
Réutilisation de données publiques

Une étude comparative entre quatre pays

Gabriella Salzano

Équipe de recherche DICEN (Dispositifs d'information et de communication à l'ère numérique) EA 4420, CNAM. Case I461 - 2, rue de Conté – 75003 Paris, France

*Université Paris Est Marne-la-Vallée
5, Bd. Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2, France
gabriella.salzano@univ-mlv.fr*

RÉSUMÉ. *La dynamique de réutilisation des données publiques (DP) est impulsée par les avancées technologiques, les volumes croissants des données publiées ou recueillies par les institutions, et les cadres législatifs nationaux et européen. Ce papier réalise une analyse comparative de cette dynamique dans trois pays d'Europe (Royaume Uni, France et Italie) et aux Etats Unis, à partir de plateformes institutionnelles se situant à plusieurs échelles territoriales (nationales, régionales, ...). En se focalisant sur les applications et leurs métadonnées, nous apportons des éléments quantitatifs et qualitatifs d'évaluation. Nous identifions aussi d'une part des leviers pour développer cette dynamique dans des contextes plus larges que ceux de la production des données, mais aussi des freins scientifiques et des initiatives encourageantes pour les dépasser.*

ABSTRACT. *The reuse of public data (PD) is driven by technological advances, by the increasing volumes of published or collected PD by institutions and citizens, as well as national and European legislative frameworks. This paper compares this dynamic in three European countries (United Kingdom, France and Italy) as well as in the United States, from institutional platforms and on several territorial scales (national, regional,...). By focusing on applications and their metadata, we provide quantitative and qualitative evaluation elements. We also identify on the one side levers to develop this dynamic in broader contexts than production data, and, on the other side, scientific obstacles and encouraging initiatives to overcome them.*

MOTS-CLÉS : *Systèmes d'Information, données publiques, réutilisation, évaluation, applications, hétérogénéité, distribution, formats, Linked Data, couverture géographique, évolution.*

KEYWORDS: *Information Systems, public data, reuse, evaluation, applications, heterogeneity, distribution, formats, Linked data, geographic coverage, evolution*

1. Introduction

La production massive des données et leur ouverture constituent des leviers phénoménaux pour la production de nouveaux services et la génération de nouvelles connaissances (Shadbolt, 2010). Ces évolutions nécessitent de nouvelles approches et techniques d'interopérabilité. L'ouverture, condition essentielle à l'interopérabilité des Systèmes d'Information (SI), concerne aussi bien les données que les logiciels, les standards et les licences. L'accessibilité, l'absence de restrictions techniques et la réutilisation figurent parmi les critères d'ouverture des données, formulés par l'Open Knowledge Foundation¹. En particulier, la propriété de réutilisation est liée à la licence d'apporter des modifications et produire des travaux dérivés, qui peuvent être distribués dans les mêmes termes que les données d'origine.

Intéressantes par leur ouverture et fiabilité, les données publiques (DP) sont des données collectées, maintenues et utilisées par les organismes publics dans le cadre de leurs missions (Martinez *et al.*, 2007). Elles sont très variées (données statistiques, sociales, économiques ou environnementales, documents d'archives, collections d'ouvrages, ...) et à priori gratuites, mais ceci n'est pas toujours le cas. Les DP peuvent ouvrir des nouveaux marchés aux développeurs, favoriser la croissance et augmenter la transparence du secteur public.

1.1 *Le contexte initial et ses évolutions*

Le développement des DP, déclenché par le programme électoral du Président Obama en 2009, s'est désormais étendu dans un grand nombre de pays. Il est régi par un cadre législatif complexe, qui se déploie avec des spécificités dans plusieurs états, mais aussi au niveau international et européen (Boustany, 2010). En France les DP sont produites ou reçues par plusieurs entités, comme l'Etat, les entreprises publiques, les collectivités territoriales, les autorités administratives indépendantes. Elles sont réutilisables librement et gratuitement dans les conditions de la « Licence Ouverte / Open Licence », ou dans les conditions d'autres licences Open Data (comme les licences OGL, ODC-BY, CC-BY 2.0, ODbL, ou autres). Au sein de l'Union Européenne (UE), la directive 2003/98/CE, facilitant la réutilisation des Informations du Secteur Public, et la directive 2007/2/CE INSPIRE, préconisant l'ouverture des données géo spatiales, accompagnées de leurs métadonnées, ont donné une très forte impulsion à l'ouverture et à la réutilisation des DP.

La progression constante des volumes de données produites ainsi que les évolutions des technologies d'analyse, d'exploitation et de traitement des données, permettent désormais la création de nouveaux services et de nouvelles applications fondés sur l'utilisation, l'agrégation ou la combinaison de données. Les besoins de réutilisation de DP dépassent de plus en plus les frontières des Etats, dans plusieurs

¹ Open Knowledge Foundation, <http://opendefinition.org/okd/francais/>

secteurs, comme les transports et la météo. Ainsi, les évolutions prévues par l'UE visent à favoriser une utilisation efficace des Informations du Secteur Public au travers de plusieurs pays. Elles préconisent aussi que les données, accompagnées de leurs métadonnées, soient mises en ligne dans un format "suffisamment structuré, pour que des applications logicielles puissent reconnaître sans ambiguïté chaque fait exposé et leurs structures internes"². La Commission Européenne a aussi invité tous les Etats membres de l'Union à se positionner sur le déploiement généralisé de l'accès ouvert aux résultats de leur recherche dans les meilleurs délais, dans la perspective du programme Horizon 2020³.

