en ligne progressive des données au rythme historique

La politique des données publiques telle qu'elle a été menée depuis la directive européenne de 2003 et jusqu'à récemment a permis de mettre en ligne une première masse critique de données. Elle a cependant maintenu jusqu'à présent l'essentiel des sources de revenus liées à la vente de données publiques, avec pour conséquence un manque de succès en termes de réutilisations innovantes.

On peut donc imaginer dans le futur une lente montée en puissance de la diffusion de la donnée publique en ligne, et le développement plus appuyé de tentatives de génération de revenus par l'État sur les marchés de la donnée.

Les tentatives de génération de revenus rendent l'expérimentation et donc l'innovation difficiles. Il n'est donc pas certain, dans un tel scénario, que la vague de fond de l'Internet mobile et des nouveaux usages puisse suffire à provoquer l'implication grandissante des différents échelons de la puissance publique dans la mise en ligne des données.

1.2. Résolution progressive et consensuelle des blocages

Cette dynamique, si elle était maintenue sur le long terme, présenterait plusieurs avantages liés à la résolution des blocages observés.

L'initiative d'ouverture des données a historiquement été laissée à chaque administration productrice, lui permettant de concentrer ses efforts sur les enjeux qui lui sont propres. Une approche graduelle permet de ne pas ajouter à la complexité dans une situation déjà mouvante pour les administrations. Enfin, le maintien des recettes de la commercialisation de certaines données facilite le financement de cette transition par les institutions publiques.

Si cette politique se poursuivait à ce rythme historique, on peut imaginer voir d'ici 3 à 5 ans une quantité plus importante de données mises en ligne. Le renforcement de l'approche patrimoniale de la donnée publique vue comme source de revenus directs risquerait

cependant de mettre à mal les bénéfices attendus de la politique de réutilisation des données, en rendant cette réutilisation difficile pour les acteurs innovants.

1.3. Risque fort de manquer les opportunités stratégiques de l'ouverture des données publiques

Le scénario de l'inertie et du développement graduel mais lent d'une offre de données publiques au rythme historique présente des risques pour l'État et la société.

(i) Perdre du temps vis-à-vis de nos partenaires

L'avancée de nos partenaires sur le sujet risquerait de réduire le poids de la France lors des discussions sur l'éventuelle révision de la directive PSI.

Les entreprises, organes de presse et acteurs technologiques de nos partenaires internationaux bénéficieraient de plus d'une avance leur permettant d'accumuler le savoir-faire et d'identifier les marchés et besoins émergents dans le domaine des données publiques, mettant par ailleurs en danger notre secteur du numérique.

(ii) Manquer une opportunité de moderniser l'État et les outils de l'administration

L'occasion d'insuffler une dimension nouvelle à la mission de l'administration est une opportunité managériale unique pour les dirigeants de notre administration, politiques comme hauts fonctionnaires.

La manquer serait non seulement une décision contre-productive à moyen terme en ces temps de recherche de réductions des dépenses publiques. Ce serait aussi se résoudre à ne pas profiter de nouvelles opportunités de services au public, créateurs de richesse et de bien-être, dont les coûts (la production des données) sont déjà supportés par l'administration dans le cadre de sa mission de services publics.

(iii) Soumettre l'ouverture des données aux seules prédictions de rentabilité

Ne pas pousser fortement pour une ouverture rapide des données, ce serait laisser les administrations ne donner accès qu'aux jeux de données dont elles peuvent percevoir l'intérêt le plus direct. Cette approche est potentiellement dommageable d'un point de vue démocratique, en cela qu'elle peut conduire à favoriser les minorités bruyantes à

l'intérêt latent et moins fortement exprimé d'une majorité.

Surtout, la valorisation des données passe le plus souvent par le croisement statistique entre différents ensembles de données, ce qui resterait impossible à moins d'une large ouverture des données.

Se restreindre à ouvrir les jeux de données intéressants *a priori*, c'est risquer de manquer des opportunités majeures de réutilisations innovantes par la société civile et les réutilisateurs.

(iv) Manquer l'évolution des technologies contextuelles

Les nouveaux services contextuels ont besoin de données urbaines et locales pour se développer. Ne pas développer leur mise en ligne, c'est ajouter une contrainte au développement de cette industrie durant la période décisive où les futurs acteurs dominants sont encore en gestation.

Le risque est fort de désavantager les acteurs français de ce marché face aux assauts de compétiteurs mieux armés et plus expérimentés, comme ce fut le cas dans le domaine de la micro-informatique, des moteurs de recherche sur le Web, et plus récemment des réseaux sociaux. La puissance publique dispose cependant dans ce cas d'un outil stratégique puissant, qu'elle se doit de saisir.

(v) Limiter le potentiel économique lié aux données publiques

La poursuite d'un objectif de rentabilité et de création de revenus directs pour l'État, via l'établissement de redevances pour réutilisation des données, risque en outre de défavoriser la création de nouvelles entreprises, car les droits de licences commerciales rendent l'expérimentation trop risquée.

Cela signifie *in fine* un déficit potentiel d'innovation dans le domaine des produits et services, dont la taxation à la valeur ajoutée et aux bénéfices pourrait élargir l'assiette fiscale et apporter des revenus potentiellement plus importants à l'État. Ces nouvelles activités pourraient par ailleurs créer des emplois et favoriser la croissance.

(vi) Fragiliser les institutions productrices de données

La frustration de citoyens et consommateurs face à la difficulté d'accès à des données qui leur importent pourrait les

pousser à développer leurs propres solutions *ad hoc*, par exemple de manière collaborative.

Bien que louable si ces solutions évoluent vers une relation public-privé efficace, ce développement comporte une part de risques, notamment celui de la perte de légitimité des institutions qui historiquement ont produit des données importantes pour les citoyens. Le risque serait alors grand de voir s'effriter un tissu d'expertises utiles, et la qualité des services aux citoyens pourrait en souffrir.

On voit donc ici se dessiner les contours d'un scénario plus sombre, qui mettrait à mal l'intérêt public et dessaisirait la puissance publique de la production et de l'ouverture à la réutilisation des données publiques.

Poussée à son terme, cette logique conduit au scénario suivant.

2. « Capture » : Des monopoles informationnels se substituent à la puissance publique

Le travail de préparation de ce rapport a mis à jour de façon répétée la difficulté de dialogue entre l'entrepreneuriat en nouvelles technologies et l'administration, des mondes qui agissent dans des temporalités différentes.

À l'inverse du scénario précédent, dans lequel on imaginait voir ces tensions se résoudre graduellement, on peut aussi esquisser en négatif un futur alternatif, où les carences de l'État deviennent un véritable attentisme qui laisse le champ libre à la capture¹⁹⁸ de l'espace citoyen des données publiques par le monopole et l'intérêt privé.

2.1. L'offre privée dépasse l'offre publique

Dans un scénario du cercle vicieux, la demande privée dépasse les moyens de soutien à l'offre de la puissance publique. Devant des conditions de réutilisation jugées comme discriminatoire, le public expert se tournerait vers les solutions alternatives, soit collectives (de type Open Street Map), soit privées.

La possibilité d'accéder à des applications basées sur les données à caractère public mais collectées en nom privé créerait auprès du public la perception d'un retard de plus en plus flagrant de l'administration. La pression pour

l'accès à ces nouveaux services augmentant, elle risquerait de soutenir le développement de monopoles et d'accords exclusifs anti-compétitifs.

(i) Les contrats d'exclusivité renforcent les monopoles informationnels privés

Face à cette forte demande du public, et dans un contexte où l'expertise accumulée serait majoritairement captée par les monopoles informationnels et les grandes entreprises dominantes du Web, il deviendrait difficile pour les entités publiques d'éviter les contrats privilégiés et les accords exclusifs pour subvenir à la mise en ligne de données encore non accessibles.

Ces contrats renforceraient la tendance, endémique aux marchés de l'information, à l'apparition de monopoles¹⁹⁹. Bien que la position dominante de ce type d'acteurs ne soit pas nécessairement dommageable au grand public dans le cas où leurs services sont ouverts à tous mais rémunérés par la publicité, ils n'en demeurent pas moins un problème pour le développement éventuel d'une industrie domestique.

(ii) Les contentieux se multiplient

Surtout, une telle situation verrait la multiplication des contentieux, soit administratifs dans le cas où des concurrents s'estimeraient lésés, soit civils dans le cas où une entreprise contractuelle considérerait que ses droits sont lésés par un autre réutilisateur.

Le climat d'incertitude juridique pourrait avoir un effet négatif sur l'innovation et la réutilisation du fait des trop grands risques perçus que des investissements dans des programmes de réutilisation présenteraient.

(iii) L'initiative privée contourne les institutions publiques

Un avantage de l'apparition des contentieux est bien sûr qu'ils permettraient *in fine* de créer la jurisprudence, et de clarifier les droits et devoirs des acteurs du marché.

De plus, face au manque perçu de réactivité de la puissance publique dans le domaine de la mise en ligne des données, et les contraintes créées d'autre part par les contrats d'exclusivités qui viendraient à se multiplier, le grand public pourrait s'organiser pour collecter ses propres données, sur le modèle du

¹⁹⁸ Kaplan, D., 2010, « Les données publiques : et après ? », Internet Actu, 9 novembre 2010 <http://www.Internetactu.net/2010/11/09/louverture-des-donnees-publiques-et-apres/>

¹⁹⁹ Wu, T., 2010, « The Master Switch : Rise and Fall of Information Empires ».

« crowdsourcing », ce qui peut être un modèle efficace de production auto-organisée.

En parallèle, des entreprises privées se positionneraient pour offrir, qui de nouveaux modes de collecte de données (par exemple à travers les données résiduelles de géolocalisation collectées par les opérateurs²⁰⁰ ou par des services applicatifs ad hoc²⁰¹), qui de nouveaux services. Ce scénario présenterait donc des avantages certains.

2.2. Valorisation des données au détriment de l'intérêt public à long terme

Cependant, le climat conflictuel dans lequel ce futur hypothétique se déroulerait tendrait à exacerber les problèmes potentiels que pouvait poser le scénario 'Inertie' et poserait un défi majeur à la puissance publique.

(i) Manquer une opportunité de moderniser l'administration

Dans ce scénario, la puissance publique laisserait échapper une occasion de moderniser le travail de l'administration. Le risque n'est donc plus seulement de passer à côté d'une occasion d'insuffler une dimension nouvelle à la mission du service public, mais bien de renforcer les rigidités dont il peut parfois faire preuve, et de voir se dégrader durablement ses relations avec ses usagers et l'écosystème de l'innovation digitale.

