



L'Open Data : universalité du principe et diversité des expériences ?

Danièle Bourcier, Primavera De Filippi

► To cite this version:

Danièle Bourcier, Primavera De Filippi. L'Open Data : universalité du principe et diversité des expériences ?. La semaine juridique - édition générale , LexisNexis, 2013, pp.1-9. <hal-01026106>

HAL Id: hal-01026106

<https://hal.archives-ouvertes.fr/hal-01026106>

Submitted on 22 Jul 2014

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

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

Administration / Citoyens Étude

L'Open Data : universalité du principe et diversité des expériences ?

L'objet de cet article est d'explorer le modèle de « gouvernement ouvert » à partir de la mise à disposition de l'information publique et des expériences d'ouverture des données publiques (*Open Data*) menées par l'administration, les collectivités locales et les établissements publics en Europe et en France. Le principe de l'*Open Data* dans le secteur public a été adopté peu à peu dans la plupart des pays occidentaux. Cependant, les modalités de sa mise en place et l'étendue des données concernées varient suivant les cultures juridiques et politiques. Lors de cette phase en effet, on observe que des particularismes locaux émergent, fondés sur des conceptions différentes de l'*Open Data*, issues des politiques d'accès à l'information, de la notion de patrimoine « public » ou des dispositions nationales du droit d'auteur.

Après avoir fait le point sur la dynamique générale du mouvement, nous comparerons différentes conceptions entre pays, et particulièrement entre la France et les pays anglo-saxons. Puis d'autres particularités seront observées au niveau national, que ce soit au niveau de l'État central ou au niveau local (collectivités territoriales, etc). Ces différences se situent tant au niveau de la définition de données publiques que des régimes de mise à disposition et de réutilisation.

Étude rédigée par

Danièle Bourcier,

Centre d'études et de recherche en sciences administratives et politiques,
CNRS, université Paris II

et :

Primavera De Filippi,

Centre d'études et de recherche en sciences administratives et politiques,
CNRS, université Paris II

Introduction

Dire que le secteur public, sous l'impulsion de l'Union européenne, s'oriente irrésistiblement vers l'ouverture des données ne suffit pas à décrire clairement les politiques qui en ont découlé au fil du temps. En France, par exemple, un premier clivage au sein du secteur public s'est manifesté entre ceux qui voudraient valoriser le patrimoine immatériel de l'État et souhaiteraient conditionner l'exploitation commerciale au paiement d'une redevance, et les tenants du principe d'une mise à disposition des informations publiques à titre gratuit et sans conditions restrictives. Malgré ces mouvements *a priori* contradictoires, la France a été pourtant le premier pays à affirmer le principe général d'ouverture des données publiques dans la loi.

Autre particularité de la situation française : face à l'État central qui élaborait et mettait en place, avec un certain retard par rapport à la Grande-Bretagne par exemple, ces politiques d'ouvertures, les collectivités locales se lançaient dans des projets concrets et ambitieux (Brest, Rennes, Région Aquitaine) au niveau territorial. En avance par rapport à l'État, ces collectivités ont affirmé et appliqué – en arguant du principe de leur autonomie – une politique relativement claire d'ouverture, jouant ainsi un rôle stratégique fondamental dans le mouvement de l'*Open Data*.

Nous analyserons la philosophie de l'*Open Data* et ses différentes déclinaisons, en comparant notamment la culture dite anglo-saxonne et la culture continentale française. Initialement vu comme contribuant efficacement au développement de l'e-démocratie et à une plus grande transparence des relations entre administrations publiques et société civile, l'*Open Data* a ensuite évolué – avec la technologie collaborative du web 2.0 – vers le partage et le traitement généralisé de données, exploitées commercialement ou de façon communautaire.

1. Qu'est-ce que l'Open Data?

A – Différents types de données (Data)

De nombreux termes sont nés autour de l'explosion des données à laquelle on assiste sur le réseau : *Open Data*, *Metadata*, *Raw Data*, *Big Data*, *Small Data*, *Linked Data*, *Data Sharing*. Nous tenterons ici d'en élucider le sens afin de mieux comprendre la façon dont ces différents typologies de données interagissent les unes avec les autres, quelles sont les relations qui se forment et les interactions qui se développent dans un déluge de données provenant de disciplines et de milieux totalement différents : données juridiques, données statistiques, données scientifiques, données administratives, données géographiques.

1) Au niveau structurel : *Data*, *Metadata*, *Big Data*, *Linked Data*

Pendant longtemps, on a distingué données et informations : alors qu'une information implique un contexte afin de pouvoir aboutir à une certaine interprétation de la réalité, une donnée est – par définition – ce qui est « donné » : une description élémentaire (brute) qui n'a pas encore été interprétée, ni contextualisée (*Raw Data*). On a aussi parlé de données factuelles (par ex. statistiques ou documentaires) et de données textuelles (par ex. textes de loi) pour en distinguer les traitements. À cet égard, de nombreux débats ont été engagés – notamment dans le milieu des bases de

données juridiques (essentiellement textuelles) – sur la notion de « données brutes » et de « données enrichies »¹. Précisons enfin que la notion de données publiques n'existe pas dans les textes juridiques.

La plupart des données que nous rencontrons sur le réseau ont été constituées au fur et à mesure. Soit elles sont restées isolées en petits ensembles de données, soit elles ont été agrégées à partir de sources différentes, pour être ensuite regroupées en des ensembles structurés qui peuvent parfois faire l'objet de grandes bases de données (*Big Data*). L'état et la qualité des données varient suivant les époques de leur constitution. Leur identité et leur source pouvaient être signalées mais souvent de façon non structurée et non homogène : aujourd'hui dans beaucoup de domaines, on constitue des métadonnées (*MetaData*) dont la structure vise à une standardisation permettant de rapprocher des données différentes, indépendamment de leur source, de leur formats ou de leur structure. Ainsi sont nées les *Linked Data* dont le principe est le suivant : ces données peuvent être de nature et de provenance variées, mais leurs métadonnées permettent de les relier suivant des points de vue différents, afin de produire des bases de données (virtuelles) incorporant des données agrégées et possiblement enrichies. Or, l'évolution de ces nouvelles techniques de traitement des données modifie la notion de données enrichies ou de données brutes : la simple connexion de données qui peut être faite par l'utilisateur peut, en effet, donner lieu à leur enrichissement. On assiste ainsi à une nouvelle suprématie du *Small Data*,² une approche qui consiste à relier de petits ensembles de données plutôt que de concevoir a priori de vastes bases de données. Ces données – souvent de nature différente – ne restent cependant pas isolées. En dépit de leurs différents sources et formats, ces données sont toujours plus connectées les unes avec les autres, associées, assemblées et interconnectées par des relations sémantiques (*Semantic Data*) dans le but de créer des écosystèmes intégrant un très grand nombre des données provenant d'une multitude de jeux de données différents. Ainsi, les subtilités d'antan s'estompent, et les distinctions entre données brutes ou enrichies, données connectées ou bases de données deviennent aujourd'hui obsolètes, ou du moins plus difficiles à cerner. Nous assistons aujourd'hui à l'émergence d'un nouvel écosystème de données, brutes mais interconnectées, élémentaires tout en étant contextualisées par leurs métadonnées, indépendantes sans pour autant être isolées, mais aussi partagées au sein de différentes communautés.