1.2 Problématique de recherche

La mise à disposition des DP est un phénomène global, très récent et en forte expansion. Ainsi, la dynamique de la réutilisation des DP est une thématique de recherche qui interpelle de façon transversale l'Ingénierie des Systèmes d'Information pour les Etats. La problématique associée couvre plusieurs questions, d'ordre méthodologique, scientifique et pratique, du type : quelles plateformes analyser ? comment peut-on les analyser et quelles mesures sont possibles ? quelles sont les plateformes dont les données sont le plus réutilisées et les caractéristiques des nouvelles applications ? quels sont les facteurs moteurs de la réutilisation des DP, les verrous scientifiques et les approches pour faire face à ces verrous ?

Pour proposer des réponses à ces questions, nous réalisons une étude comparative de la dynamique de la réutilisation des DP dans trois pays d'Europe et aux Etats Unis.

2. Démarche

L'adoption d'une approche empirique, pertinente pour analyser par une observation directe un objet d'étude très contemporain et en rapide progression (Yin, 2009), nous conduit à dégager les étapes suivantes :

- Analyse des travaux liés à la réutilisation des DP, selon trois axes : a) évaluation et modélisation de portails de mise à disposition de DP ; b) DP et Linked Data ; c) exemples de réalisations basées sur la réutilisation des DP (section 3).
- Choix des plateformes. Ce choix a porté sur des sites institutionnels, à divers niveaux de couverture géographique. Au stade actuel de la réutilisation des DP, cette

² Proposition de directive du Parlement Européen et du Conseil, modifiant la directive 2003/98/CE concernant la réutilisation des informations du secteur public, Page 17, http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive_proposal/2012/fr.pdf

³ Recommandation de la Commission aux Etats Membres, relative à l'accès aux informations scientifiques et à leur conservation, 17/7/2012, <http://goo.gl/uNTSq>

étape est délicate, car il n'existe pas de catalogue centralisé de tous les portails de mise à disposition des DP, les classifications existantes⁴ sont très hétérogènes et le portail *Open Data* de l'UE⁵ vient d'ouvrir (décembre 2012, en version *beta*).

Le panel initial a concerné le niveau national, avec deux sites pionniers, véritables références, data.gov (US) et data.gov.uk (UK), et deux sites plus récents, data.gouv.fr (FR) et dati.gov.it (IT). Relevant d'autorités gouvernementales, ces sites disposent d'un pouvoir institutionnel pour imposer leur choix à l'échelle d'un pays.

Pour saisir des tendances globales de la réutilisation, on a complété ce panel par des nombreux portails de collectivités territoriales en France et Italie, ainsi que deux sites anglo-saxons, concernant l'état de New York et la ville de Londres.

– Observation des plateformes (sections 4, 5 et 6). Elle se focalise sur les *applications basées sur la réutilisation des données publiques*, le terme *application* représentant schématiquement un outil qui contribue à réaliser une tâche ou à chercher une information⁶. Cette étape est aussi délicate : compte tenu de l'hétérogénéité des plateformes (ancienneté, organisation, ...), les modes de collecte d'informations et les éléments recueillis diffèrent parfois de l'une à l'autre.

Pour chaque portail, si l'onglet "Applications" n'existe pas, on utilise d'autres méthodes de navigation pour réaliser un recueil global d'informations. Celui-ci porte sur les volumes, les thématiques prioritaires, les plateformes techniques utilisées et les producteurs. On analyse ensuite les métadonnées détaillées des applications ainsi que les démarches de "captation" des développeurs (concours, incitations). Les sites étant très dynamiques, on exprime des ordres de grandeur pour les applications, plus que des volumes précis. Les volumes de données mis à disposition par les sites ne sont pas renseignés, car pas du tout comparables : des plateformes considèrent comme unitaires des jeux de données qui en contiennent plusieurs autres, tandis que d'autres, comme la plateforme (FR), comptabilise chaque fichier.

– Analyse comparative de la réutilisation des DP entre les quatre pays, en termes de couvertures géographiques, thématiques, plateformes, métadonnées, ainsi que d'exploitation du potentiel de DP, de leviers et freins (section 7).