(ii) Fragiliser les institutions productrices de données

Le climat de défiance dans lequel un développement rapide de ce type se déroulerait pourrait ainsi avoir des conséquences fortement négatives sur le lien numérique entre l'administration et les citoyens.

Si cette politique se poursuivait, on peut imaginer l'apparition dans les 3 à 5 ans d'une véritable fracture numérique, cette fois entre l'État et les administrés. Cela mènerait à la fragilisation des institutions publiques liées à la collecte de données, ce qui pourrait déboucher sur un problème de légitimité pour l'action de l'État, dans un contexte²⁰² (appels à la transparence, montée en force du débat

public) où celle-ci est parfois déjà remise en question.

(iii) Surcharger dangereusement les capacités de la CADA

Le risque est fort qu'une augmentation sensible des contentieux vienne grever les services de la Commission d'une charge de travail supérieure à ses ressources actuelles. Outre un retard mécanique des développements autour de la réutilisation des données, ceci pourrait mettre en danger le droit des citoyens à accéder aux documents administratifs.

(iv) Soumettre l'ouverture des données aux exigences de rentabilité des grandes entreprises de l'information

Un défi spécifique que poserait ce scénario est la montée en puissance des seules entreprises privées dans le champ de la diffusion des données publiques.

Dans un monde numérique, la communication et la réutilisation des données ont une portée politique qui participe au « vivre ensemble » entre citoyens. Si l'État ne remplissait pas, ou pas assez rapidement aux yeux du public, le rôle de producteur et distributeur de données qui permettraient de développer les services attendus par les citoyens, la sphère privée remplirait seule ce rôle de décisionnaire quant aux données qui ont vocation à être collectées et mises à disposition.

Cela impliquerait un recul des options stratégiques disponibles à la puissance publique d'une part, et d'autre part pourrait créer des difficultés pour les citoyens ou pour d'autres entreprises, notamment les startups et entreprises de croissance, qui risqueraient de ne pas avoir accès aux données qui leur sont nécessaires dans des conditions aussi faciles que les entreprises établies productrices de ces données.

Surtout, cela risquerait de revenir à concéder les marchés porteurs, et les emplois qualifiés et souvent difficiles à délocaliser que ces nouvelles activités sont à même de créer, aux acteurs issus de marchés où l'accès plus facile aux données publiques leur aurait donné un avantage concurrentiel initial.

(v) La prolifération des licences pourrait bloquer le développement de la réutilisation

Le statut juridique de certains ensembles de données est souvent compliqué par l'implication d'acteurs de droit privé en parallèle à l'administration (par exemple dans

²⁰⁰ Urban Mobs, 2009, projet de visualisation géographique du trafic cellulaire mené conjointement par Orange et FaberNovel <http://www.urbanmobs.fr/fr/france/>

²⁰¹ Waze, Inc. http://world.waze.com/guided_tour/

²⁰² Landier, A. et Thesmar, D., 2010, « La Société Translucide : Pour en Finir avec le Mythe de l'État Bienveillant »

le cas où un contrat de service public ne délimite pas clairement les droits liés aux données générées par un opérateur privé de réseaux publics).

Si ce type de situation se multipliait, il serait à craindre que cela ne mette en danger la capacité des entreprises à innover dans le domaine des données publiques.

De même que l'on observe dans le domaine des biotechnologies que la prolifération de brevets croisés sur certaines molécules et ses applications empêche parfois l'innovation de prendre place par peur pour les nouveaux entrants de ne pouvoir valoriser les résultats sans passer par un contentieux coûteux²⁰³, on risquerait de voir des applications utiles au public perçues comme des investissements trop risqués du simple fait du manque de clarté quant aux conditions sous lesquelles les données sous-jacentes pourraient être réutilisées.

(vi) Focaliser sur les échecs et frustrations.

Un scénario dans lequel la mise en ligne des données publiques et leur valorisation se feraient dans la douleur mettrait de plus l'accent de façon exagérée sur les ratés inévitables, les contentieux qui émergeraient sans doute et les frustrations des différents acteurs face à telle ou telle décision politique ou administrative.

Cela risquerait de mettre à mal la légitimité de cette politique tout entière, et de repousser les bonnes volontés en cristallisant les quelques problèmes qui viendront naturellement à se poser.

Même si ce scénario avait l'avantage d'accélérer dans un premier temps l'apparition de certains services utiles aux citoyens, il serait probablement néfaste à moyen et long terme.

3. « Symbiose » : Catalyser le développement d'un écosystème

Au contraire des deux scénarios précédents, on peut s'inspirer des expériences les plus réussies jusqu'à présent, à l'étranger comme sur le territoire national, pour tracer les lignes d'un scénario dans lequel l'État accompagnerait avec succès les collectivités

locales, les institutions publiques et les acteurs privés ou citoyens dans la création d'un écosystème en développement harmonieux.

3.1. Un écosystème soutenu par une volonté politique forte

Dans un tel scénario, l'administration impliquerait et fédérerait les volontés existantes autour de la question de la réutilisation des données publiques comme projet de modernisation de la société.

Elle déploierait une volonté politique forte pour harmoniser et mettre en cohérence les politiques menées par les différentes institutions productrices de données.

Cela supposerait une communication accrue entre administration et réutilisateurs, et l'établissement de ponts et de points de contacts entre les différents types d'expertises.

En contrepartie de cet investissement modeste, les bénéfices attendus dans ce futur hypothétique sont une plus grande fluidité des échanges autour de la question des données, la circulation des meilleures pratiques, et une multiplication des applications utiles aux citoyens développées à partir de ces données.

Tous ces éléments pourraient concourir à accélérer le déploiement d'une politique nationale forte de réutilisation des données publiques, et la réalisation concrète de son potentiel.

L'exemple britannique est à ce titre édifiant. Lors de la conférence de presse à laquelle nous avons pu assister le 19 novembre 2010 à Londres, à l'occasion de la mise en ligne de nouvelles données de dépenses par l'État britannique (l'ensemble des données d'exécution des budgets nationaux détaillés dépense par dépense à partir de 25 000 livres, avec mise à jour mensuelle), les avantages d'une telle approche étaient manifestes : la richesse des réutilisations présentées, l'anticipation des besoins des utilisateurs, la coproduction mise en œuvre à cette occasion.

Nous pouvons en tirer trois leçons sur les éléments constitutifs d'un scénario de symbiose. Il nécessiterait une coopération technique rapprochée entre l'État, le privé et l'associatif. La production de services par la société civile à partir des données publiques deviendrait la norme. Elle pourrait jouer un rôle d'exemple, et inciterait à mutualiser des solutions techniques, accélérant plus encore la mise en ligne des données.

De plus, une symbiose réussie repose sur la coopération avec des journalistes formés aux

²⁰³ Kenneth A. Oye and Rachel Wellhausen, "The Intellectual Commons and Property in Synthetic Biology," in Markus Schmidt et al (eds), *Synthetic Biology: The Technoscience and its Societal Consequences*, Springer, 2009, pp 121-140.

techniques numériques. Le maniement des données, leur analyse et la présentation des résultats sont des compétences bien spécifiques et qui évoluent rapidement avec les nouvelles technologies. Cette coopération serait nécessaire pour réaliser pleinement l'objectif de transparence de l'action de l'État concernant le mouvement *Open Data*.

Enfin, la symbiose entre acteurs de l'écosystème permettrait d'intégrer avec confiance la mise en ligne des données publiques dans l'effort de pédagogie des administrations sur leur travail et dans leur stratégie de renforcement du lien numérique aux usagers.

3.2. Nouveau rôle et nouveaux risques pour l'Administration

La mise en ligne des données afin de les voir réutilisées dans le cadre du développement de services innovants est une stratégie qui reste encore très novatrice. Le travail de fédération des initiatives que l'État mènerait dans ce scénario doit être l'occasion de prendre en compte les risques inhérents à une telle approche.

Il est important notamment d'éviter l'écueil qui consisterait à abandonner entièrement la mission de réutilisation des données publiques à la société civile. La production d'information synthétique ou de services à destination des citoyens est une activité légitime de l'administration.

De plus, l'accélération de la mise en ligne des données pose des problèmes spécifiques par rapport aux autres scénarios. En particulier, une certaine anarchie dans les données disponibles à la mise en ligne risque d'apparaître initialement, ce qui suppose un effort particulier d'indexation et la mise en ligne d'un portail identifié d'accès aux données publiques.

3.3. Mise à disposition rapide et à faible coût de la majorité des données publiques

En regard de l'effort de soutien politique nécessaire au développement d'un tel scénario, les avantages qu'il pourrait présenter sont nombreux et particulièrement attractifs.

(i) Demande peu de moyens et peut être mis en œuvre rapidement

La clé d'un tel scénario passe avant tout par la clarification des attentes des réutilisateurs auprès des administrations, et la mise en communication de ces différents acteurs.

Cela suppose certes un investissement dans l'analyse et la compréhension des enjeux, mais l'expertise qui s'est portée sur le sujet est déjà forte, notamment dans le monde associatif (Regards Citoyens, LiberTIC, FING), institutionnel (Délégation aux Usages de l'Internet, APIE, Ministère de la Justice, collectivités locales, Commission Européenne) et privé (Cap Digital). La création récente de la mission Etalab pourrait permettre la fédération de ces énergies et expertises.

Les moyens nécessaires à mettre en œuvre pourront être mobilisés par la réorientation de l'effort qui portaient jusqu'ici sur la tarification de la diffusion.

(ii) Catalyse les énergies et les bonnes volontés

L'implication proactive de l'État dans un réseau d'acteurs d'origines différentes valorise fortement l'activité de chacun. Par opposition aux scénarios précédents qui reposaient sur une tension forte entre demande de réutilisation et offre de données, les frictions éventuelles seraient ici fortement tempérées par la reconnaissance des bonnes volontés de part et d'autre de l'écosystème.

(iii) Favorise l'expérimentation

Dans ce scénario, l'expérimentation serait facilitée par la mise en cohérence des conditions tarifaires, juridiques et techniques conditionnant la réutilisation d'une grande partie des données publiques. Il est raisonnable d'imaginer un cercle vertueux par lequel les innovations à succès attirent l'attention d'autres scientifiques, ingénieurs ou entrepreneurs, et encouragent à leur tour encore plus d'innovation.