2) Au niveau des sources : *Government Data, Corporate Data, Community Data*

Chaque jour, un nombre croissant de données sont produites ou collectées par différents acteurs opérant dans différents secteurs d'activité et qui se distinguent par leurs objectifs et finalités réciproques.

L'État, tout d'abord, est un important producteur de données. Toujours plus de données sont produites, reproduites, collectées, diffusées ou rediffusées par les administrations publiques dans le cadre de leurs missions institutionnelles. Il s'agit notamment de données démographiques, géographiques, météorologiques, économiques, financières, culturelles, touristiques, etc. qui visent à assurer la qualité et la continuité du service public.

Le secteur marchand contribue lui aussi à la production et à la collecte d'un nombre toujours plus important de données, dans le but de les exploiter pour des finalités commerciales. Aujourd'hui, un grand volume de données est saisi chaque jour par des entreprises ou des opérateurs privés qui opèrent des transactions (*e-commerce*) et fournissent des services en lignes, tels que des *Webmails* (Gmail, Hotmail), des services de messagerie en ligne (Skype, MSN), des réseaux sociaux (Facebook, LinkedIn), des moteurs de recherche (Google, Yahoo), etc. Dans les années à venir, le développement des médias sociaux et de l'Internet des objets (*the Internet of things*) portera sans doute à une croissance exponentielle des données collectées par le secteur privé.

Enfin, les communautés numériques issues de la société civile assument un rôle toujours plus fondamental dans la production et la collecte de données. De nombreuses communautés se concentrent en effet sur le développement de plateformes et d'applications – dites « applications citoyennes » – qui visent à encourager et faciliter la collecte ou l'enrichissement de données de la part des citoyens³. De nature souvent communautaire, ces applications s'appuient sur la contribution de « citoyens capteurs »⁴ : des individus qui collectent ou génèrent des informations géolocalisées dans le but de contribuer à un projet commun (*par ex.* OpenStreetMap.org), améliorer la qualité de vie d'une communauté donnée (*par ex.* Beecitiz.com), ou signaler un problème (*par ex.* SeeClickFix.com).

Dans cet article, nous nous limiterons à étudier l'*Open Data* concernant les données publiques, ce qui ne couvre qu'une partie de la notion de données ouvertes. On s'occupera notamment de données publiques (*Government Data*) telles qu'elles ont été définies par la directive européenne sur la réutilisation des informations du secteur public (*Directive PSI*)⁵ qui se réfère à tous les documents existants détenus par des organismes du secteur public (*i.e.* tous contenus, informations, ou données – y compris les données culturelles et de recherche qui ont été désormais incorporées dans le

1 Le service public de l'information juridique soutenu et maintenu par le Secrétariat général du Gouvernement devait, suivant les accords avec les éditeurs juridiques (négociés au sein de SPIDI), se limiter à fournir des données brutes et non enrichies. L'enrichissement recouvrait les outils d'aide aux ontologies et les thésaurus, mais aussi les liens hypertextes entre articles de loi et jurisprudence qui citait ces articles et inversement. Quant à la doctrine, elle devait rester dans le circuit des données privées (base DOCTRINAL).

2 Le concept de *Small Data* a été popularisé par Rufus Pollock (fondateur de l'Open Knowledge Foundation) dans un *blog post* décrivant les avantages du *Small Data* par rapport aux *Big Data*, disponible à l'adresse suivante : <http://blog.okfn.org/2013/04/22/forget-big-data-small-data-is-the-real-revolution/>.

3 Pour plus de détails, voir Danièle Bourcier, Primavera De Filippi (2012), *Vers un nouveau modèle de partage entre l'administration et les communautés numériques*, Nicolas Matyjasik, Philippe Mazuel (eds.) *Génération Y et gestion publique : quels enjeux ?*, Institut de la gestion publique et du développement économique (IGPDE), Octobre 2011.

4 Le concept de « citoyen capteur » a été présenté par Boris Mericskay, chercheur à l'université canadienne de Laval, lors du 80^e congrès de l'ACFAS – Colloque 448 « Science et démocratie » avec une présentation sur « Les citoyens capteurs au service des sciences citoyennes ».

5 V. la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 sur la réutilisation des informations du secteur public, ainsi que la directive 2013/37/UE du Parlement européen et du Conseil du 26 juin 2013 modifiant la directive 2003/98/CE.

régime général de réutilisation, même si leur réutilisation peut être parfois sujette à des exceptions), à l'exception des documents dont la fourniture est une activité qui ne relève pas de la mission de service public (article 2(a)), des documents dont des tiers détiennent les droits de propriété intellectuelle (article 2(b)) et des documents qui, en vertu du droit en vigueur, ne sont pas accessibles pour des raisons de protection de la vie privée, de confidentialité ou de sécurité nationale (article 2(c)).

Ainsi, nous analyserons non seulement le cadre réglementaire régissant les données au sens strict, mais aussi celui régissant tout document, information ou contenu qui a été produit et/ou qui est actuellement détenu par les administrations publiques⁶.

B – Caractéristiques des données publiques ouvertes (*Open Government Data*)

D'après Wikipédia, une donnée ouverte est « *une donnée numérique, d'origine publique ou privée, publiée de manière structurée selon une méthodologie qui garantit son libre accès et sa réutilisation par tous, sans restriction technique, juridique ou financière* ». On pourrait résumer en disant que l'*Open Data* se réfère au monde des données déjà constituées et accessibles et qui deviennent disponibles sous un certain régime juridique et sous un certain format technique qui en permettent ou en facilitent la réutilisation.

Une définition plus précise nous est fournie par l'*Open Definition* de l'Open Knowledge Foundation (<http://opendefinition.org>), qui considère qu'une donnée n'est ouverte que s'il est possible de l'utiliser, de la réutiliser et de la redistribuer librement – avec, comme seules conditions admissibles, d'une part, l'obligation d'en mentionner la source et, d'autre part, la nécessité de permettre la réutilisation de toute base de donnée dérivée sous les mêmes conditions que la base de donnée originale (*share-alike*).

Une caractéristique centrale réunit donc le mouvement de l'*Open Data* : une donnée ouverte implique sa libre réutilisation.

Or, cette définition ne rend qu'insuffisamment compte du mouvement et des politiques dont on parle dans le secteur public, ou dans le secteur subventionné par des fonds publics. Surtout on s'aperçoit que, aussi bien dans la loi que dans la pratique, une donnée publique dite ouverte ne correspond pas toujours à l'intégralité des caractéristiques relevées dans cette définition. Tout d'abord, dans de nombreux pays – y compris en France – la loi garantit un droit aux administrations publiques de maintenir la paternité de leurs données, d'en préserver l'intégrité, et, dans certains cas, de soumettre la réutilisation de ces données au paiement d'une redevance. Ensuite, à l'exception des administrations publiques centrales de l'État, de nombreuses institutions publiques ne sont pas soumises à l'obligation de mettre à disposition leurs données de manière libre et gratuite et sont donc libres d'imposer un certain nombre de conditions ou de restrictions sur les modalités de réutilisation (c'est le cas, par exemple, en France ainsi que dans de nombreux autres pays d'Europe, des collectivités territoriales, des SPIC, des EPIC, des institutions culturelles et de recherche). Ainsi, une donnée du secteur public au sens large – sauf si elle est dans le domaine public – est rarement totalement ouverte (*open*) et intégralement gratuite (*free*).