3. Travaux liés et réalisations

Le cadre théorique qui guide notre évaluation de la dynamique de réutilisation des DP, dans un contexte global d'ouverture des données, s'inscrit dans le thème de l'« ingénierie des évolutions » des SI. Ce thème comprend l'analyse et la conception de nouveaux SI pour répondre aux évolutions des législations, des organisations, des données et des processus métiers (Akoka *et al.*, 2009). Les DP mises à la disposition

⁴ Ces classifications sont établies par exemple par l'European Public Sector Information Platform, (<http://epsiplateform.eu>) ou par la Semantic Web Company, <http://goo.gl/bVaDv>

⁵ Portail Open Data de l'Union Européenne, <http://open-data.europa.eu/>

⁶ data.gov, <http://www.usa.gov/Mobile/AppsGallery/FAQs.shtml>

par des portails institutionnels sont vues comme des sources de données hétérogènes et distribuées à très forte valeur ajoutée pour la réutilisation et l'interopérabilité.

Cette étude de la littérature, nécessairement synthétique, analyse des travaux de recherche liés à notre problématique et concernant l'évaluation et la modélisation de portails de mise à disposition de DP ainsi que les liens entre DP et Linked Data, avant de recenser des applications significatives de réutilisation de DP.

3.1 Evaluation et modélisation de portails de données publiques

Nous analysons différentes approches d'évaluation et modélisation des portails de mise à disposition des DP, avant de positionner notre travail.

Le séminaire international Using Open Data (W3C, 2012), réalisé sous l'égide du consortium W3C, synthétise l'analyse d'applications basées sur la réutilisation des DP et des travaux de recherche en ce thème.

En écho au modèle à 5 étoiles établi par T. Berners-Lee (Heath *et al.*, 2011) pour classer les niveaux d'ouverture des données, (Davies, 2012) modélise l'engagement des administrations dans le processus de mise à disposition des données selon 5 niveaux: 1) être guidé par la demande ; 2) décrire complètement les données (métadonnées relatives à la création, aux mises à jour, aux utilisations) ; 3) supporter des activités conversationnelles avec les utilisateurs (annotations, commentaires, remarques, contacts) ; 4) fournir des outils, des tutoriels, des plateformes facilitant l'utilisation de ces données ; 5) collaborer autour des données avec d'autres organismes (partage de ressources communes).

(Ferro *et al.*, 2011) s'intéressent aux aspects commerciaux de la réutilisation des DP. Leurs questions de recherche traitent de création de la valeur et concernent : l'existence d'entreprises qui tirent des profits de la réutilisation de DP ; des mécanismes mis en œuvre pour donner du sens aux données (agrégation, géo-référencement, validation, combinaison de contenus et de services, visualisation, ...) ; des différents rôles joués par les DP dans la détermination des prix (ingrédient clé ou marginal d'un produit ou d'un service), et des modes de rémunération possibles (payant, soumis à conditions de temps ou de volumes, gratuit).

L'Open Data Index⁷, élaboré au sein du portail Open Data de la Commune de Bologne, en Italie, mesure le potentiel des sites à publier des données. Il réunit des critères prenant en compte 5 paramètres : le potentiel d'utilité des données pour le citoyen, les entreprises ou les institutions ; la facilité d'exporter des informations ; les types de formats des données exportables ; la qualité des données disponibles ; la fraîcheur des données par rapport à la finalité.

⁷ Open Data Index, <http://dati.comune.bologna.it/node/154>

Au niveau de l'UE, un groupe de travail vient de voir le jour, avec l'objectif d'harmoniser les métadonnées, en définissant les vocabulaires contrôlés et les classes essentielles pour faciliter les échanges de données entre portails⁸.

Une étude comparative sur la politique d'accès aux documents publics aux Etats Unis, au Royaume Unis et en France (Boustany, 2010) analyse l'incidence sur la vie politique de l'usage des TIC, associé aux réformes politiques. Le Royaume-Uni, suivi par les Etats-Unis, apparaissent comme les pays le plus en avance dans la mise à disposition des DP et la sollicitation citoyenne.

(Salzano, 2012) propose des éléments de conception d'un observatoire de recherche sur les DP en santé. Les dimensions de recherche concernent : a) l'organisation, pour l'implication des institutions et des utilisateurs ; b) les méthodologies, pour analyser l'offre actuelle de plateformes de DP, avec ses limites, et identifier les approches de médiation et de recherche d'information sur le Web pour pallier à leur fragmentation ; c) les architectures et les technologies, pour relever les défis de conception de ces portails, en termes d'ergonomie, de support des volumes et de puissance des services proposés.

L'approche ici proposée se positionne dans le cadre de cet observatoire de recherche, notamment en termes d'objectifs (analyse des plateformes actuelles de DP). Elle se focalise sur la réutilisation des DP et étend la portée de l'étude à toutes les thématiques. Elle apporte une attention particulière aux métadonnées des applications, ainsi qu'aux leviers et freins pour la réutilisation.