(iv) Facilite la modernisation de l'administration et promet des économies de fonctionnement

Contrairement aux scénarios précédents dont l'un des risques majeurs était de ne pas profiter de l'opportunité de modernisation de l'État que représente la mise en ligne des données publiques, une évolution en symbiose permettrait d'accompagner plus efficacement l'adaptation de l'administration à un monde où la mise en ligne des données de fonctionnement des institutions serait la norme.

En utilisant des outils standards de mise en ligne comme des plateformes mutualisées basées sur des logiciels libres, la puissance publique effectuerait des économies substantielles. Elle pourrait par ailleurs favoriser l'émergence de nouveaux outils et

standards d'interopérabilité, repris par l'ensemble de la communauté.

(v) Facilite la circulation de l'information technique et juridique.

Un aspect malheureux mais inévitable de la réutilisation des données publiques est la complexité des sujets que le thème recoupe. La nouveauté même du concept et son émergence récente, du fait de développements technologiques qui commencent seulement à s'esquisser aux yeux du grand public, entraînent une certaine difficulté à communiquer les enjeux en question.

En particulier, la question de la réutilisation des données suppose des connaissances étendues et fréquemment réactualisées en informatique, en finances publiques, en droit des documents administratifs, et en droit de la propriété intellectuelle. Il est probablement illusoire de compter sur l'apprentissage exhaustif de ces pratiques par l'ensemble des acteurs publics concernés par ces questions.

Le développement d'un réseau informel d'acteurs en relation fréquente permettrait à l'inverse aux compétences de chacun de se renforcer mutuellement.

(vi) Permet de communiquer l'importance du sujet aux administrations et collectivités

L'implication de l'État dans un processus de collaboration avec la communauté a comme effet direct de souligner l'importance du sujet aux administrés ainsi qu'aux agents de l'État et des collectivités. C'est donc à la fois un outil politique et un outil de gestion puissant.

(vii) Un nouvel outil de communication pour l'État

Enfin, l'émergence d'un tel écosystème autour de la mise en ligne des données publiques pourrait permettre à des journalistes experts en maniement, interprétation et visualisation des données de travailler directement sur les données issues des administrations et des collectivités.

Cela ouvrirait de nouvelles possibilités pour éclairer le débat public, et de nouvelles techniques journalistiques utiles à la fois pour expliquer les politiques qui sont menées et en débattre, pour mettre en valeur le travail de l'administration et pour souligner les enjeux des arbitrages réalisés au niveau national ou territorial.

Les caractéristiques propres à chaque scénario sont synthétisées dans le tableau ci-contre.

Sur la base de l'état des lieux et des trois scénarios étudiés dans ce chapitre, nous proposons au chapitre suivant une série de 16 propositions pour une politique ambitieuse des données publiques, qui pourraient permettre de concrétiser les bénéfices de la réutilisation des données publiques.

Critères		Scénarios		
		Inertie	Capture	Symbiose
Répondre à la demande	Rapidité	-	+	+
	Saisir l'opportunité des nouvelles technologies	--	-	++
	Valorisation des données	--	-	++
Mise en œuvre	Moyens nécessaires	++	++	++
	Facilité	++	+	-
Impact sur l'administration	Renforcer le rôle l'administration	-	--	+
	Modernisation de l'administration	-	--	++
	Valorisation de l'action de l'administration	-	--	+
Intérêt public	Court terme	--	+	++
	Long terme	-	--	++

Chapitre VIII. Seize propositions pour une politique ambitieuse des données publiques

À partir des leçons des premières expériences d'*Open Data* et des scénarios que nous avons détaillés au chapitre précédent, nous proposons trois axes stratégiques pour une politique ambitieuse qui viserait à réaliser le plein potentiel de la réutilisation des données publiques.

Simplifier pour accélérer : réduire les efforts nécessaires d'une part à la diffusion des données par les acteurs publics, et d'autre part à leur réutilisation, afin d'accélérer le développement de services utiles à partir des données publiques.

Faciliter l'expérimentation : adapter les conditions de réutilisation des données publiques pour faciliter l'expérimentation par les citoyens, les associations et la communauté des développeurs et des entrepreneurs.

Favoriser l'émergence d'un écosystème : favoriser l'émergence d'un écosystème de producteur et de réutilisateurs de données publiques en France, autour de la mission Etalab et du futur portail data.gouv.fr.

Ce chapitre suggère des modalités concrètes de développement de ces trois axes stratégiques. Il présente seize propositions issues de notre analyse, qui visent à faciliter la réutilisation la plus large des données publiques et à permettre la réalisation la plus rapide possible de l'ensemble des bénéfices que la société peut en attendre.

1. Simplifier pour accélérer

1.1. Faciliter la diffusion de données publiques

Proposition 1 : Formaliser les bonnes pratiques de diffusion de données publiques attendues par la communauté des réutilisateurs.

Expliciter l'effort minimal attendu des différents acteurs publics. Pour répondre aux préoccupations très fortes des institutions publiques amenées à diffuser leurs données en ligne pour la première fois, il serait souhaitable de rappeler que l'effort demandé est minimal.

Dans la grande majorité des cas, la mise en ligne des données n'a pas de raison de mobiliser excessivement les agents du service, qui plus est en regard des économies de temps qui découleront de la possibilité pour les citoyens d'accéder aux données directement sans devoir établir une demande officielle auprès de l'institution.

Assurer la formation des agents et leur accès à des guides clairs et explicatifs. Dans la continuité du travail de préfiguration réalisé par l'APIE, nous suggérons de poursuivre l'effort de mise à jour de guides et documents à destination des agents. Notamment, détailler les raisons qui poussent à vouloir mettre en ligne les données publiques, l'importance des formats réutilisables, les étapes à suivre et les grands enjeux juridiques, administratifs et techniques serait d'une aide importante aux institutions publiques amenées à diffuser des données.

Simplifier les formats de diffusion. Nous suggérons d'explicitement clairement la simplicité potentielle du processus de mise en ligne de données. De simples fichiers CSV suffisent à participer activement à la mise en ligne des données publiques, qui n'a donc pas lieu de mobiliser outre mesure les ressources techniques et humaines de l'administration.

1.2. Réduire les barrières à la réutilisation

Proposition 2 : Mettre en cohérence les conditions légales de réutilisation des données issues de sources différentes.

Établir une « licence données ouvertes » pour réutilisation gratuite qui s'appliquerait le plus largement possible. Nous suggérons d'harmoniser le plus possible les licences utilisées au sein des différentes administrations de l'État, des établissements publics et des collectivités territoriales. Il est nécessaire d'éviter la prolifération de licences, qui rendraient la réutilisation de données issues de sources différentes trop complexe.

Nous suggérons pour cela un travail d'harmonisation vers une même licence de

réutilisation à titre gratuit, la « Licence Données Ouverte » de l'État, qui serait utilisée par défaut par les administrations centrales, déconcentrées, et dont les collectivités territoriales pourraient s'inspirer.

Appliquer cette licence de la manière la plus large aux données disponibles sur data.gouv.fr permettrait de simplifier substantiellement le régime juridique qui s'applique à ces données.

Ce travail pourrait être mené en concertation entre les différents acteurs et sur la base du travail effectué par le Ministère de la Justice (licence « Information Publique »), des conditions générales d'utilisations rappelées par l'APIE, et des licences issues du monde du libre (par exemple licence ODbL).

Encourager un marquage graphique des droits de réutilisation. Afin de permettre aux réutilisateurs d'identifier facilement les données réutilisables et les conditions qui s'y appliquent, nous suggérons de généraliser le marquage des fiches descriptives de jeux de données disponibles à la réutilisation avec des pictogrammes spécifiques, sur le modèle du marquage « Information Publique » du ministère de la Justice, qui expliqueraient les droits afférents à ces données.

Proposition 3 : Encourager l'utilisation de formats facilement réexploitables, qui respectent les normes d'interopérabilité et de « lisibilité machine ».

Privilégier les formats ouverts respectant les normes d'interopérabilité. L'utilisation d'un format fermé et difficilement réexploitable peut limiter très fortement l'utilité d'un jeu de données pour ses réutilisateurs, en rendant sa réutilisation trop complexe.

Communiquer sur l'importance des formats réexploitables auprès des agents. Il est crucial que les agents de l'État en charge de l'accès aux données publiques soient bien sensibilisés à cette question, et disposent de bases de connaissances et d'outils simples pour privilégier l'utilisation de formats ouverts et respectant les normes d'interopérabilité.

Maintenir une liste de formats correspondants à ces critères. À cet effet, il sera utile de définir une liste des formats facilement lisibles par des machines, à destination des agents.

Les institutions publiques privilégieront idéalement les formats simples et non propriétaires de type CSV (simples valeurs séparées par des virgules, outil standard parmi les tableurs), et éviteront de communiquer des jeux de données au format PDF (qui rend les données difficilement manipulables) sauf pour des rapports de type traitement de texte.

Au-delà des recommandations du Référentiel Général d'Interopérabilité²⁰⁴ (RGI) aux administrations, des évolutions seront à prévoir dans la perspective de la réutilisation des données publiques. Il est à noter par ailleurs que la Commission européenne est en discussions avec le W3C dans le cadre d'une réflexion liée aux questions de standards.²⁰⁵

Assurer la pérennité des ontologies et structures de données dans le temps. Une des difficultés liées à la réutilisation de jeux de données remis à jour fréquemment tient aux changements potentiels d'ontologie ou de structuration des données qui peuvent intervenir dans le temps. Dans le cas des jeux de données mis à jour périodiquement, les administrations productrices devront veiller à reproduire le format et la structure de données utilisée, afin d'en faciliter une réutilisation automatisée.

1.3. Anticiper les nouveaux usages des données publiques

Proposition 4 : Faire en sorte que chaque contrat public clarifie à l'avenir le financement de l'ouverture et les droits de réutilisation liés aux données produites dans son cadre.

Clarifier le financement de l'ouverture et les droits de réutilisation de données dans les contrats publics. Une quantité importante de données d'intérêt public sont générées par des acteurs qui sortent du champ de la loi CADA de 1978. Plutôt qu'une modification de la réglementation qui viserait à intégrer l'ensemble des cas particuliers, nous suggérons de privilégier dans un premier temps un travail d'incitation et de communication auprès des entités concernées sur l'importance de ce sujet.