Dans le secteur public, l'*Open Data* pourrait donc être défini comme la mise à disposition *libre* et (quasi) *gratuite* de données qui implique la possibilité de les réutiliser de façon la moins contraignante possible. Quasi-gratuite signifie par exemple qu'un certain coût de restitution peut être facturé – par ex. pour couvrir les coûts de gestion ou de mise à disposition des données – bien que cette redevance ne corresponde en aucune façon au coût réel de production (comme le Journal officiel version papier est quasi-gratuit). L'idée sous-jacente du principe de gratuité est que, puisque la production des données issues des administrations publiques a été financée, entièrement ou en partie, par l'argent des contributeurs ou contribuables (i.e. les taxes), il serait injuste de demander aux citoyens de contribuer à nouveau à couvrir les coûts de production. Il est cependant acceptable, dans certains cas, de soumettre la réutilisation de ces données à une redevance, qui ne doit toutefois pas dépasser le strict minimum afin de rembourser les coûts supplémentaires qui ont été pris en charge par l'administration afin de rendre ces données accessibles au public.

Ainsi, on peut dire que l'*Open Data* dans le secteur public est, en fin de compte, une question de degré. Les données sont neutres, mais les politiques gouvernementales ne le sont pas. L'*Open Data* n'est donc pas un domaine entièrement technique, c'est un domaine ancré dans des cultures différentes. Différents pays ou régions, ainsi que différentes institutions sont susceptibles d'adopter des démarches *Open Data* différentes, avec différents degrés d'ouverture selon leurs spécificités et besoins respectifs. On s'intéressera ici à la mise en pratique de ces expériences.

2. Open Data, universalité du principe

A – Portée géopolitique de l'*Open Data*

Au cours des dernières années, nous avons assisté à une forte croissance des initiatives *Open Data* à travers le monde, notamment en ce qui concerne l'ouverture des données publiques.

Suite aux initiatives des États-Unis (data.gov) et du Royaume-Uni (data.gov.uk), de nombreuses initiatives ont été lancées à échelle locale, régionale et nationale ayant pour objectif l'ouverture des données publiques et le développement de portails en ligne pour la mise à disposition de ces données. Appliqués pour la première fois en 2009 (date de lancement du premier portail data.gov), les principes et les valeurs de l'*Open Data* se sont rapidement répandus à travers le monde⁷.

⁶ Cela signifie que, pour nous, les bases de données juridiques (par ex. Légifrance) peuvent être considérées comme entrant dans le jeu de l'*Open Data*.

⁷ Une liste de tous les portails *Open Data* qui se sont développés dans le monde est disponible sur le site : www.data.gov/opendatasites.

Aujourd'hui, l'*Open Data* est devenu un phénomène universel : on compte désormais des centaines d'initiatives locales ou régionales, avec plus de 40 pays qui ont entrepris une démarche d'*Open Data*⁸. En Europe, suite à la directive européenne sur la réutilisation des informations du secteur public⁹ la majorité des pays ont adopté une politique visant à promouvoir l'ouverture des données publiques. Au sein de chaque pays, de nombreuses villes, régions et collectivités territoriales ont décidé de se lancer – sans y être obligées – dans une démarche d'*Open Data*. Mais l'Europe n'est pas l'exception. Au cours des années, des initiatives *Open Data* ont surgi sur tous les continents : en Amérique, au-delà des États-Unis et du Canada, on retrouve le Mexique, le Brésil, l'Argentine, le Chili, l'Uruguay et le Pérou. En Asie, on y retrouve – entre autres – la Russie, la Chine, la Corée du Sud, la Malaisie, l'Indonésie et l'Inde qui a récemment lancé son portail *Open Data* développé en partenariat avec le Gouvernement américain. L'Océanie n'en est pas exclue, avec des initiatives *Open Data* provenant aussi bien de l'Australie que de la Nouvelle-Zélande. Enfin, bien que moins répandues dans le continent africain, on y retrouve toutefois des initiatives *Open Data* dans les Émirats Arabes, en Tunisie, au Maroc, au Kenya, et même dans le Royaume de Bahreïn.

B – Portée socio-économique de l'Open Data

La notion de « gouvernement ouvert » porte en elle une évolution radicale par rapport à la mise à disposition des données publiques.

D'abord, ce ne sont plus seulement les données officielles (données juridiques, données administratives) qui sont concernées mais toutes les données produites par les personnes publiques.

Ensuite, l'objectif de l'accès à ces données concerne non seulement le principe traditionnel de transparence dans les rapports administration-citoyens mais aussi celui de réutilisation généralisée de ces données, au-delà du service public traditionnel. Elle rejoint des initiatives que l'on peut regrouper sous le terme d'accès à l'information publique et qui sont fondées sur un principe nouveau, lié à la valeur économique des données : c'est ainsi que les citoyens, qui ont « contribué » par l'impôt à produire ces données, se donnent le « droit » (*empowerment*) de les valoriser pour leur propre intérêt. Ce droit, à l'heure du web 2.0 et des téléphones mobiles, ne concerne plus les grands opérateurs qui, dans le schéma du service public des grandes banques de données juridiques¹⁰, représentaient les diffuseurs concessionnaires prenant en charge des investissements lourds. Les données à traiter sont aujourd'hui plus légères et leurs usages sont liés à des applications variées voire éphémères.

Un autre aspect concernant le rôle de l'État doit être souligné dans le cadre de cette évolution. Les principes de défense du service public de l'information qui mettaient l'État au centre du dispositif visaient à garantir la neutralité des sélections (certaines banques de données n'étaient intéressées que par des fonds rentables, comme les arrêts des Cours suprêmes) et l'égalité des institutions et des citoyens à l'accès aux données. Actuellement, une grande partie de ces garanties ont été acquises mais aussi, et surtout, l'État n'a plus les moyens de s'ingérer dans la valorisation de ces nouveaux dépôts de données. Pour cela, le gouvernement ouvert favorise l'externalisation du processus d'intégration et de qualification des données. Cette attitude pourrait susciter une concurrence entre des acteurs privés afin d'obtenir des données plus fiables. En d'autres termes, le gouvernement ouvert vise à une mise à disposition des données, mais cela ne peut se faire que par une standardisation de leur mise en forme voire une prise en charge du traitement des données avant leur diffusion.