3.2 Données publiques et Linked Data

La mise à disposition des DP issues des institutionnels constitue un levier exceptionnel et de qualité au développement du Web des Données (Linked Data Web, LDW). Le LDW contient des données identifiées par des URI et représentées dans le format RDF, apte à relier ces données entre elles par des traitements automatiques. (Shadbolt *et al.*, 2012) observent que le LDW est actuellement largement peuplé par les initiatives DBpedia, DBLP Computer Science Bibliography, Comprehensive Knowledge Archive Network (CKAN) et par des journaux, comme la London Gazette et le New York Times. Leur analyse du portail data.gov.uk, pionnier dans la diffusion de DP en RDF, conduit à identifier des axes, notamment technologiques, pour une plus large évolution du LDW : mise à disposition d'interfaces réellement conviviales et adaptées aux utilisateurs finaux, avec possibilité de définir des points de jointure pour construire des vues cohérentes issues de diverses sources. Les points de jointure prioritaires sont fournis par des

⁸ Communiqué "European Commission Creates a Working Group to Define Common Metadata", 22/03/2013, <http://epsiplatform.eu/content/european-commission-creates-working-group-define-common-metadata>

entités de nature géographique ou temporelle, ainsi que par les producteurs de données.

(Ding *et al.*, 2011) formalisent les étapes pour migrer des données ouvertes publiées par les institutionnels vers le LDW : catalogage des sources disponibles ; capture d'un état des DP input pour la conversion, ainsi que de leurs métadonnées, notamment provenance et version des données ; conversion de ces données au format RDF. Le portail Linking Open Government Data⁹ a adopté cette démarche pour publier des versions Linked Data de données gouvernementales. Il fait état de la création d'environ 1900 jeux de données au format RDF à partir de 119 sources, aux formats natifs .xls ou .csv, issues des portails data.gov et data.gov.uk. Environ 1700 de ces jeux de données ont été enrichis en utilisant 9500 liens vers d'autres données liées. La gestion des versions des données permet d'évaluer les changements dans les données publiées dans les plateformes.

3.3 Réalisations basées sur la réutilisation de données publiques

Eurostat¹⁰ et l'OCDE (Organisation de Coopération et de Développement Economiques¹¹) fournissent des exemples "historiques" de réutilisation de DP issues de différents pays :

– Eurostat, Direction Générale de la Communauté Européenne depuis 1958, est l'Office statistique de l'Union Européenne chargé de fournir des statistiques au niveau européen. Cette Direction a mis en place une méthodologie statistique rigoureuse pour assurer une très haute qualité des informations diffusées, permettre des comparaisons entre les pays et les régions et répondre à des besoins opérationnels. Eurostat est le producteur principal du portail de l'UE opendata.eu.

– l'OCDE, officiellement née en 1961, a pour mission de fournir des analyses et des avis de grande qualité aux gouvernements et responsables publics, "*dans l'intérêt des citoyens, et conformément à sa mission qui est de proposer « des politiques meilleures pour une vie meilleure »*". L'OCDE envisage de rendre accessibles, utilisables et gratuites, d'ici au milieu de l'année 2015, toutes ses données et d'instaurer un système de gestion de l'information et des connaissances.

Parmi les réalisations plus récentes de réutilisation de DP issues de différents pays, nous citerons le programme ESPON¹² et le projet Reegle¹³ :

⁹ Linking Open Government Data, <http://lod2.tw.rpi.edu/>

¹⁰ Eurostat, <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>

¹¹ OCDE, <http://www.oecd.org/fr/> et OCDE, Rapport "Semaine de l'OCDE 2012", <http://www.oecd.org/fr/rcm/50452390.pdf>

¹² ESPON, désigné en français avec l'acronyme ORATE (Observatoire en réseau de l'aménagement du territoire européen et de la cohésion territoriale), <http://www.espon.eu/main/>

– ESPON (European Spatial Observatory Network on territorial development and cohesion), est un programme européen visant à soutenir une politique régionale axée sur la cohésion territoriale et la compétitivité. Les 27 Etats membres de l'UE, ainsi que l'Islande, le Liechtenstein, la Norvège et la Suisse y participent. Pour identifier les potentiels de chaque région, ESPON produit des outils destinés aux analystes et aux décisionnaires de tous niveaux (régional, national, européen). Ainsi, un portail de bases de données, alimenté aussi par des données d'Eurostat, fournit des outils décisionnels, de contrôle de qualité et des données territoriales, tandis que un HyperAtlas¹⁴ permet de comparer et analyser multiples indicateurs, démographiques ou économiques, sur plusieurs territoires.

– Reegle (Search Engine for Renewable Energy and Energy Efficiency), portail d'information initié en 2006, a été complètement reconçu en 2010 par la Semantic Web Company, sous l'égide d'organismes internationaux. Opérant dans le domaine des énergies renouvelables et de l'efficacité énergétique, il s'adresse à une pluralité de parties prenantes en Europe : institutionnels, industriels, citoyens, chercheurs, gestionnaires, associations. Il analyse les contenus de plus de 550 sites web concernant la législation, les acteurs, les projets et les rapports dans ce domaine, et fournit des "profils" énergétiques par pays (production consommation, programmes, ...), des statistiques actualisées automatiquement et un moteur de recherche. Reegle illustre la mise en œuvre de bonnes pratiques d'exploitation des Linked Data et des standards du W3C (format RDF, langage SPARQL)¹⁵ : son Catalogue des Acteurs relie plus de 1700 initiatives, partenaires et organisations.