²⁰⁴ Référentiel Général d'Interopérabilité : http://references.modernisation.gouv.fr/sites/default/files/RGI_Version1%200.pdf

²⁰⁵ Commission Européenne, 16 décembre 2010, "W3C annonce les bonnes pratiques pour un web mobile", CORDIS actualités. http://cordis.europa.eu/fetch?CALLER=FR_NEWS_FP7&ACTION=D&DOC=2&CAT=NEWS&QUERY=012cf488f5e8:c26c:0ac2e8d6&RCN=32879

Nous préconisons notamment que l'ensemble des Contrats d'Objectifs et de Moyens des Établissements publics et des Délégations de Service Public contractées avec des entreprises abordent explicitement le statut et les droits de réutilisation afférant aux données générées dans leur cadre.

Ces contrats devraient ainsi prévoir la réorientation des moyens déployés pour facturer et diffuser des données protégées vers l'organisation de l'infrastructure nécessaire à l'ouverture et à la diffusion gratuite ou à coût marginal de ces données.

Il serait bien évidemment souhaitable que ceci se fasse dans le sens d'une généralisation la plus large de la réutilisation des données publiques, de façon gratuite, à titre personnel ou commercial.

Établir des bonnes pratiques et des clauses types. L'État pourrait établir des bonnes pratiques, notamment sur les types de données concernées, les formats à privilégier, et les modèles d'accès préférables, en vue du renouvellement des COM et DSP concernés, secteur par secteur (par exemple transport, eau, environnement). Il pourrait par ailleurs préparer des clauses types que les différentes institutions pourraient reprendre et adapter à leur situation spécifique.

Élargir le périmètre des données réutilisables aux ressources essentielles. Nous suggérons enfin de réfléchir à l'opportunité d'octroyer, sur la base de la doctrine des ressources essentielles en droit de la concurrence, un droit de réutilisation à titre personnel ou commercial sur certaines données générées par des acteurs privés mais qui pourraient avoir une importance forte pour le public, et pour l'établissement d'un marché concurrentiel aval, notamment d'applications et services.

Proposition 5 : Faciliter l'enrichissement des données et l'intégration de corrections éventuelles par la communauté.

Faciliter l'intégration des métadonnées. Une des craintes exprimées par les administrations est le risque que les métadonnées explicatives nécessaires à une utilisation judicieuse des données (par exemple l'incertitude de mesure) se perdent lorsque les données sont recoupées et ne soient pas intégrées aux réutilisations.

Nous suggérons, en réponse à cette crainte, de développer une solution technique de marquage sémantique permettant l'interprétation par les moteurs de recherche ou les outils de collecte de données des métadonnées accompagnant le jeu de données, et leur enrichissement en fonction du contexte. Une norme « microformat²⁰⁶ » pourrait constituer une solution technique à cet effet.

Permettre l'enrichissement collaboratif des jeux de données. Un des avantages de la mise en ligne des jeux de données de l'État consiste en la possibilité, pour les réutilisateurs qui remarqueraient des erreurs dans les jeux de données de les signaler voire de les corriger.

Par exemple, les données sur le système national de transports publics en Grande-Bretagne, mises en ligne en 2010, contenaient près de 6 % de localisations d'arrêts de bus erronées, que les utilisateurs ont pu détecter et signaler²⁰⁷. Le défi pour la puissance publique consiste à mettre en place les outils techniques et réglementaires lui permettant de réintégrer éventuellement de telles corrections, afin d'en bénéficier elle-même.

L'État devra donc réfléchir à la possibilité, à terme, de passer d'un modèle « à sens unique » (diffusion des données du secteur public vers la société civile) à un modèle d'écosystème où les données de l'État et des collectivités, ouvertes à la société civiles, pourraient être enrichies en retour de façon collaborative (« crowdsourcing »).

Proposition 6 : Impliquer des acteurs publics pilotes dans les projets de recherche sur le Web sémantique.

Bien que prématurées pour un déploiement initial à grande échelle, les technologies du Web sémantique et des données liées (« *linked data* ») pourraient devenir un élément majeur de l'Internet de demain. Participer à ses premières évolutions pourrait donc être d'une importance stratégique pour l'administration française.

²⁰⁶ Chaumont, J., 2010, « Qu'est-ce que le Web Sémantique ? », section 3.1 et suivantes dans *Social Commerce : Quand le E-Commerce Rencontre le Web d'Aujourd'hui*.

²⁰⁷ Shadbolt, N., 2011, allocution au colloque « Open Data : Et Nous, Et Nous, Et Nous ? » du 17 mars 2011 organisé par Regards sur le Numérique. Présentation : <http://www.slideshare.net/RSLN/opendata-datagovuk-how-did-we-do-it>

Nous suggérons d'identifier un certain nombre d'acteurs publics en pointe du mouvement de réutilisation des données, et de soutenir l'utilisation de leurs jeux de données dans le cadre des projets de recherche internationaux sur le sujet²⁰⁸.

2. Faciliter l'expérimentation pour encourager l'innovation

2.1. Généraliser la réutilisation gratuite

Proposition 7 : Rendre la réutilisation des données gratuite pour toute réutilisation à titre personnel.

D'après les études économiques déjà présentées (étude de Strasbourg), la propension des citoyens à payer un droit d'accès aux données publiques pour en faire ensuite un usage personnel est généralement très faible.

Par défaut, et en dehors des exceptions de bon sens (par exemple des ensembles de données de très grande taille occasionnant des coûts de distribution conséquents), une tarification pour usage personnel réduirait donc fortement la diffusion des données, et donc les bénéfices qu'on peut en attendre.

Ceci irait qui plus est à contre-courant des fortes attentes des citoyens en terme de transparence de l'action de l'État. Nous suggérons donc d'établir un principe formel de gratuité pour toute réutilisation de données publiques à titre personnel.

Proposition 8 : Encadrer strictement les cas dans lesquels une redevance est légitime, et charger les producteurs de données d'établir l'absolue nécessité de toute tarification.

Dépasser une valorisation purement commerciale des données publiques et encourager leur réutilisation la plus large. Dans quelques cas spécifiques et bien identifiés, le revenus issus des données publiques participent aux principes d'indépendance de notre démocratie. Toutefois, les institutions produisant des données publiques au sens de la loi de 1978

doivent dépasser un mode de valorisation des données qui serait purement commercial.

Dans la vaste majorité des cas, l'établissement d'une redevance n'apporterait aucun revenu significatif à l'échelle des institutions productrices, et ne justifierait de toute façon pas les coûts liés à la mise en œuvre d'une tarification.

Il est donc souhaitable que les acteurs publics se concentrent sur la réalisation de leurs missions de service public, et mettent les données créées dans ce cadre à la disposition des entrepreneurs et des startups plutôt que de viser à les valoriser eux-mêmes à travers une redevance inefficace.

Généraliser la réutilisation libre et gratuite la plus large possible. Dans le cas de jeux de données nouvellement mis en ligne, débouchant sur des marchés en développement et dont les applications sont encore méconnues, le potentiel d'innovation lié à la réutilisation est particulièrement fort.

Il apparaît donc nécessaire d'encourager la gratuité à l'ensemble des jeux de données mis en ligne, même pour des réutilisations à titre commercial, à moins que les producteurs de données puissent établir, sur la base d'arguments fondés (notamment liés à des besoins d'investissements lourds de diffusion des données), l'absolue nécessité d'une redevance.

La décision récente du Premier ministre François Fillon soumettre à un décret toute possibilité d'établir une tarification des données publiques va d'ailleurs dans ce sens.

Restreindre la tarification à coûts moyens aux marchés matures. Une tarification qui recouvrerait les coûts moyens de mise à disposition n'est justifiable que pour les marchés matures des données à valeur ajoutée, où le potentiel d'innovation est plus limité et les modèles d'affaires mieux connus, et seulement si cela ne réduit pas excessivement la diffusion des données.

Rappelons cependant que la théorie économique établit qu'il n'est légitime de facturer aux réutilisateurs que les coûts engendrés *du fait et uniquement du fait* de la mise à disposition des données, et non pas des frais inhérents au fonctionnement normal des institutions, tels que des frais de production des données qui rentrent dans le cadre de leur mission de service public.

²⁰⁸ « LOD2 : Creating Knowledge out of Interlinked Data », projet de recherche supporté par la Commission Européenne : <http://lod2.eu>

2.2. Faciliter l'expérimentation pour les citoyens et les acteurs de l'innovation

Proposition 9 : Conditionner toute redevance pour réutilisation de données à valeur ajoutée à la mise à disposition gratuite des données brutes qui les sous-tendent.

S'il existe un marché pour des données à valeur ajoutée, son optimalité suppose qu'il soit le plus possible ouvert à la concurrence.

Pour que l'établissement d'une redevance pour réutilisation de données à valeur ajoutée produite par une institution publique ne soit pas anticoncurrentielle, il est nécessaire que les données brutes sur lesquelles cette activité est construite, éventuellement recueillies dans le cadre d'une mission de service public, soient disponibles et réutilisables commercialement.

Les institutions qui fixent une redevance pour la réutilisation de données à valeur ajoutée devraient donc mettre gratuitement les données brutes sous-jacentes à disposition des réutilisateurs, au risque sinon de créer des distorsions de concurrence dommageables et sources de contentieux potentiels.

Proposition 10 : Explorer des conditions alternatives de réutilisation pour laisser carte blanche à l'expérimentation.

. Même lorsque des redevances pour réutilisations sont jugées légitimes, il nous apparaît très important d'encourager les administrations à se montrer favorables à des expérimentations. Notamment, elles devraient viser à permettre aux acteurs innovants, comme les Jeunes Entreprises des Pôles, d'expérimenter de nouvelles activités commerciales sans devoir s'engager sur un contrat de long terme qui créerait des barrières à l'entrée discriminantes. Nous suggérons dans ce cas de rechercher des modes de couverture des coûts alternatifs qui permettent de favoriser l'innovation.

On privilégiera donc la recherche de partenariats avec des pôles de compétitivité et des incubateurs pour favoriser l'émergence d'applications innovantes. Ces partenariats devront prévoir un partage par la puissance publique du savoir-faire et de la connaissance pointues de ses données avec des startups rompues aux techniques de programmation les plus pointues.

Explorer les licences gratuites avec partage des conditions initiales à l'identique (« share alike »). Lorsqu'une redevance couvrant la réutilisation de données à titre commercial ne saurait être remise en question, l'institution productrice pourrait explorer la mise en place d'une licence gratuite pour réutilisation commerciale restreinte avec partage des conditions initiales à l'identique (ce qui exclut notamment la revente directe des données).