Enfin, une des dernières étapes que l'on peut observer, quels que soient les pays, c'est la naissance de nouveaux acteurs autres que l'utilisateur ou le citoyen final. Les premiers ré-utilisateurs concernés proviennent en général du secteur marchand, et leurs initiatives ont été à certains moments vues de façon restrictive comme pouvant priver l'administration du revenu sur son « patrimoine » ou d'un « petit équilibre » financier. Mais il existe un autre type d'utilisateur, qui, communautaire au sens large, poursuit une finalité collaborative plutôt que commerciale¹¹. Une des principales raisons du silence sur ces dernières initiatives vient du fait que les missions de l'administration sont représentées plutôt dans un face à face vis-à-vis de l'administré, le secteur commercial se situant en dehors du champ du service public (à travers la concession, la délégation ou d'autres dispositifs de partenariats). Ce sont donc de nouvelles relations administration-administrés, étendues à d'autres acteurs collectifs dont il est question.

Dans le contexte actuel de restriction des missions de l'administration (crise politique, déficit budgétaire, réduction de postes, etc.), il est devenu fondamental de passer du modèle administratif du face à face avec le public à un modèle d'interactions du secteur public avec des groupes issus de la société civile ou de communautés de citoyens. On peut ainsi désenclaver la réflexion sur l'intérêt pour le secteur public d'ouvrir ses données : il s'agit pour lui de répondre à des besoins qui peuvent ne pas être assurés par l'administration ou par le secteur privé mais qui pourraient l'être avantageusement par une communauté d'internautes. L'utilisateur n'est plus seulement intéressé par l'accès à l'information ou par l'usage des données brutes. L'enjeu principal est de rendre ces données utiles et utilisables pour mettre en commun des ressources et compléter un service qui serait insuffisamment assuré par l'administration. Ainsi, l'ouverture des données devient-elle un véritable enjeu politique pour certains groupes ou communautés¹².

Pour conclure, d'une façon générale, le mouvement d'accès, d'ouverture, de partage des données et de réutilisation est un mouvement irrésistible qui concerne tous les domaines. En science sous la pression des chercheurs, le mouvement

8 Le site <http://datacatalogs.org> maintient une liste de tous les jeux de données qui ont été mises librement à disposition du public par des initiatives locales, régionales et nationales.

9 Directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 sur la réutilisation des informations du secteur public.

10 Voir, suite au rapport *Leclercq* (1984) non publié, le décret n°84-940 24 octobre 1984 (*JO 25 oct. 1984, p. 3336*) qui introduit une sorte de droit – non d'accès – mais de « consultation notamment par télématique » des banques de données juridiques.

11 *Primavera De Filippi, Danièle Bourcier, Vers un nouveau modèle de partage entre l'administration et les communautés numériques, préc.*

12 Ainsi *Liberté pour les données de santé* <http://www.opendata.com/qui-sommes-nous/> [???Pour auteurs : le lien semble invalide???

du *Data Sharing* se renforce avec les déclarations et recommandations successives qui dépassent le cadre des États¹³ mais aussi grâce à des plateformes comme Sciences Commons¹⁴ visant à encourager la diffusion de la recherche scientifique. Même dans le secteur du droit d'auteur, pourtant soucieux des droits patrimoniaux, maints mouvements se sont développés autour du logiciel libre, des contenus ouverts et de la culture libre (*free culture*)¹⁵.

3. Open Data : diversité des expériences

A – État des lieux

L'*Open Data* est peut-être en train de devenir un principe universel, il reste cependant assez hétéroclite quant à son implémentation dans différents pays. Une vision globale des politiques d'*Open Data* dans le monde nous est fournie par l'*Open Data* census de l'Open Knowledge Foundation (census.okfn.org), une application en ligne qui permet d'identifier, de comparer ou d'évaluer différents jeux de données selon leur degré d'ouverture respectifs.

Il est, en effet, important de comprendre que la démarche *Open Data* d'un pays ne dépend pas (uniquement) de la quantité de données qui ont été mises librement à la disposition du public, mais aussi (et surtout) de la qualité de ces données.

Il s'agit, tout d'abord, d'identifier les jeux de données qui sont les plus urgents à ouvrir dans la mesure où ils contiennent des informations critiques pour les citoyens (tels que les finances publiques, ou les données relatives aux transports publics). On remarquera ainsi que certains pays, tels que les États-Unis, le Royaume-Uni, l'Australie, le Japon, le Danemark et la Norvège sont beaucoup plus avancés en matière d'ouverture que d'autres pays, tels que l'Égypte, le Kenya, l'Angola, le Mozambique ou la Finlande, qui – bien qu'ayant adopté une politique d'*Open Data* – restent cependant plus limités dans leur démarche d'ouverture, notamment en ce qui concerne le budget et les dépenses de l'État, les données environnementales et les données de transport qui ne sont pas toujours librement accessibles, ou qui ne sont réutilisables que de manière limitée.

Le format des données est tout aussi important, dans la mesure où il va influencer les opportunités d'extraction et de réutilisation des données. Des données publiées dans un format ouvert et interopérable (par ex. RDF) seront ainsi plus susceptibles d'être réutilisées que des données publiées en un format non structuré (par ex. PDF).

B – Diversité au niveau international : comparaison des approches en termes de gouvernement ouvert

Pourquoi les politiques de l'*Open Data* se sont-elles développées d'abord aux États-Unis, en Australie et en Grande-Bretagne ? Certes il s'agit de promouvoir la transparence, mais surtout de développer un secteur où tout le monde serait à égalité pour créer de la valeur ajoutée à un projet fondé sur ces données ouvertes et, pourquoi pas, en tirer un profit.

Aux États-Unis et dans les autres pays anglo-saxons, on observe une logique d'ouverture et de qualité des données régulée par le marché plus que par l'État. En effet, la mauvaise qualité des données aurait des effets négatifs sur la commercialisation d'un service fourni aussi bien par le secteur public que par le secteur privé. Une donnée erronée compromettrait la fiabilité du produit et nuirait ainsi à la confiance du « client » vis-à-vis du traitement global des données. Le réutilisateur n'a donc pas « intérêt » à diffuser un service défectueux. La qualité se régule « par le bas ».

La notion d'écosystème des données publiques en tant que source de profits est toutefois appréhendée d'une manière différente aux États-Unis et en Grande Bretagne. Dans le premier cas les données publiques diffusées gratuitement sont sources d'innovation, dans le second elles sont sources d'expertise qui deviendront ultérieurement sources de profits. Dans les deux cas, la notion de revenus patrimoniaux prônés par l'APIE est absente.

En France, la préoccupation budgétaire est centrale dans l'évolution de la conception de gouvernance numérique. À l'inverse du monde anglo-saxon en général, et britannique en particulier, l'État (centralisé) a conservé une primauté en matière de développement des réseaux numériques, de diffusion des informations ou de développement de l'administration numérique. C'est à l'aune des contraintes budgétaires ou tout au moins de politiques industrielles qu'est repensée la place de l'État dans l'organisation du paysage numérique. L'espace numérique français a ainsi été appréhendé en termes patrimoniaux. Le rapport *Jouyet/Levy* de 2006 et la constitution de l'APIE ont mis en avant la notion de valorisation des données publiques comprises comme « compléments de ressources » pour les administrations.¹⁶ Parallèlement, la notion de territoire numérique largement issue de la notion d'écosystème telle qu'elle a été développée aux États-Unis suscite de nouvelles expériences aussi bien au niveau local (comme l'espace numérique de la ville de Rennes) qu'au niveau national (comme le portail Etalab).