4. Analyse de la réutilisation des DP aux Etats Unis et au Royaume Uni

On analyse en détail les portails nationaux de ces pays pionniers de l'Open Data, avant de survoler deux sites majeurs ayant une couverture territoriale restreinte.

4.1 Etats Unis

La plateforme data.gov, démarrée en mai 2009 sous l'impulsion du gouvernement fédéral, part du principe que les données produites par les administrations sont des biens publics et doivent être mises à la disposition des contribuables. Elle publie plus de 1500 applications reparties dans trois catégories¹⁶ :

¹³ Reegle, <http://www.reegle.info/>

¹⁴ L'HyperAtlas est développé dans le cadre d'un contrat coordonné par le laboratoire LIG de l'Université Joseph Fourier, de Grenoble, en France

http://www.espon.eu/main/Menu_ToolsandMaps/ESPONHyperAtlas/index.html

¹⁵ Linked Open data : the essential, <http://www.reeep.org/sites/default/files/LOD-the-Essentials.pdf>

¹⁶ Les catalogues de ces catégories sont consultables respectivement aux adresses : <https://explore.data.gov/catalog/apps>, <http://www.data.gov/developer-apps-showcase> et <http://apps.usa.gov>

– Les *government apps* (1200), produites par l'U.S. Federal Executive Branch, comprennent environ 850 extracteurs de données, des fils RSS (270) et des outils interactifs (*widgets*, 75), de type cartes interactives, informations sur les transports ou la météo. Les thématiques les plus représentées sont : geography and environment (205), transportation (85), information and communication (82), sciences and technologies (72), education (42), business enterprise (52), law enforcement, courts and prisons (35), energy (29), health and nutrition (14). L'application *Airline On-Time Performance and Causes of Flight Delays*¹⁷ figure parmi les plus consultées du public.

– Les *citizen-developed apps* (230) comprennent celles sélectionnées à l'issue de concours internationaux. Leur valeur ajoutée consiste à transformer plusieurs données brutes en information, en rendant cette informations plus accessibles par le public, à l'aide de *mashups* et d'outils de visualisation sur PC ou téléphone. Ces applications sont utilisées pour des prises de décisions dans des domaines aussi divers que l'analyse du marché du travail, au niveaux locaux, régionaux ou des états (*Employment Market Explorer*¹⁸) ou la planification d'activités en fonction de la qualité de l'air actualisée en temps réel (*AIRNow*¹⁹). Cependant, comme ces applications ne sont pas rattachées au gouvernement fédéral, data.gov ne peut pas être garant des implications de leurs utilisations ni de la qualité de l'information diffusée.

– Les *mobile apps* (100) cataloguées dans l'*Apps Gallery* au sein de USA.gov ont pour objectif d'apporter au plus large public des informations ou des services sur des supports mobiles (Android, Apple, Mobile Web, Blackberry). L'accès aux données personnelles de l'utilisateur n'est pas systématique. Ces applications concernent une multitude de catégories et poursuivent divers objectifs, comme chercher des centres de vaccination, des bibliothèques ou comparer les prix du gaz sur un territoire.

Le *Developers' Corner* incite les développeurs à rejoindre la communauté de l'*Open Data* via des ressources pour interconnecter des données gouvernementales, à l'aide d'approches relevant du Web sémantique ou d'outils de visualisation, de gestion de contenus, comme Dupral, et d'interrogation, comme SPARQL. La communauté *data.gov Semantic Web*, en conjonction avec le W3C, recommande des standards pour les données institutionnelles et leurs métadonnées, et sollicite les développeurs à créer une nouvelle génération des mashups de données liées.

Les métadonnées associées aux applications sont très disparates. Les métadonnées les plus exhaustives concernent les *government apps* et renseignent sur la catégorie, les droits, les tags, le site du producteur, les dates de mise à disposition, de modification et éventuellement la fréquence de mise à jour, ainsi que sur des aspects d'activité (visites, téléchargements, commentaires). Les datasets (jeux de données) utilisés par ces applications sont aussi détaillés très finement, en termes de couverture

¹⁷ Cette application est publiée à l'adresse <http://goo.gl/uXE4K>.

¹⁸ Employment Market Explorer : <http://pujaplicaciones.javeriana.edu.co/Employment/>

¹⁹ AIRNow : <http://m.epa.gov/apps/airnow.html>

(unité d'analyse, granularité, couverture géographique), d'acquisition (modalités et moyens techniques, dictionnaire de données), de documentation technique et éventuellement de méthodologie statistique. Des flags indiquent si l'application a été suggérée par le public, la présence de labels et chartes de qualité, ainsi que de clauses de confidentialité, si cela est pertinent. Les métadonnées des applications recensées dans les deux autres catalogues décrivent de façon succincte l'application elle-même mais pas les données utilisées.

Concernant des portails de DP à couverture géographique plus restreinte, le catalogue *opendatasites*²⁰ récence aux USA 39 portails pour les Etats et 34 portails de villes et comtés. A titre d'exemple, le portail de l'état de New York (<http://goo.gl/z2jkf>) répertorie une dizaine d'applications, décrites très succinctement et concernant notamment la météo et les transports.