Ce modèle de licence dite « *share alike* » a par exemple été mise en place par la Mairie de Paris, et permet de soutenir la création de nouveaux usages, par exemple des services aux utilisateurs financés par la publicité, tout en excluant du champ de la gratuité les activités commerciales de revente directe (et donc en assurant le maintien des sources de revenus liés à la donnée publique).

Explorer la rémunération par l'accès aux données d'usage. Une piste à explorer est la vente de données d'usages sur certains jeux de données ou applications spécifiques, qui pourraient avoir une valeur prédictive forte ou intéresser les acteurs d'un marché particulier. Cela permettrait d'alléger en adéquation la tarification de la réutilisation elle-même.

Explorer la rémunération par le service client. Dans le cas des données à valeur ajoutée commerciale, nous suggérons d'imaginer une tarification double, liée au niveau de service promis aux réutilisateurs. Une option gratuite mais offrant des garanties moindres faciliterait l'expérimentation.

Faciliter l'expérimentation avec les Web Services. Au-delà de la tarification pour réutilisation des données elles-mêmes, se pose la question de l'accès à des services informatiques (notamment des API ou des « Web Services ») permettant un accès facilité ou en temps réel à des données ou des services basés sur les données.

Lorsque des contraintes techniques mènent le producteur de ce Web Service à fixer une rémunération pour service rendu, nous suggérons de laisser la possibilité à des acteurs innovants d'expérimenter gratuitement avec le service en dessous d'un seuil de volume suffisamment élevé pour permettre une mise en production à l'échelle. Cela permet d'encourager l'innovation tout en respectant les contraintes techniques du producteur.

3. Soutenir l'émergence d'un écosystème

3.1. Animer et la communauté des réutilisateurs

Proposition 11 : Impliquer la communauté des réutilisateurs dans les initiatives de mise à disposition de données publiques.

Créer un point d'accès centralisé et mis à jour automatiquement. Nous suggérons, pour permettre une réutilisation efficace des données mises en ligne, de construire le portail data.gouv.fr sur le modèle d'un moteur de recherche, indexant les ensembles de données publiques réutilisables et leurs métadonnées en un point d'accès identifié.

Il serait souhaitable à terme d'automatiser la mise à jour de l'indexation des données autant que cela sera possible techniquement.

Développer une offre grand public qui mette en valeur des applications utiles. Un écosystème de réutilisation des données publiques se doit d'impliquer le grand public. Il serait donc souhaitable, à travers le portail data.gouv.fr, de mettre en avant des réutilisations pratiques et utiles aux citoyens réalisés à partir des données publiques.

Le danger d'un portail axé uniquement sur l'agrégation de liens vers des jeux de données brutes est qu'il soit ignoré du grand public, qui n'y trouverait pas son intérêt.

Le portail data.gouv.fr devra être un lieu focal avec une double vocation : portail d'accès aux données pour les réutilisateurs, et site de destination pour consommateurs recherchant des applications utiles à partir des données publiques.

Nous suggérons donc de créer une section « Applications » et de mettre ces dernières en valeur sur la page d'accueil elle-même.

Investir dans l'accompagnement de la communauté. Afin de faciliter l'échange de bonnes pratiques et la formation d'une véritable communauté, il est souhaitable d'investir dans la création d'un forum ou d'une base de connaissances évolutive, solution aujourd'hui classique sur le Web. Afin de préserver la qualité des échanges qui y auront lieu, et pour impliquer la communauté dans les activités autour de la donnée publique, nous suggérons de plus de créer un ou plusieurs postes d'animateurs (« *community managers* »).

Démultiplier le service aux utilisateurs. Le futur portail pourrait aussi lister et mettre à jour les coordonnées des personnes à même de répondre aux questions des réutilisateurs au sein de chaque administration, et agréger les échanges au sujet de chaque jeu de données, afin d'éviter la répétition de questions aux producteurs. Cela permettrait de démultiplier l'effort que les producteurs de données réalisent pour expliquer le contexte des données collectées aux réutilisateurs.

Organiser des séminaires ou ateliers de travail ponctuels sur les sujets techniques d'intérêt pour la communauté. La technicité du sujet nécessite de façon occasionnelle la tenue de séminaires à destination de la communauté des producteurs et des réutilisateurs de données publiques.

Proposition 12 : Mettre en place des assises des données publiques et animer la communauté des réutilisateurs.

Regrouper l'ensemble de l'écosystème pour des assises des données publiques. Au-delà d'ateliers de travail ponctuels, il paraît important de regrouper, à intervalles réguliers, les différents acteurs concernés par la politique de réutilisation des données publiques, pour des colloques et des sessions de travail permettant l'échange de points de vue et la prise de conscience des évolutions nécessaires.

Il semble préférable d'impliquer tout à la fois les acteurs de l'innovation (pôles de compétitivité, capitaux risqués, secteur numérique), les associations, les institutions publiques et les collectivités territoriales, mais aussi l'administration elle-même, dont le travail est modifié et mis en valeur par la réutilisation des données publiques.

Animer la communauté des réutilisateurs. Une des conclusions des premières expériences de réutilisation de données publiques est que, sans animation, la mise à disposition des données publiques produit peu²⁰⁹.

Le rôle d'animateur de la communauté des développeurs et des réutilisateurs de données publiques, à travers la participation à des événements et la mise en avant des réalisations marquantes ou utiles aux citoyens, est donc central à une politique ambitieuse des données publiques.

²⁰⁹ Entretien avec Charles Népote, FING.

3.2. Encourager la réutilisation des données publiques

Proposition 13 : Encourager le développement d'applications à partir des données publiques.

Organiser un concours national de développement d'applications. L'effet de levier potentiel d'un concours de développement d'application est tel (estimé à 50 pour 1 pour la première expérience en 2008 à Washington, D.C.) qu'il apparaît comme un élément indispensable d'encouragement.

Nous suggérons de s'inspirer du modèle des concours de développement d'applications, notamment ceux organisés par l'IGN (été 2010) et les mairies de Rennes ou de Paris (automne 2010 et hiver 2010-2011), au niveau national, et la Banque Mondiale (hiver 2010-2011) et l'Open Knowledge Foundation (avril à juin 2011) au niveau international. Il est clairement nécessaire de consacrer des moyens matériels assez importants pour créer une masse critique de postulants.

Dans un second temps des concours thématiques pourraient être mise en place dans le domaines de la santé, de l'éducation, de l'environnement, etc. à l'initiative des ministères et agences concernés et encourageant des partenariats d'organisation avec le secteur privé et la société civile.

Intégrer la réutilisation des données publiques aux appels à projets de R&D numérique. Nous suggérons d'encourager dès à présent la prise en compte de l'utilisation judicieuse qui pourrait être faite des données publiques dans l'ensemble des appels à projets de recherche et développement lancés en soutien au développement du secteur du numérique.

Pérenniser l'offre d'applications créées à partir des données publiques. Nous suggérons d'investir dans le suivi de l'offre d'applications créées à partir des données publiques, et dans la communication de leur utilité auprès du public.

Un appui en terme de communication et un soutien financier permettraient notamment de multiplier l'impact de concours de développement d'applications éventuels, et de pérenniser les applications utiles aux citoyens qu'ils auraient contribué à faire émerger.

Proposition 14 : Intégrer au débat démocratique les données publiques et leur réutilisation.

Communiquer aux administrations l'importance des données pour mettre en lumière leurs métiers. Pour permettre la mise en ligne d'une offre de données publiques riche et qui éclaire le débat public, il est en premier lieu nécessaire de convaincre les administrations du potentiel des données publiques pour permettre la mise en lumière de leurs métiers et de leur domaine d'action. Un travail de pédagogie sur la notion de réutilisation paraît aussi indispensable. Il semble notamment nécessaire de mettre à mal la notion selon laquelle « mes données ne peuvent intéresser personne d'autre », et de faire prendre conscience du retard qui risque de s'accumuler si les institutions publiques tardent à mettre en ligne leurs données publiques.

Illustrer la réutilisation des données publiques par le développement de services aux citoyens. Il semble important que le public soit tenu au courant, à la fois des efforts consentis par l'administration, et des bénéfices concrets qui en découlent. On pourra par exemple labelliser des applications rendues possibles grâce à la réutilisation des données et au travail de l'administration.

L'État lui-même aura un rôle d'exemple et de démonstration à jouer en la matière, et pourrait choisir de développer lui-même à cet effet un certain nombre de services qui illustrent certains jeux de données publiques qui touchent les citoyens et leurs préoccupations.

Faire des données et de leur réutilisation un outil d'éclairage du débat public. Nous suggérons d'encourager par exemple la création, par des développeurs ou des journalistes, de visualisations et d'infographies interactives à partir des jeux de données publiques, qui pourraient avoir un rôle de pédagogie auprès d'un public bien plus large que la simple cible des réutilisateurs, et permettraient de mettre en perspective les grands enjeux des politiques publiques.

3.3. Élargir le champ de l'Open Data

Proposition 15 : Développer une stratégie nationale à partir du rôle de coordination des institutions publiques de la mission Etalab.

Mutualiser des solutions de mise en ligne pour les administrations et les collectivités.

Pour faciliter l'ouverture des données des collectivités, il serait souhaitable de mettre en place des solutions techniques mutualisées, en particulier pour éviter la multiplication de dépenses superflues.

Par exemple, un des rôles de la future plateforme de données publiques de l'État data.gouv.fr pourrait être de proposer des espaces dédiés aux collectivités qui souhaiteraient y publier leurs données. Ces espaces pourraient enfin être exportables, afin de figurer ensuite sur les sites des mairies et collectivités qui voudraient en disposer.

Développer une stratégie nationale pour les données publiques. La mission Etalab a parmi ses prérogatives un rôle de coordination des administrations. Pour favoriser la réutilisation des données publiques de l'ensemble des institutions publiques et des collectivités, ce rôle de coordination pourrait être étendu à l'établissement d'une stratégie nationale visant à faciliter la mise à disposition de toutes les données publiques françaises, et d'encourager à leur réutilisation.

Cette stratégie pourrait bien sûr s'appuyer sur le portail data.gouv.fr, mais aussi sur un choix de licences gratuites, sur des recommandations techniques notamment en terme de formats de diffusion des données, et sur l'animation de l'ensemble de la communauté des réutilisateurs.

Elle permettrait de faciliter non seulement la réutilisation des données publiques de l'État, mais aussi de celles des données des communes et des départements, particulièrement importantes pour les citoyens du fait de leur plus grande proximité.