Autre différence fondamentale entre l'approche anglo-saxonne et l'approche continentale : elles reposent sur deux conceptions du droit moral, lié au droit d'auteur. Ces conceptions divergentes deviennent visibles dans la mise en œuvre de l'*Open Data*.

En France, la visée de l'*Open Data* est orientée vers une protection « par le haut » de la qualité des données. L'État veut s'assurer et se porter garant de l'authenticité et de la qualité des données dont il est « auteur ». Il aurait, par assimilation

13 PLOS Lettre ouverte (2001) ; Déclaration de Budapest (2002) ; Charte ECHO (2002) ; Déclaration de Berlin (2003).

14 Science Commons est un projet de Creative Commons créé pour définir des stratégies et des outils pour une diffusion de la recherche scientifique plus rapide et efficace telle qu'internet le permet. Plus d'information sur le site <http://creativecommons.org/science>.

15 L'exemple le plus notable sont les licences *Creative Commons*, des licences conçues pour libérer les contenus de certains droits exclusifs, à la demande des auteurs. Plus d'information sur le site <http://creativecommons.org>.

16 Le rapport *Jouyet/Levy*, consacré à l'économie de l'immatériel, suggérait la création d'une agence pour la valorisation du patrimoine immatériel des institutions publiques en France (tels que les brevets, les marques, les licences, mais aussi les savoir-faire et les bases de données). Cette idée a conduit à la création en 2007 de l'Agence du patrimoine immatériel de l'État (APIE) qui a pour rôle d'optimiser l'impact de la gestion du patrimoine immatériel français. C'est ainsi que l'APIE intervient dans l'élaboration d'un cadre réglementaire, et met à disposition des administrations publiques une série d'outils destinés à favoriser une gestion plus dynamique et efficace du patrimoine immatériel de l'État.

à l'auteur d'un écrit ou d'une œuvre, une sorte de droit moral ou éthique sur ses données – de même que le droit moral dans le régime du droit d'auteur protège le nom de l'auteur (l'attribution) et l'intégrité du contenu. Ce sont bien ces deux caractéristiques qui restent présentes dans les conditions d'utilisation des licences sur les données publiques. N'y aurait-il pas là un héritage (implicite) du droit de la propriété intellectuelle dans les conditions fixées par les licences françaises? Bien qu'il existe une loi qui affirme haut et fort un principe d'ouverture (contrairement aux pays anglo-saxons), cette loi introduit en même temps des éléments de contrôle de qualité, mais aussi de patrimonialisation, qui rapprocheraient le régime des données publiques d'une conception issue de la propriété intellectuelle fondée sur la valorisation du patrimoine. Ainsi, à côté d'un souci légitime de protection de l'information publique, le régime des données publiques en France a introduit un modèle économique dans un modèle administratif.

1) L'expérience anglo-saxonne de l' « Open-Government »

Les premières expériences d'*Open-Government* ont été lancées dans les pays comme les États-Unis¹⁷, la Grande Bretagne¹⁸ et l'Australie¹⁹.

Avec le site *data.gov*, le Gouvernement américain a lancé le début d'un mouvement général d'ouverture des données publiques à travers le monde. Le but de ce site est de faciliter l'accès aux informations et aux données publiques en publiant toutes les ressources générées par le gouvernement fédéral des États-Unis – dans la mesure où elles ne sont pas protégées par la loi et que leur accès n'est pas limité pour des raisons de sécurité nationale. Encouragés par l'initiative nationale du président Obama en matière de gouvernement ouvert²⁰, la transparence du gouvernement fédéral et l'ouverture des données publiques sont aujourd'hui perçues comme des étapes nécessaires pour stimuler l'innovation aux États-Unis et ailleurs dans le monde. Ainsi, avec seulement 47 jeux de données en 2009, *data.gov* accueille en 2013 plus de 450 000 jeux de données provenant de 172 agences fédérales.

En Grande-Bretagne, la notion de gouvernement ouvert est essentiellement liée à la libération des données publiques, promue à travers le portail *data.gov.uk*. La gouvernance et le contrôle de ce portail ne sont pas centralisés au niveau de l'État. Introduisant un nouveau partage de responsabilité entre le domaine public et le domaine privé, la gestion du site est confiée au cabinet du Premier ministre et les archives nationales d'une part, et des sociétés sous-traitantes liées au web sémantique d'autre part, avec, au milieu, l'Open Knowledge Foundation²¹, organisation agissant comme conseiller et agence de promotion et de normalisation.

En Australie, le Gouvernement s'est engagé à construire le site *data.gov.au* servant de point d'accès central pour les données du Gouvernement. Dans un fameux discours, la sénatrice australienne Kate Lundy a rappelé que les trois piliers fondamentaux d'un gouvernement ouvert sont : les services axés sur le citoyen, la démocratisation des données et le gouvernement participatif²². Elle souligne que « les données du Gouvernement sont une ressource publique » qui est donc destinée à la diffusion publique, à moins qu'il y ait une raison spécifique de ne pas le faire. En Australie, l'ouverture des données a été fortement facilitée par les licences Creative Commons²³. Le Gouvernement australien recommande en effet que toute information produite ou reçue par le secteur public soit publiée par défaut sous une des licences Creative Commons²⁴. Aujourd'hui, un grand nombre de ressources culturelles et de documents administratifs – y compris le budget fédéral – ont été publiés sous une licence des plus permissives qui ne demande que l'attribution de la source (CC-BY).

2) L'expérience française en matière de « gouvernement ouvert »

En France, l'approche du « gouvernement ouvert » menée au sein des ministères a été tout d'abord de nature juridique. Contrairement à ce qui se passe en matière de droits de propriété intellectuelle – où l'interdiction est la règle et la réutilisation n'est permise que dans certains cas exceptionnels – la loi a introduit un régime général favorisant l'accès et la libre réutilisation des données publiques, où l'interdiction n'est possible que par dérogation à la règle générale. Ce régime s'est construit en plusieurs étapes, à travers une longue succession de règles de droit (y compris les licences) plutôt qu'à travers l'observation de la pratique et des usages. Parallèlement à cette évolution du droit au niveau national, les collectivités locales ont défini leur stratégie de façon plus collaborative et avec une plus forte proximité avec les citoyens.

17 *Data.gov* est un site web lancé en mai 2009 pour mettre à disposition du public les données ou informations publiques produites ou issues du gouvernement des États-Unis.

18 *Data.gov.uk* est un site web lancé en septembre 2009 pour mettre à disposition du public les données ou informations publiques produites ou issues du gouvernement du Royaume-Uni.

19 *Data.gov.au* est un site web lancé en Mars 2011 pour mettre à disposition du public les données ou informations publiques produites ou issues des administrations publiques australiennes.

20 Pour plus de détails sur l'Open Government U.S. National Action Plan du président Obama : « The Open Government Partnership », National Action Plan for the United States of America, septembre 2011.