4.2 Royaume Uni

Le site data.gov.uk a démarré en janvier 2010, sous l'impulsion du Premier Ministre, prônant la libération massive des données publiques pour promouvoir la transparence, l'efficacité et l'innovation sociale. Il publie environ 230 applications et incite fortement les développeurs à en soumettre des nouvelles pour les partager.

Les tags les plus représentés et les applications les mieux placées (*rating*) concernent la météo, les transports, la santé, mais des thématiques liées à la criminalité sont aussi présentes. Environ les trois quarts des applications s'adressent au secteur privé, les autres étant destinées au secteur public et aux services.

Le tag *rdf data* identifie quelques applications, comme *GB Roads Visualisation* et *WeatherChart*²¹. La plateforme présente un onglet *Linked data*, riche de nombreuses ressources et annonçant les API qui permettent de charger progressivement des données au format RDF, notamment dans les domaines des transports et de l'éducation.

Les métadonnées décrivant chaque application renseignent sur le développeur, les commentaires, les tags, les datasets utilisés et les liens vers ces données, les applications liées et les applications similaires. Ces informations sont complétées par une description textuelle et des écrans d'exemples.

En ce qui concerne des portails à couverture territoriale plus restreinte, le catalogue *opendatasites* en récence une quinzaine, au niveau des villes, comtés ou districts. Le *datastore* du portail londonien²² réunit environ 75 applications. Il incite

²⁰ <http://www.data.gov/opendatasites>

²¹ Ces applications sont publiées respectivement aux adresses <http://data.gov.uk/apps/gb-roads-visualisation> et <http://data.gov.uk/apps/weather-chart>.

²² <http://data.london.gov.uk/datastore>

la créativité des potentiels contributeurs avec des *inspirational uses*, en mettant en vitrine des applications sur le recensement, les élections ou la location de vélos.

5. Analyse de la réutilisation des données publiques en France

Au niveau national, la plateforme data.gouv.fr (décembre 2011) vise à rassembler et à mettre à disposition librement les DP de l'Etat, de ses établissements publics administratifs et, si elles le souhaitent, des collectivités territoriales et d'autres structures chargées d'une mission de service public²³. Fin 2012 a été créé le secrétariat général pour la modernisation de l'action publique, auquel est rattachée la mission Etablab, chargée de promouvoir et développer data.gouv.fr.

Le site ne présente pas d'onglet spécifique pour les Applications. La recherche des "application web" à partir de l'onglet "Data" conduit à 2 applications²⁴ : leurs métadonnées indiquent une brève description, le producteur, la date de publication, les données utilisées. L'onglet "Articles" conduit aussi à ces applications, ainsi qu'aux annonces de l'utilisation des données de data.gouv.fr par des plateformes, comme *OpenStreetMap*²⁵, pour compléter ses cartes interactives, et *data.publica*²⁶, pour l'accidentologie routière.

Pour stimuler la réutilisation des DP, la mission Etablab a lancé deux concours *dataconnections*. Les applications lauréates du 1^{er} concours n'apparaissent pas dans le site data.gov.fr. Elles se répartissent dans les 3 catégories : Projets Grand Public : a1) Fourmisanté, a2) Home'n go; Projets Utilité Publique : b1) CitizensUp, Voxe, qui a reçu les félicitations du jury ; Projets Professionnels : c1) La place de l'immobilier, c2) Webshell, c3) Open Data Ware²⁷. Le deuxième concours *dataconnection* reconduit ces catégories et introduit la thématique principale "Mobilité et territoires", qui comprend le transport, le tourisme, la vie ou l'information locale, notamment géographique.

Au niveau des collectivités territoriales, un nombre limité de sites disposent d'un onglet "Applications" ou "Cas d'utilisation".

Les plateformes Paris Data²⁸ et Data Rennes Métropole²⁹ sont celles qui recensent le plus grand nombre d'applications, de l'ordre de 80 et 20 respectivement.

²³ <http://www.data.gouv.fr/A-propos>

²⁴ data.gouv.fr, onglet "Data", puis recherche de "application Web".

²⁵ OpenStreetMap, <http://www.data.gouv.fr/Articles/OpenStreetMap-complete-sa-carte-du-monde-interactive>

²⁶ Application d'accidentologie routière, <http://goo.gl/DTySJ>, publiée dans le portail www.data-publica.com

²⁷ Ces applications sont publiées aux adresses Web suivants : a1) www.fourmisante.com, a2) www.homengo.com, b1) www.voxe.org, c1) www.laplaceleimmobilier-pro.com, c2) www.webshell.fr, c3) www.opendatasoft.com.

²⁸ Paris Data, <http://opendata.paris.fr/opendata/jsp/site/Portal.jsp>

Toutefois, la présence de l'application *SNCF* parmi les premières et d'une application *test2* parmi les secondes suscite des interrogations sur les critères conduisant à ces sélections.