Une véritable stratégie nationale aurait enfin pour bénéfice de permettre le croisement de données de l'État avec celles des collectivités, ce qui pourrait donner lieu à de nouveaux services plus riches et plus innovants.

Proposition 16 : Élargir le champ de l'Open Data.

Élargir le champ de la libre réutilisation des données publiques. Face aux régimes dérogatoires et aux exceptions au principe de la réutilisation la plus libre des données publiques, nous recommandons d'élargir le plus possible le champ de la réutilisation qui entre dans les critères de l'Open Data. En

particulier, les données qui bien qu'à caractère industriel et commercial représentent un intérêt fort pour le public devraient être largement ouvertes à la réutilisation libre et gratuite.

Explorer l'application de l'Open Data aux entreprises privées. La question de la réutilisation des données privées, encore à un stade prospectif, a vu émerger le concept de « *Data Portability* », l'idée que les utilisateurs de services (par exemple des services en ligne, ou un abonnement de téléphonie mobile) pourraient demander d'accéder à leurs données d'usage dans des formats réexploitables.

Une généralisation de ce principe serait une avancée en faveur des consommateurs. Ils sont sensibles à la fois aux questions de vie privées sous-jacentes et à la possibilité de suivre leur consommation de la façon la plus transparente, afin de réduire l'asymétrie de l'information. Cela doit ainsi leur permettre à terme d'optimiser leurs choix économiques.

L'obligation de communiquer les données d'usage aux consommateurs dans des formats réexploitables serait surtout un enjeu fort de libre choix des consommateurs, car il permettrait des comparaisons beaucoup plus faciles entre offres commerciales et entre fournisseurs de service. Ce serait donc un élément potentiel à part entière des politiques de concurrence.

Développer une stratégie européenne et internationale. Pour réaliser le plein potentiel des données publiques et de leur réutilisation, il nous paraît indispensable que la France noue des liens forts avec les autres États autour de la question des données publiques.

Au niveau Européen, les évolutions éventuelles de la Directive PSI pourraient constituer un levier d'action puissant en faveur de conditions de réutilisation plus propices à l'innovation et au développement de nouveaux services aux citoyens.

Enfin, l'établissement de normes d'interopérabilité, ou tout au moins l'échange de bonnes pratiques quant à la structuration des données, aux ontologies utilisées et aux formats de diffusion, pourraient permettre, au niveau international le plus large, de poser les bases du Web des Données et de l'Internet de demain.

Nous espérons que ces suggestions pourront s'avérer utiles dans le processus de mise en place, à tous les niveaux, de l'ouverture et de la réutilisation des données publiques.

L'ouverture des données de l'État et de la sphère publique est un projet de longue haleine. C'est un premier pas vers l'Internet de demain et un élément important du développement du numérique. C'est un enjeu stratégique pour l'avenir de notre économie.

C'est enfin un sujet de société novateur et important, dont les conséquences, notamment sur la transparence de l'État, ne seront pas sans créer des bouleversements.

Mais l'*Open Data* est avant tout une transformation profonde de la société, et un outil puissant de politique publique pour faire face aux grands défis de demain. L'État et les gouvernants doivent s'en saisir pleinement et avec ambition.

Bibliographie

- AEC (Aquitaine Europe Communication), 2010, « Les données publiques : Guide pratique et juridique », Guide juridique n°4, l'Agence des Initiatives Numériques.
<http://www.aecom.org/>
- APPLE INSIDER, 2010, « Apple says App Store has made developers over \$1 billion ». Juin 2010.
http://www.appleinsider.com/articles/10/06/07/apple_says_app_store_has_made_developers_over_1_billion.html
- ARROW, K., 1962, « Economic Welfare and the Allocation of Resources for Innovation ».
- BAUDREZ, M-L., CHANTEPIE, P., MEURISSE, J., 2009, « Partager notre patrimoine culturel », Ministère de la Culture et de la communication.
- BANCILHON, F. et GANS, B., 2011, « Size and Structure of the French Public Sector Information »
http://gov.opendata.at/site/sites/default/files/Size%20and%20Structure%20of%20French%20PSI_0.pdf
- BENGHOZI, P.-J., BUREAU, S. et MASSIT-FOLLEA, F., 2009, « L'Internet des objets : Quels enjeux pour l'Europe ? ». Maison des sciences de l'homme, Paris.
- BENHAMOU, B., 2010, « Eléments pour une stratégie numérique européenne », Délégation aux Usages de l'Internet, 15 mars 2010.
- BENHAMOU, B., 2010, « Les concours d'applications mobiles dans le monde », Délégation aux Usages de l'Internet, 21 octobre 2010.
- BERNERS LEE T., 1994, World Wide Web Consortium (W3C). « Uniform Resource Locators (URL): A Syntax for the Expression of Access Information of Objects on the Network ».
<http://www.w3.org/Addressing/URL/url-spec.txt>
- BESSON, E., 2008, « France numérique 2012 : Plan de développement de l'économie numérique », octobre 2008.
<http://lesrapports.ladocumentationfrancaise.fr/BRP/084000664/0000.pdf>
- BILLAUT, J-M., 26 novembre 2010, « Connaissez-vous Eduardo Larrain from Paris ? KelQuartier, premier utilisateur d'Open Data ? »,
<http://billaut.typepad.com/jm/2010/11/connaissiez-vous-eduardo-larrain-from-paris-.html>
- BLANC, S., 2010, « La libération des données : On ne reviendra pas en arrière », OWNI.fr, <http://owni.fr/2010/06/14/liberation-des-donnees-on-ne-reviendra-pas-en-arriere/>
- BROWN, J., DUGUID, P., 2000, « The social life of information », Boston: Harvard Business School Press.
- CAMERON, D., 2010, Allocution à l'occasion de la mise en ligne des données de dépenses publiques à partir de 25,000 livres.
<http://www.guardian.co.uk/politics/blog/2010/nov/19/government-spending-files-live-blog?intcmp=239>
- CHAUMOND J., 2010, « Social Commerce : Quand le E-Commerce Rencontre le Web d'Aujourd'hui ».
- CISCO, 2010, « Cisco Visual Networking Index Global Mobile Data Traffic Forecast Update, 2009-2014 ».
http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html
- COMMITTEE ON GEOPHYSICAL AND ENVIRONMENTAL DATA, 1995, « On the full and open exchange of scientific data », National Research Council, Washington, D.C.
<http://www.nap.edu/readingroom.php?book=exch&page=summary.html>
- CONSEIL D'ANALYSE STRATEGIQUE, « Rapport thématique France 2025 : Production et emploi ». Gilbert Cette, Président de groupe.
<http://www.strategie.gouv.fr/IMG/pdf/4Productionemploi.pdf>
- CRAGLIA M., Coordonateur technique d'Inspire pour la Commission Européenne (centre commun de recherche). Séminaire « Innovations et développement du secteur de l'information géographique » organisé le 5 octobre 2009 à la Grande Arche de la Défense par le Conseil National de l'Information Géographique (CNIG), l'AFIGÉO et l'Institut des Sciences et des Techniques de l'Équipement pour le Développement (ISTED).

- DAVID P., 2000, « A tragedy of the public knowledge 'commons' ? », Oxford & Stanford University, <http://129.3.20.41/eps/dev/papers/0502/0502010.pdf>
- DEKKERS, M. et al., 2006, « MEPSIR, Measuring Public Sector Information Resources, Final Report of Study on Exploitation of public sector information », http://ec.europa.eu/information_society/policy/psi/docs/pdfs/mepsir/final_report.pdf
- DIRECTION GENERALE DE LA MODERNISATION DE L'ÉTAT, 2010, « Innover au service du citoyen », [http://www.rgpp.modernisation.gouv.fr/index.php?id=51&tx_ttnews\[tt_news\]=531&tx_tnews\[backPid\]=11&cHash=89b9c6970f](http://www.rgpp.modernisation.gouv.fr/index.php?id=51&tx_ttnews[tt_news]=531&tx_tnews[backPid]=11&cHash=89b9c6970f)
- FAURE, P., Vice Président du Conseil Général de l'Industrie, de l'Energie et des Technologies, explique l'importance du « green IT » aux Assises du Numérique 2009 le 1er octobre 2009. http://www.dailymotion.com/video/xaov8c_pascal-faure-assises-du-numerique-2_news
- FIORETTI, M., 23 octobre 2009, « Walkability: Check It Before Choosing Your Next Home », <http://stop.zona-m.net/2009/10/walkability-check-it-before-choosing-your-next-home/>
- GALLON, C., 2011, « Carte contributive sur l'ouverture des données en France », Libertic, <http://libertic.wordpress.com/2011/02/11/carte-contributive-sur-louverture-des-donnees-publiques-en-france/>
- GARRETT, J.J., 2005. « Ajax : A New Approach To Web Applications », <http://www.adaptivepath.com/ideas/essays/archives/000385.php>
- GFII, Groupement Français des Industries de l'Information. 20 mai 2010. « L'ouverture des données publiques : un enjeu pour le développement de l'économie numérique. », Colloque tenu à la Maison de l'Europe, http://www.gfii.asso.fr/article.php?id_article=3226
- GFII, Groupement Français des Industries de l'Information, 2010, « PSI Re-use in France : Overview and Recent Developments », http://www.gfii.asso.fr/article.php?id_article=3315
- GIGAOM, 2010, « GeoLocal : The Rise of Consumer Location-Based Services », 28 mai 2010. <http://gigaom.com/mobile/geolocal-the-rise-of-consumer-location-based-services/>
- GIGAOM, 2010, « MyCityWay Dives Into Data, Comes Up With Gold », 19 novembre 2010. <http://gigaom.com/2010/11/19/mycityway-dives-into-data-comes-up-with-gold/>
- JOSPIN, L., 1997, « Préparer la France pour la Société de l'Information ». Discours à Hourtin le 25 août 1997. <http://www.admiroutes.asso.fr/action/theme/politic/lionel.htm>
- JUPPE, A., ROCARD, M., 2009, « Investir pour l'Avenir : Priorités stratégiques d'investissement et emprunt national ». http://www.commission-investissement.fr/IMG/pdf/Rapport_191109.pdf
- KAPLAN, D., 2010, « Les données publiques : et après ? », Internet Actu, 9 novembre 2010 <http://www.Internetactu.net/2010/11/09/louverture-des-donnees-publiques-et-apres/>
- KIRKPATRICK, M., 2010, « Objects Outpace New Human Subscribers to AT&T, Verizon », ReadWriteWeb, 10 août 2010, http://www.readwriteweb.com/archives/objects_outpace_new_human_subscribers_to_att_veri_z.php
- LACOMBE, R., 2010, « Tocqueville 2.0 and Open Government Data ». LeWeb 10 Ignite Session. <http://www.youtube.com/watch?v=6qj8jS2Xsfk>
- LAFFONT, J. et TIROLE, J., 1993, « A Theory of Incentives in Procurement », MIT Press : Cambridge, MA.
- LANDIER, A. et THESMAR, D., 2010, La Société Translucide : Pour en Finir avec le Mythe de l'État Bienveillant.
- LESSIG, L., 2009, « Remix : Making art and commerce thrive in the hybrid economy ». <http://remix.lessig.org/>
- L'INFORMATICIEN, 2011, « Interview de Séverin Naudet, Directeur d'Etalab : Encourager l'innovation par toute la communauté des développeurs et des entrepreneurs pour soutenir le développement de l'économie numérique ». Numéro 91, mai 2011.
- M. LE PREMIER MINISTRE, 2011, « Création de la mission Etalab, chargée de la mise en ligne de data.gouv.fr », communiqué de presse du 22 février 2011. http://www.gouvernement.fr/sites/default/files/communiqués/02.22_Communique_de_presse_-_Etalab.pdf
- M. LE PRESIDENT DE LA REPUBLIQUE, 2011, « Discours à l'occasion de l'installation

du Conseil National du Numérique le 27 avril 2011».