21 L'Open Knowledge Foundation est une association à but non lucratif fondée en 2004 à Cambridge au Royaume-Uni. Son but est de promouvoir la culture libre dans des domaines aussi variés que l'éducation, la science, ou tout autre secteur d'activités produisant des données ouvertes. Plus d'information sur le site <http://okfn.org>. L'Open Knowledge Foundation a développé le logiciel CKAN, actuellement utilisé par de nombreux portails, dont *data.gov.uk* et, plus récemment, *data.gov* aux États-Unis. Plus d'information sur le site <http://ckan.org>.

22 Discours de la sénatrice Kate Lundy, secrétaire parlementaire du Premier ministre et secrétaire parlementaire de l'immigration et la citoyenneté, délivré le 1er mars 2011 lors du Citizen Centric Service Delivery à Canberra sur le thème de : « Services axés sur les citoyens : un principe nécessaire pour parvenir à une véritable transparence du gouvernement ».

23 Creative Commons est une organisation à but non lucratif qui a pour dessein de faciliter la diffusion et le partage des œuvres tout en accompagnant les nouvelles pratiques de création à l'ère numérique. Plus d'informations sur www.creativecommons.org.

24 Suite aux recommandations du Gouvernement 2.0 Taskforce (disponible sur <http://gov2.net.au/report/>), le Gouvernement australien s'est engagé à promouvoir un gouvernement ouvert en mettant en disposition les informations du secteur public sous une des licences Creative Commons.

Le cadre juridique et réglementaire de l'*Open Data* en France est régi par la loi n° 78-753 du 17 février 1978 conférant à tout citoyen un droit d'accès aux documents administratifs. Cette loi a été modifiée par l'ordonnance n° 2005-650 du 6 juin 2005, transposant en droit français la directive européenne 2003/98/CE sur la réutilisation des informations du secteur public en introduisant un droit de réutilisation des données publiques.

Aujourd'hui, la loi reconnaît donc à tout citoyen un droit d'accès et de réutilisation des données publiques, définies comme toute information produite ou reçue « dans le cadre de leur mission de service public, par l'État, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission » (article 1).

Ce principe – on l'a souligné précédemment – est en soi une innovation juridique puisqu'aucun autre pays ne l'avait affirmé de façon aussi claire dans la hiérarchie des normes. Mais il est important de rappeler l'écart entre la liberté abstraite conférée à chaque individu par la loi, et la liberté concrète qui implique la possibilité pour chaque individu d'exercer effectivement cette liberté. La liberté abstraite des individus ne peut se transformer en une liberté concrète que si la loi prévoit un mécanisme capable d'assurer que cette liberté – ou ce droit – puisse être effectivement appliqué partout et par tous. Ainsi, le droit à l'accès et à la réutilisation des données publiques peut être demandé soit directement par le citoyen, soit indirectement – en cas de refus – par l'intermédiaire de la Commission d'accès aux documents administratifs (CADA).

Dans sa forme initiale, la loi autorisait les autorités administratives à subordonner la réutilisation des données publiques au paiement d'une redevance, tant que les coûts n'étaient pas excessifs par rapport à ceux de la collecte, de la production et de la mise à disposition du public. La situation a changé avec la circulaire du 26 mai 2011, imposant aux administrations centrales d'autoriser la réutilisation gratuite de leurs données à partir du 1er juillet 2012. La fixation de redevances n'est possible que si elle est « dûment justifiée par des circonstances particulières » et doit se faire par le biais d'un décret, suite à l'approbation du Conseil d'orientation de l'édition publique et de l'information.

C'est ainsi qu'on voit apparaître une autre finalité qui n'est plus seulement juridique ou démocratique mais aussi économique : « les données publiques, financées par l'impôt, doivent pouvoir être réutilisées par d'autres acteurs, au service de la qualité des services sur le territoire, de la croissance économique, de la connaissance et du débat démocratique » (article 1). Ces dispositions ne s'appliquent pas cependant aux données relevant de la sécurité nationale (article 6), aux documents produits ou reçus par un service public à caractère industriel ou commercial et aux informations sur lesquelles des tiers détiennent des droits de propriété intellectuelle (article 10).

En outre, la loi prévoit que, sauf accord de l'administration, la réutilisation des données publiques ne peut se faire qu'à condition qu'elles ne soient pas altérées, que leur sens ne soit pas dénaturé et que leurs sources et la date de leur dernière mise à jour soient mentionnées (article 12).

La dénaturation des données est un concept qui mérite d'être analysé de plus près. Etymologiquement, le procédé de dénaturation consiste à modifier la nature ou les qualités naturelles des choses, indépendamment de la modification de leur fonction ou de leurs qualités non intrinsèques. Dans le cas des données, cela signifie que les données peuvent être changées, par exemple en modifiant leur format ou en les enrichissant avec plus de détails, sans pour autant provoquer une altération de leur sens. L'interdiction de dénaturer les données n'est donc pas directement reliée avec la possibilité ou non de réutiliser les données pour la réalisation d'œuvres ou de bases de données dérivées, mais se réfère uniquement à la prohibition de modifier le sens et la signification de ces données quel qu'en soit le support. La question reste de savoir si une certaine réutilisation des données peut être considérée comme une dénaturation des données au sens de la loi.

C – Diversité au niveau national : le cas de la France

Fondée sur une utilisation étendue du web 2.0, la notion de territoire numérique²⁵ propose un nouveau modèle de production, d'utilisation, de partage et de diffusion des données. Là aussi émerge un modèle de gouvernance où s'estompent les frontières, les responsabilités et les compétences entre l'État, les collectivités locales et la société civile. Si l'on analyse comment ce mouvement est parti - de la ville de Brest, à travers les politiques de diffusion et de partage de données publiques sous licences Creative Commons de la région Aquitaine - les territoires semblent avoir été les premiers à s'engager et à vouloir explorer les avantages d'une politique d'ouverture des données publiques. Depuis, de nombreuses collectivités territoriales se sont dirigées vers l'*Open Data*, telles que les villes de Paris, Rennes, Arles, Montpellier, Toulouse et Bordeaux, ainsi que les départements de la Gironde, Saône-et-Loire, les régions de Nord-Pas-de-Calais et des Pays de la Loire. Loin de vouloir obtenir des gains de nature financière, la plupart de ces initiatives visent à encourager le développement de nouveaux services basés sur la réutilisation et l'intégration des données publiques afin de proposer des informations plus détaillées et des services plus spécialisés aux citoyens.

Ainsi, défendant l'idée simple qu'il faut ouvrir les données pour mieux les partager, les données publiques de la ville de Paris sont aujourd'hui à la disposition de tous sur opendata.paris.fr, celles de la ville de Toulouse sont accessibles sur data.grandtoulouse.fr, alors que les données publiques de la ville de Montpellier peuvent être consultées librement sur opendata.montpelliernumerique.fr. La ville de Brest représente peut-être l'un des exemples le plus clair, avec une politique visant à promouvoir l'accès et la réutilisation des données publiques par le développement d'un grand nombre de projets destinés à simplifier ou améliorer la vie des citoyens²⁶. La ville de Rennes n'est pas de reste avec plus de 150 jeux de données (informations sur les réseaux de transports en commun, sur l'emplacement des équipements urbains, ou, plus récemment, sur les données budgétaires municipales) mises à disposition librement sur son portail²⁷ dans l'espoir que les citoyens s'en emparent. De même, les départements de la Gironde et d'Aquitaine se sont orientés vers une meilleure exploitation et valorisation des données publiques par la mise en place d'une plateforme de publication mutualisée qui cherche à maximiser l'accessibilité et la réutilisation des données²⁸. Ces initiatives ne sont que des

²⁵ Ce terme vient de l'expérience suédoise, pionnière en la matière.