Montpellier Territoire Numérique³⁰ publie 14 applications, pouvant être triées par éditeur et reparties dans les catégories suivantes : Transport (2), Environnement (3), Arts & Culture (2), Proximité (1), Localisation (4), Sports & Loisirs (1), Patrimoine / Tourisme (1). L'application gratuite *Handimap.org*, présente dans les deux portails de Rennes Métropole et de Montpellier Territoire Numérique, permet de calculer des itinéraires accessibles aux personnes à mobilité réduite et d'afficher différents points d'intérêts liés à l'accessibilité. Elle est basée sur la réutilisation des DP et une démarche systématique de *cartoparties*³¹, permettant de construire et enrichir des fonds de cartes.

VisitProvence³² publie 7 applications pour Android et Iphone, sur le thème du tourisme et utilisant des données diffusés sur l'Open Data des Bouches-du-Rhône.

La Rochelle³³ publie une API pour télécharger des DP depuis sa plateforme, ainsi qu'une application gratuite pour Smartphones et tablettes de type Android, dans la catégorie Espace publique, concernant les déplacements des personnes à mobilité réduite. Grâce à cette application, la ville a reçu le trophée Ville citoyenne 2012 de l'APAJH (Association Pour Adultes et Jeunes Handicapés).

Le portail Proxima Mobile³⁴ coordonné par la Délégation aux Usages de l'Internet (DUI), rattachée au Ministère chargé de l'Industrie, de l'Énergie et de l'Économie numérique, recense plusieurs de ces applications. Reconnu comme le premier portail européen de services d'intérêt général sur terminaux mobiles, il a comme principal objectif de sélectionner et labelliser des nouveaux services qui répondent aux quatre critères de sa charte (intérêt général, ergonomie, gratuité pour l'utilisateur final, absence de publicité).

6 Analyse de la réutilisation des données publiques en Italie

Au niveau national, la plateforme dati.gouv.it (octobre 2011) est rattachée au ministère en charge de la modernisation de l'administration publique. Elle publie environ 140 applications classées par : a) type de plateforme technique, principalement iPhone (99), Android (64) et Web (22) ; b) niveau administratif : central (29), régional (24), provincial (11) et communal (37) ; c) thématiques prioritaires : sport et temps libre, art et culture, politique et institutions,

²⁹ Data Rennes Métropole, <http://www.data.rennes-metropole.fr>

³⁰ Montpellier Territoire Numérique, <http://opendata.montpelliernumerique.fr/>

³¹ Cartoparties, <http://montpellier.territoirenumerique.org/opendata/cartoparty/>.

³² VisitProvence, <http://data.visitprovence.com/>

³³ La Rochelle, <http://www.opendata.larochelle.fr>

³⁴ Proxima Mobile, <http://www.proximamobile.fr/>

environnement, transports et infrastructures ; d) producteur, privé ou public. Il est possible de signaler une application et de la décrire selon ces critères, compatibles avec les métadonnées Dublin Core.

Le site publie deux applications basées sur RDF et les Linked data : a) LodLive³⁵, définie comme le premier navigateur de ressources RDF, est basé uniquement sur des *endpoints* SPARQL et opère sur plusieurs sources, comme dbpedia.org, yago-knowledge.org, data.nature.com ou worldbank.270a.info ; b) ComunicaPA³⁶, application mobile permettant aux citoyens et aux entreprises d'obtenir des informations en Linked Data sur l'administration publique et de les utiliser pour naviguer sur les sites officiels, téléphoner ou envoyer des messages.

Au niveau des collectivités territoriales, un nombre limité de portails disposent d'un onglet "Applications" ou "Cas d'utilisation"³⁷. Le tourisme et les transports sont les thèmes prioritaires de ces portails.

Nous avons remarqué que : a) la commune de Rome publie une quarantaine d'applications dans les catégories "Grandes Entreprises", "Petites et moyennes entreprises" ou "Développeurs Individuels", 6 de ces applications ayant reçu des prix ; b) la région de la Vénétie publie 4 applications pour Android et Iphone, dans les thèmes suivants : météo, balnéation, itinéraires et tourisme. Cette région pilote aussi un projet international collaboratif au sein d'une initiative d'harmonisation des données ouvertes dans les pays de la Méditerranée (projet Homer³⁸) ; c) d'une façon originale, la région Piémont publie, parmi 8 "cas d'utilisation", des applications sur le thème de l'éducation.

Tous ces portails, ainsi que ceux des Communes de Florence et Turin et des régions Emilia Romagne, Ligurie, Lombardie, Toscane, sont très actifs dans le lancement de concours pour inciter le développement d'applications réutilisant les données qu'elles diffusent ainsi que celles d'autres institutions.

7. Analyse comparative de la réutilisation des DP dans les quatre pays

De l'analyse de la réutilisation des DP à partir des plateformes analysées, nous pouvons affirmer que la réutilisation de DP est un processus clairement mais différemment enclenché dans les pays étudiés.