<http://www.elysee.fr/president/les-actualites/discours/2011/discours-sur-Internet-et-l-economie-numerique.11256.html>

MARTINEZ, R., 2010, «PSI Re-use in France: Overview and Recent Developments »,

http://www.epsiplatform.eu/topic_reports/topic_report_no_10_psi_re_use_in_france_overview_and_recent_developments

MAUDE, F., Ministre du Cabinet Office, allocution du 19 novembre 2010, conférence de presse à l'occasion de la mise en ligne des entrées de 25,000 livres et plus du budget de fonctionnement de l'État britannique :

<http://www.guardian.co.uk/commentisfree/2010/nov/19/francis-maude-government-data-published?intcmp=239>

MCKINSEY & COMPANY, 2011, « Impact d'Internet sur l'économie française : Comment Internet transforme notre pays » :

http://www.mckinsey.com/locations/paris/home/Rapport_complet_Impact_final.pdf

NEW ENGLAND JOURNAL OF MEDICINE, 2007, « Public Reporting and Pay for Performance in Hospital Quality Improvement ». January 26, 2007.

<http://www.nejm.org/doi/full/10.1056/NEJMsa064964#t=articleBackground>

NEW YORK TIMES, 23 avril 2011. « Show Us The Data (It's Ours After All) ». Richard H. Thaler.

<http://www.nytimes.com/2011/04/24/business/24view.html>

ORANGE LABS, 2010, « Web des Données et Données Ouvertes » : <http://www.orange-innovation.tv/dreamorange/resources/documents/4022/>

[OpenData OrangeLabs Sofrecom 2009.pdf](http://www.orange-innovation.tv/dreamorange/resources/documents/4022/)

O'REILLY, T., 2010, Open Government, O'Reilly editions, disponible sur :

<http://opengovernment.labs.oreilly.com/>

OTTERY, C., 2010, « Partage des Données pour Mieux Réagir Face aux Catastrophes Naturelles », 12 décembre 2010, OWNI.

<http://owni.fr/2010/12/12/partager-des-donnees-pour-mieux-reagir-face-aux-catastrophes-naturelles/>

OYE, K.A., WELLHAUSEN, R., 2009, « The Intellectual Commons and Property in Synthetic Biology ».

PARSONS, E., 2009. Geospatial Technologist for Europe, Middle East and Africa chez Google. Séminaire « Innovations et développement du secteur de l'information

géographique » organisé le 5 octobre 2009 à la Grande Arche de la Défense par le Conseil National de l'Information Géographique (CNIG), l'AFIGÉO et l'Institut des Sciences et des Techniques de l'Équipement pour le Développement (ISTED).

PENIN, J., et al., « La valorisation des informations du secteur public (ISP) : un modèle économique de tarification optimale », https://www.apiefrance.fr/sections/acces_thematique/reutilisation-des-informations-publiques/etude-economique/view

PETTIFER, R. 25 août 2009. « Europe loses 300 million Euro in taxation per year. » ePSI Platform.

http://www.epsiplatform.eu/guest_blogs/europe_looses_300_million_euro_per_year_in_taxation

PEUGEOT, V., 2010, « Les enjeux publics, économiques et citoyens de l'ouverture des données : l'expérience britannique »,

http://academia.edu.documents.s3.amazonaws.com/1850987/CommunicationDocSoc2010_VPeugeot_opendata.pdf

POLLOCK, R., 2008, « The Economics of Public Sector Information. », Oxford University Department of Economics.

http://www.rufuspollock.org/economics/papers/economics_of_psi.pdf

RAO, L., 2010, « Boom ! 5 Million Users Check-In to Foursquare », Techcrunch, 1er décembre 2010.

<http://techcrunch.com/2010/12/01/foursquare-hits-5-million-users/>

RATTI, C., 2009. « Digital Cities : 'Sense-able' urban design. », 2 octobre 2009, Wired UK Magazine,

<http://www.wired.co.uk/magazine/archive/2009/11/features/digital-cities-sense-able-urban-design>

REFLECTION GROUP ON BRINGING EUROPE'S CULTURAL HERITAGE ONLINE, 2011, « The New Renaissance: Report of the 'Comité des sages' ».

http://ec.europa.eu/information_society/activities/digital_libraries/doc/refgroup/final_report_cds.pdf

REGARDS SUR LE NUMERIQUE, 17 mars 2011, « Data.gouv.fr : La France va se lancer en version bêta avant fin 2011 ».

<http://www.rslnmag.fr/blog/2011/3/17/data-gouv-fr-la-france-va-se-lancer-en-version-beta-avant-fin-2011/>

RIESTER, F. et al., 2010, « Améliorer la Relation Numérique à l'Usager : Rapport Issu des Travaux du Groupe Experts Numériques ».

http://www.modernisation.gouv.fr/fileadmin/Mes_fichiers/pdf/RapportGroupeExpertsNumeriques.pdf

RONAI, M., 1996, « Données publiques : accès, diffusion, commercialisation. » Revue Problèmes Politiques et Sociaux : Dossiers d'Actualité Mondiale. Nos 773-774, 1^{er} novembre 1996. La Documentation Française.

SEGARAN, T., 2009, « Programming the Semantic Web », O'Reilly éditions.

SHADBOLT, N., 2011. Allocution au colloque « *Open Data* : Et Nous, Et Nous, Et Nous ? » du 17 mars 2011 organisé par Regards sur le Numérique. Présentation : <http://www.slideshare.net/RSLN/opendata-datagovuk-how-did-we-do-it>

SHAH, S., 2010, « Cannibalize Business Development by Popularizing your API », <http://shaivalshah.com/cannabilize-business-development-by-populariz>

SHAPIRO, C. et VARIAN, H., 1998, « Information Rules: A Strategic Guide to the Network Economy », HBS Press.

SIEGLER, M.G., 2010, « Marissa Mayer's Next Big Thing : 'Contextual Discovery' – Google Results Without Search », 8 décembre 2010, Techcrunch, <http://techcrunch.com/2010/12/08/googles-next-big-thing/>

STIGLITZ, J. et ORSZAG, P., 2000, « The Role of Government in a Digital Age ». Octobre 2000. http://archive.epinet.org/real_media/010111/materials/stiglitz.pdf

TAPSCOTT, D. et WILLIAMS, A. 2006. « Wikinomics : How Mass Collaboration Changes Everything. »

TESTER, J.W. et al, 2005. « Sustainable Energy: Choosing Among Options », MIT Press, Cambridge, MA. <http://mitpress.mit.edu/catalog/item/default.asp?tid=10566&ttype=2>

THE ECONOMIST. 25 février 2010. « The Data Deluge : Businesses, Governments and Society are Only Starting to Tap Its Vast Potential ». <http://www.economist.com/node/15579717>

THOENIG, J-C., 4 septembre 2010, Conférence aux élèves du Corps des Ponts, des Eaux et des Forêts.

U.S. NATIONAL RENEWABLE ENERGY LAB, International Solar Resource Data. http://www.nrel.gov/csp/troughnet/solar_data.html

VERDIER, H., 2010, « Big Data : Making sense at scale. », 14 décembre 2010. <http://www.henriverdier.com/2010/12/big-data-making-sense-at-scale.html>

WAGNER, R., 2003, « Information wants to be free: intellectual property and the mythologies of Control ». http://papers.ssrn.com/sol3/papers.cfm?abstract_id=419560

WATTERS, A., 2010. The Age of Exabytes : Tools & Approaches for Managing Big Data, ReadWriteWeb. <http://www.readwriteweb.com/reports/big-data/>

TECHCRUNCH, 2010. « Smartphone App Market Reached More Than \$2.2 Billion in the First Half of 2010 », 20 août 2010, <http://techcrunch.com/2010/08/20/smartphone-applications-market-size/>

TECHCRUNCH, 2010, « Global Smartphone App Download Market Could Reach \$15 Billion by 2013 : Report », 5 mars 2010, <http://techcrunch.com/2010/03/05/global-smartphone-app-download-market-could-reach-15-billion-by-2013-report/>

WU, T., 2010, « The Master Switch: The Rise and Fall of Information Empires », Knopf éd.