²⁶ Plus d'informations sur <http://www.wiki-brest.net/http://www.wiki-brest.net/>.

²⁷ Voir <http://data.rennes-metropole.fr>

²⁸ Voir <http://www.datalocale.fr/>

exemples figurant parmi les nombreuses initiatives *Open Data* qui se sont développées jusqu'à présent sur le territoire français²⁹.

1) L'hétérogénéité des licences

L'ouverture des données ne peut se faire qu'avec une licence qui spécifie les droits et les devoirs associés à la consultation et à la réutilisation des données. Cependant, jusqu'au lancement de la licence Etalab en octobre 2011, il n'y avait, en France, aucune licence de référence pour la mise à disposition des données.

C'est pour cela que le ministère de la Justice a créé la licence information publique librement réutilisable (LIP)³⁰, une licence permettant la libre diffusion et réutilisation des données, sous réserve d'en mentionner la source et d'en respecter l'intégrité, avec comme condition ultérieure que toute diffusion commerciale des données ne peut se faire que si celles-ci font l'objet de « nouveaux traitements ». Cela signifie que les données du ministère de la Justice ne peuvent être librement réutilisées que dans un contexte différent et pour des finalités différentes de celles pour lesquelles elles ont été collectées.

En ce qui concerne les administrations locales, le manque de consensus sur une licence unique de réutilisation des données publiques en France a obligé les collectivités territoriales à effectuer une décision de caractère individuel en matière d'ouverture des données afin d'adopter la licence la plus adaptée au contexte politique local. Puisque le gouvernement ne leur avait pas proposé une licence standard de référence, ni même ouvert largement la notion de domaine public (ce qui aurait pu être une solution), les collectivités territoriales et les administrations publiques françaises ont dû déterminer par elles-mêmes le plan d'action le plus approprié au contexte politique local, indépendamment du fait que les mêmes réflexions avaient déjà été menées par ailleurs. C'est ainsi que les départements de la Gironde, de l'Aquitaine, et du Nord-Pas-de-Calais, ainsi que les villes de Nantes, Toulouse, Angers, La Rochelle, et – plus récemment – la commune de Brocas ont choisi la licence ODbL (*Open Database Licence* de l'Open Knowledge Foundation)³¹ pour la mise à disposition de leur données, alors que la ville de Paris a opté pour un modèle mixte d'ouverture des données (ODbL Paris)³² qui concilierait mieux citoyenneté et patrimonialité. Le département de Saône-et-Loire, ainsi que les villes d'Arles, de Montpellier et de Bordeaux ont initialement préféré utiliser les licences de l'APIE qui ne permettent que la réutilisation non commerciale des données, tandis que la ville de Rennes, elle, a opté pour la création de sa propre licence « Rennes Métropole en accès libre », une adaptation de la licence APIE qui répondrait mieux aux attentes de la communauté du libre. Les licences de l'APIE ont, cependant, été rapidement abandonnées, aussi bien par le département de Saône-et-Loire qui – après avoir expérimenté avec la licence APIE V2, considérée comme « une licence d'état rassurante » - a finalement adopté la licence LIP, que par la ville de Bordeaux, qui a finalement adopté une des licences proposées par Creative Commons (CC-BY-NC-SA France 2.0)³³ pour la mise à disposition de ses données.

Bien que les initiatives prises par ces collectivités territoriales soient la preuve d'un intérêt croissant pour l'ouverture et la réutilisation des données locales en France, l'utilisation – ou la création – de licences différentes risque de nuire à l'interopérabilité de ces données et de limiter ainsi l'efficacité de ces initiatives. La multiplication des licences pour la mise à disposition de données publiques constitue en effet un frein à leur réutilisation ce qui risque de compromettre, par exemple, la création de plateformes 2.0 intercommunales ou interterritoriales. En effet, à l'exception de la licence ODbL, ces licences ne peuvent pas être considérées comme des « licences libres » (telles qu'elles ont été définies par l'Open Knowledge Foundation)³⁴ car elles imposent des restrictions plus ou moins importantes sur la réutilisation des données.

L'arrivée d'une licence standard de référence pour tout le territoire français se présente donc comme une opportunité pour réduire cette autre complexité juridique qui limite le potentiel de valorisation des données publiques.

2) Une licence de référence pour la France : La licence ouverte

Au regard des spécificités de la loi française en matière d'ouverture des données, les administrations publiques se méfient souvent des licences internationales qui n'ont pas été adaptées à la situation française. Ainsi, bien qu'il existe aujourd'hui plusieurs licences reconnues au niveau international – telles que les licences Creative Commons ou la licence ODbL – ces licences sont toutefois incapables de rassurer les administrations qui diffusent leurs données sur le territoire français. Il était donc extrêmement important de créer une licence personnalisée qui se rapporte expressément à toutes les obligations réglementaires du droit français.

C'est exactement cette lacune que l'Agence pour le patrimoine immatériel de l'État (APIE) a essayé de combler en élaborant des contrats-type conçus pour répondre aux besoins des administrations publiques et des opérateurs institutionnels. Ces licences reprennent les termes de la loi du 17 juillet 1978 en fixant les conditions de réutilisation des informations publiques, avec ou sans redevance. L'APIE a récemment publié la version V2 de sa licence de réutilisation libre et gratuite, qui a été renommée licence informations publiques³⁵. En dépit de leurs noms, aucune de ces licences n'est pourtant compatible avec les critères établis par l'Open Definition de l'Open Knowledge Foundation.

Ce manque d'homogénéité est dû notamment à la différence qui existe entre les licences publiques (qui concernent le public au sens large) et les licences individuelles (qui concernent des individus identifiés). Les licences de l'APIE qui ne confèrent qu'un « droit personnel » au licencié établissent donc une relation contractuelle de nature individuelle. Cette caractéristique rend ces licences potentiellement incompatibles avec les licences libres existantes qui mettent sur un pied d'égalité toutes les personnes qui ont téléchargé les données et rendent les données indépendantes de leurs sources.

29 Une liste de bases de données publiques librement accessibles peut être consultée à l'adresse suivante :

http://publicdata.eu/package?extras_eu_country=FR.

30 http://www.rip.justice.fr/information_publicque_librement_reutilisable.

31 <http://opendatacommons.org/licenses/odbl/>.

32 http://opendata.paris.fr/opendata/jsp/site/Portal.jsp?page_id=10.

33 <http://creativecommons.org/licenses/by-nc-sa/2.0/fr/>.

34 <http://opendefinition.org/okd/francais/>.

35 Voir <https://www.apiefrance.fr/>.

C'est dans ce contexte que le 18 octobre 2011, le comité Etalab a publié la « licence ouverte/Open Licence »³⁶ précédant de quelques semaines l'ouverture de la plateforme nationale des données publiques *data.gouv.fr*. Prévues par la circulaire du 26 mai 2011, l'arrivée de cette licence constitue une avancée importante dans le mouvement *Open Data* en France. L'objectif est celui de faciliter la consultation et d'encourager la réutilisation des données publiques, en fournissant une licence de référence qui peut être facilement adoptée par toute collectivité territoriale ou organisme public qui souhaiterait ouvrir ses données au public.

En dépit des critiques que l'on pourrait lui adresser, cette licence est finalement plus ouverte que la plupart des licences qui ont été utilisées jusqu'à présent. En effet, la licence ouverte favorise la réutilisation des données provenant du secteur public, en autorisant non seulement la reproduction, la redistribution, et l'adaptation des données, mais aussi leur exploitation pour des finalités commerciales.³⁷ La seule restriction est celle de toujours mentionner la paternité des données, sans que cela ne soit vu comme une reconnaissance ou une caution morale de la part du producteur des données. Ainsi, la licence ouverte se déclare explicitement compatible avec l'Open Government Licence (OGL) du Royaume-Uni, la licence Creative Commons Attribution 2.0 (CC-BY 2.0) de Creative Commons, et l'*Open Data Commons Attribution* (ODC-BY) de l'Open Knowledge Foundation, ainsi que toute autre licence qui demande uniquement que la paternité des données soit mentionnée.³⁸

La licence précise aussi que le producteur des données n'est pas responsable de la manière dont les données sont réutilisées ; c'est aux ré-utilisateurs de s'assurer que leur réutilisation ne risque pas d'induire en erreur des tiers en ce qui concerne la source, la date ou le contenu des données (la non-dénaturation des données).

Actuellement, au-delà d'être utilisées par les administrations publiques, la licence ouverte a aussi été adoptée par des villes et des collectivités ou institutions territoriales, telles que la ville de Montpellier, Marseille, Arles, le conseil général des Hauts-de-Seine, l'observatoire de l'économie et des territoires de Loir-et-Cher, la Marie de Versailles, etc. ainsi que par des établissements publics tel que, notamment, la Bibliothèque nationale de France (BnF) pour son projet *data.bnf.fr*.

La région Provence-Alpes-Côte-d'Azur a adopté, elle, un système de licences mixtes en acceptant sur son portail *opendata.regionpaca.fr* tout jeu de données mis à disposition du public sous une licence permettant leur réutilisation gratuite - y compris la licence ouverte, l'ODbL ou la licence Creative Commons CC-BY (Attribution)³⁹.

Ainsi, malgré l'existence de cette licence de référence, l'utilisation des licences en France reste des plus hétérogène. Cela est illustré dans le catalogue des informations publiques de la Direction de l'information légale et administrative (DILA) qui recense l'ensemble des fonds réutilisables dans le cadre des conditions prévues par la loi (les données juridiques, économiques et financières, les débats des assemblées, les données administratives, etc.) en précisant, le cas échéant, que la réutilisation de ces données est soumise aux conditions d'une licence. Dans ce catalogue, diverses icônes signalent les modalités de réutilisation des données (soumises au paiement d'une redevance ou à des conditions particulières, ou bien soumises à la licence ouverte et donc librement accessibles et réutilisables). Ces explications fort utiles confortent notre hypothèse de la complexité des règles à prendre en compte pour un réutilisateur qui doit faire une distinction – parfois difficile – entre les données publiques qui sont librement réutilisables et celles qui ne le sont pas.

Conclusions

En détaillant les expériences concrètes de l'*Open Data* et les politiques nationales ou locales mises en œuvre, nous avons cherché à montrer que le principe est en train de se généraliser mais qu'il est diversement interprété et appliqué. Cette diversité est susceptible de poser de nouvelles difficultés car la raison même de l'*Open Data* – à savoir la réutilisation des données – risque de ne pas être effective. En effet, dans la mesure où elles n'appartiennent pas au domaine public, les données publiques sont soumises à des licences différentes susceptibles de rendre les données juridiquement incompatibles (notamment dans le cas des licences de partage à l'identique). Bien qu'une standardisation soit en train de s'organiser *de facto* entre les licences de réutilisation des données, il reste un détail important pour assurer la concrétisation du principe d'*Open Data*. Les formats peuvent, en effet, mettre en échec ces principes en limitant l'interopérabilité technique des données.

Ainsi, l'interopérabilité – légale et technique – des données constitue une précondition fondamentale de l'*Open Data*. Au-delà de la compatibilité des licences, il est important que les données soient mises à disposition du public dans des formats ouverts et interopérables, afin de donner à tout acteur la possibilité d'extraire, mais surtout de croiser ou de « remixer » les données. L'enjeu principal est de rendre ces données utiles et utilisables pour mettre en commun des ressources et assurer un service qui prolonge l'action publique. On peut alors désenclaver la réflexion sur l'intérêt pour l'administration d'ouvrir ses données : il s'agit pour elle de répondre à des besoins *différenciés* qui peuvent ne pas être

36 Plus de détails sur le site <http://blog.etalab.gouv.fr/>

37 Or, un détail important risque d'échapper à une première lecture superficielle de la licence. Alors que la Licence Ouverte autorise la reproduction, la redistribution, l'adaptation et la réutilisation des données quelle qu'en soit la finalité, elle ne permet cependant l'exploitation des données que pour des finalités commerciales. Bien que cela ne soit probablement qu'un oubli plutôt qu'un acte délibéré destiné à prévenir l'exploitation non commerciale des données, cette lacune est cependant significative. Il semblerait que, lors de la rédaction de cette licence, le comité Etalab n'ait pas pris en compte la possibilité que le secteur privé puisse intervenir dans l'exploitation des données sur une base non commerciale. Ceci est d'autant plus inadéquat dans le contexte numérique, où le nombre d'initiatives visant à la réutilisation des données sans finalité commerciale ne cesse d'augmenter et que la proportion de ces initiatives mises en œuvre directement par les membres de la société civile devient de plus en plus importante.

38 Sans entrer dans les détails, cette compatibilité n'est pas réelle dans la mesure où la licence ouverte exige non seulement la mention de la source (attribution) mais aussi l'indication de la date de dernière mise à jour.

39 <http://creativecommons.org/licenses/by-sa/2.0/fr/>.

assurés par elle seule (une plateforme collaborative Web 2.0 par exemple) ou qui pourraient l'être avantageusement par une entreprise ou une communauté d'internautes⁴⁰.

Or, aucune des licences *Open Data* jusqu'à présent développées ne comporte une clause indiquant que les données doivent être mises à disposition du public dans un format ouvert, interopérable et, si possible, structuré.

40 Développer la collaboration entre l'Administration et ses publics est devenu un nouveau mot d'ordre que l'administration Obama avait promu dès 2009 pour développer le site gouvernemental : « *Nous ne pouvons relever les défis du XXIe siècle avec une bureaucratie du XXe siècle* ». *Enabling collaboration : three priorities for New Administration, Report, N Y, NAPA, 2009 www.collaborationproject.org*.