Textes réglementaires

(i) Textes européens :

- Directive 90/313/CEE du Conseil, du 7 juin 1990, concernant la liberté d'accès à l'information en matière d'environnement. Abrogée par :
- Directive 2003/4/CE du Parlement et du Conseil du 28 janvier 2003 concernant l'accès du public à l'information en matière d'environnement.
- Transposition en droit français : Loi n°2005-1319 du 26 octobre 2005 portant diverses dispositions d'adaptation au droit communautaire dans le domaine de l'environnement.
- Directive PSI du Parlement européen et du Conseil n°2003/98/CE du 17 novembre 2003 concernant la réutilisation des informations du secteur public.
- Transposition en droit français par l'ordonnance n°2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques.
- Directive Inspire du Parlement européen et du Conseil n° 2007/2/CE du 14 mars 2007 établissant une infrastructure d'information géographique dans la Communauté européenne.
- Transposition en droit français dans les articles L. 127-1 à L. 127-10 du code de l'environnement introduits par l'ordonnance n° 2010-1232 du 21 octobre 2010.
- Décision n° 247769 du Conseil d'État en date du 24 juillet 2006.
<http://www.easydroit.fr/jurisprudence/Conseil-d-Etat-10eme-et-9eme-sous-sections-reunies-24-07-2006-247769-247769/J49598>
- Circulaire du 18 octobre 2007 relative à la mise en œuvre des dispositions régissant le droit d'accès à l'information relative à l'environnement.
http://www.toutsurlenvironnement.fr/files/circulaire_aarhus.pdf.
- Arrêté du 22 Janvier 2009 fixant le montant des rémunérations dues en contrepartie de la cession des licences de réutilisation de données de la base de données informatique du Ministère de l'économie, de l'industrie et de l'emploi relative aux prix des carburants.
http://www.prix-carburants.gouv.fr/document/Arrete_du_22_janvier_2009.pdf
- Arrêté du 15 juin 2010 relatif à la mise à disposition du public pour un usage de rediffusion de la base de données électronique de l'Institut national de la statistique et des études économiques dénommée SIRENE.
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000022377461&fastPos=1&fastReqId=448289598&categorieLien=cid&oldAction=rechTexte>
- Décret n° 2011-194 du 21 février 2011 portant création d'une mission « Etalab » chargée de la création d'un portail unique interministériel des données publiques.
<http://legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000023619063&categorieLien=id>
- Circulaire du 26 mai 2011 relative à la création du portail unique des informations publiques de l'Etat « data.gouv.fr » par la mission « Etalab » et l'application des dispositions régissant le droit de réutilisation des informations publiques.

(ii) Textes nationaux :

- Loi n°78-753 du 17 juillet 1978 sur l'accès aux documents et données publiques.
- Ordonnance n°2005-650 du 6 juin 2005 modifiant la loi n°78-753 du 17 juillet 1978 sur l'accès aux documents et données publiques.

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000024072788&dateTexte=&categorieLien=id>

- Décret n° 2011-577 du 26 mai 2011 relatif à la réutilisation des informations publiques détenues par

l'Etat et ses établissements publics administratifs.

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000024072772&categorieLien=id>

Interlocuteurs

(i) Rencontres dans le cadre d'entretiens particuliers

APIE : Mme Danielle Bourlange, directrice-adjointe, le 5 novembre 2010.

CanalTP : M. Guillaume Crouigneau, directeur délégué, le 12 novembre 2010.

Cap Digital : M. Henri Verdier, président, le 12 novembre 2010.

FING (Fondation Internet Nouvelle Génération) : M. Charles Nepote, le 25 novembre 2010.

IGN : M. Jean-Yves Bréard, directeur par intérim de la Maîtrise d'Ouvrage Déléguée du Service Public, le 17 décembre 2010.

INSEE : M. Pierre Audibert, directeur de la diffusion et de l'action régionale, et M. Thierry Aouizerate, chef de département INSEE info service, le 26 novembre 2010.

LATTS (Laboratoire Techniques, Territoires et Sociétés) de l'ENPC : M. Sylvain Parasio, chercheur en sociologie, le 18 octobre 2010.

Maurice Ronai, auteur du rapport sur les données publiques à la Documentation Française en 1996, le 3 décembre 2010.

Ministère de la Justice : M. Thomas Saint Aubin, auteur de la Licence Information Publique, le 23 décembre 2010.

NavX : M. Jean Cherbonnier, PDG, le 22 décembre 2010.

Regards Citoyens : M. Tanguy Morlier, et M. Benjamin Ooghe, cofondateurs de l'association, le 9 novembre 2010.

Secrétariat Général du Gouvernement : M. Stéphane Cottin, chargé de mission auprès du cabinet du Secrétaire général, le 22 octobre 2010.

(ii) Rencontres à Londres le 18 et 19 Novembre 2010 (« Open Government Data Camp »)

Data Publica : M. François Brancillon, Directeur et Fondateur.

ONU : M. Keping Yao, Division for Public Administration and Development

Management, United Nations Department of Economic and Social Affairs.

Open Knowledge Foundation : M. Jonathan Gray, Community Manager.

Suez Environnement : Guillaume Féry, responsable Innovation et Protection de l'Environnement.

Timetric : Andrew Walkingshaw, co-founder and CEO, and Simon Briscoe, VP for Product.

UK Cabinet Office : James Forrester, Transparency & Digital Engagement, Efficiency and Reform Groupe, in charge of Data.gov.uk.

US Office of the Chief Information Officer : Joey A. Hutcherson, Data.gov.

World Bank : Eric Swanson, Program Manager for Global Monitoring & WDI, Development Data Group.

(iii) Rencontres à Rennes le 29 novembre 2010

Commission Européenne : Andrej Osterman, direction générale Société de l'information et médias.

InCité Solution : Pierre-Emmanuel Muller, directeur associé.

LiberTIC : Claire Gallon, fondatrice.

Ministère du Budget : Mme Frédérique Summing, Service innovation à la direction Générale de la Modernisation de l'État.

Rennes Métropole : M. Xavier Crouan, directeur des informations.

(iv) Contacts par téléphone ou par courriel

MEEDDM : M. Laurent Coudercy, ancien responsable du portail de l'Environnement.

(v) Portails régionaux

ADACL40 (Aide aux Collectivités Locales du Département 40) : Mme Marie-Hélène Lalande, responsable du projet IGECOM 40.

CETE Ouest : Mme Catherine Chamard-Bois, consultant méthodologie et géomatique Division Ville et Territoires.

CIGAL : Guillaume Ryckelynck, chargé de mission.

CRAIG (Centre Régional Auvergnat de l'Information Géographique).

CRIGE PACA : Mme Christine Archias, directrice.

Géobretagne : Mme Lydie Vinsonneau, Conseil régional de Bretagne, Pôle Système d'Information Géographique direction de l'aménagement.

Région Basse Normandie : M. Benjamin Lecointe, Dir. Aménagement et D.D.

Liens et références

(i) Différents acteurs

<http://www.afi.geo.asso.fr/>
<https://www.apiefrance.fr/>
<http://www.apur.org/>
<http://www.cada.fr/>
<http://www.cnil.fr/la-cnil/>
<http://www.data-publica.com>
<http://data.gouv.fr/>
<http://blog.etalab.gouv.fr/>
<http://fing.org/>
http://www.gfii.asso.fr/rubrique.php3?id_rubrique=60
<http://libertic.wordpress.com/libertic/>
<http://okfn.org/>
<http://www.opendatafoundation.org/>
<http://owni.fr>
<http://www.regardscitoyens.org/>

(ii) Site de diffusion de données

<http://opendata.paris.fr/>
<http://canadp-archivesenligne.paris.fr/>
http://www.certu.fr/fr/catalogue/product_info.php?products_id=3817&language=fr
<http://www.geocatalogue.fr/>
<http://www.insee.fr/fr/default.asp>
<http://www.ign.fr/>
http://www.loire.fr/display.jsp?id=c_828986
<http://professionnels.ign.fr/2/telechargement.htm>
<http://www.toutsurlenvironnement.fr/>
<http://www.govdata.eu/>

(iii) Licences

<http://www.paris.fr/portail/viewmultimediacument?multimediacument-id=94658>
http://www.legifrance.gouv.fr/html/licences/licences_notice.htm

http://www.rip.justice.fr/information_publicque_librement_reutilisable

<http://www.opendatacommons.org/licenses/odbl/>

(iv) Réutilisations intéressantes

<http://caselex.com>
<http://www.guardian.co.uk/>
http://www.paris.fr/accueil/actualites-municipales/du-neuf-sur-opendata-paris-fr/rub_9656_actu_99478_port_23785
<http://www.husetseb.dk/about>
<http://lexisnexis.com>
<http://www.météo-paris.com>
<http://www.navx.com>
<http://www.openstreetmap.fr/>
<http://www.rte-france.com/fr/developpement-durable/maitriser-sa-consommation-electrique/eco2mix-consommation-production-et-contenu-co2-de-l-electricite-francaise>

(v) Références diverses

<http://148apps.biz/app-store-metrics/?mpage=appcount>
<http://www.ac-grenoble.fr/ecogest/pedago/administration/prodpeda/mcd.htm>
<http://www.acteurspublics.com/article/28-09-10/les-portails-etrangers-au-banc-d-essai>
<http://www.advancity.eu/>
<http://aws.amazon.com/ec2/>
<http://biobricks.org>
<http://www.canaltp.fr/>
http://cordis.europa.eu/fp7/i2010_fr.html
<http://creativecommons.org.au/weblog/entry/269>
<http://datalift.org/>
<http://www.dbpedia.org>
<http://www.debatpublic.fr/>

<http://georss.org>

<http://www.henriverdier.com>

<http://lod2.eu>

<http://loopt.com>

<http://www.moma.org/interactives/exhibitions/2008/elasticmind/>

<http://okfn.org/>

<http://www.opengovdata.org/home/8principles>

<http://www.opendefinition.org/okd/>

<http://en.openei.org/>

<http://www.pcinpact.com/actu/news/57602-telephonie-mobile-Internet-arcep-orange.htm>

<http://www.quora.com>

<http://re.jrc.ec.europa.eu/pvgis/apps4/pvest.php>

<http://www.startuplessonslearned.com/>

<http://www.urbanmobs.fr/fr/france/>

<http://www.w3.org/People/Berners-Lee/>

<http://www.w3.org/Protocols/>

http://en.wikipedia.org/wiki/Raw_data

http://en.wikipedia.org/wiki/Virtual_globe

http://fr.wikipedia.org/wiki/Loi_de_Metcalf

<http://fr.wikipedia.org/wiki/BarCamp>

http://world.waze.com/guided_tour/

<http://www.youtube.com/watch?v=6qj8jS2Xsfk>



École des Ponts ParisTech. 2011. *Pour une politique ambitieuse des données publiques : Les données publiques au service de l'innovation et de la transparence*. Rapport à la Délégation aux usages de l'Internet, remis au Ministre de l'Industrie, de l'Energie et de l'Economie numérique le 13 juillet 2011.

Vous pouvez contacter les auteurs de ce rapport sur
<http://rapport-opendata-2011.fr>



École des Ponts ParisTech. 2011. *Pour une politique ambitieuse des données publiques : Les données publiques au service de l'innovation et de la transparence.* Rapport à la Délégation aux usages de l'Internet, remis au Ministre de l'Industrie, de l'Energie et de l'Economie numérique le 13 juillet 2011.