Au niveau national, les catalogues d'applications aux Etats Unis et au Royaume Uni sont de loin les plus riches et les mieux structurés. La plateforme nationale

³⁵ LodLive, <http://en.lodlive.it/>

³⁶ ComunicaPA, <http://www.dati.gov.it/content/comunicapa>

³⁷ Les portails de la commune de Rome, de la région de la Vénétie et dans la région Piémont sont respectivement aux adresses <http://dati.comune.roma.it/applicazioni?page=7>, <http://dati.veneto.it/applicazioni> et <http://www.dati.piemonte.it/catalogodati/esempi-uso.html>

³⁸ Projet Homer, http://www.homerproject.eu/images/downloads/HOMER_Newsletter1.pdf

italienne, démarrée au même temps que l'analogue en France, présente cependant un portefeuille d'applications bien plus vaste, structuré selon les dimensions thématiques, économiques et de couverture géographique.

Aux niveaux des couvertures territoriales plus restreintes, nous interprétons les plateformes numériques des collectivités territoriales en France et en Italie comme des *aiguillons* des initiatives nationales, souvent plus tardives. Les sites anglo-saxons de ces niveaux nous apparaissent complémentaires des sites nationaux.

Les thématiques prioritaires comprennent les transports, la santé et les loisirs. Ceci est sûrement lié au rôle joué par les données géo spatiales, conforté en Europe par la mise en place de la directive INSPIRE. Des thématiques comme la criminalité ou la sécurité, représentées de façon significative dans les plateformes anglo-saxonnes, sont absentes dans les portails étudiés en France et en Italie.

Les plateformes techniques mobiles sont ciblées par les applications des quatre pays et contribuent à créer des services distribués dans tous les secteurs d'activité.

L'ouverture vers le Web sémantique, via les Linked Data, qui facilitent l'interopérabilité dans des contextes élargis, apparaît surtout dans les applications des plateformes anglo-saxonnes, mais aussi dans des applications en Italie.

Les métadonnées décrivant les applications aux USA et en UK sont les plus complètes et détaillées. Elles couvrent de nombreux critères objectifs, mais aussi subjectifs, comme les tags et les appréciations des utilisateurs.

Les principaux leviers de la réutilisation des DP mis en évidence sont :

- les avancées technologiques qui favorisent la production de masses de données, l'intégration et la visualisation d'informations sur plateformes mobiles. La diffusion de services et d'outils de cartographie et de mashup facilitent des développements rapides basés sur l'utilisation de données géo spatiales.

- les gouvernements, qui mènent des actions dans plusieurs axes comme : a) accords de collaborations entre pays, dans des secteurs prioritaires, notamment les politiques publiques territoriales ou l'énergie ; b) allocation de budgets pour harmoniser et diffuser des données institutionnelles sous licences ouvertes, avec leurs métadonnées ; c) foisonnement de concours pour inciter des développeurs à participer de la communauté Open Data.

Les plateformes institutionnelles ne permettent pas d'analyser les demandes des citoyens. Celles-ci s'expriment néanmoins dans les cadres juridiques, au sein des différents pays et de l'Europe.

Exploitation du potentiel des DP. Actuellement, les applications publiées par les différentes plateformes étudiées n'exploitent pas tout le potentiel des données institutionnelles, car elles utilisent très majoritairement un seul jeu de données, plutôt que plusieurs. A plus forte raison, les applications à couverture internationale, envisagées par la législation européenne et attendues par les citoyens, munies d'interfaces multi langues, suivant une classification thématique intégrée et un modèle économique concordé, sont très peu nombreuses.

Les principaux freins se situent selon nous au niveau stratégique (priorité donnée aux projets d'ouverture et de réutilisation des DP, allocation de budgets) et organisationnel, avec la prolifération de catalogues hétérogènes contenant des ressources potentiellement partageables (données et applications). Ces difficultés se combinent et s'amplifient avec les changements d'échelle, lors du passage du niveau local au niveau régional, voire du national à l'international. Elles rendent difficiles aussi bien des recherches automatiques de données que leur réutilisation dans des contextes différents de ceux pour lesquels elles ont été produites.

Les hétérogénéités sémantiques et structurelles des métadonnées (thématisques, temporelles, économiques, d'utilisation) qui décrivent les applications concernent par exemple : a) les catégories de classement, de différents types : thématique (transports, tourisme, santé, ...), territoriale (niveau européen, d'un pays, d'une région, d'un département, d'une commune, ...) ou encore économique, pouvant représenter des aspects d'accessibilité, liés aux licences, ou encore des aspects sectoriels (comme secteur public, privé, tertiaire) ; b) les indicateurs, concernant le référencement sur les réseaux sociaux, les commentaires, le degré de satisfaction ou le nombre de téléchargements. Ces indicateurs, porteurs d'information sur des activités des utilisateurs, ont cependant une faible fiabilité, les commentaires étant souvent anonymes et les téléchargements n'étant pas synonymes d'utilisation.

8. Conclusions et perspectives

La réutilisation de données publiques est un phénomène très récent, complexe, international et en rapide évolution. Cette étude comparative analyse la dynamique de ce phénomène, à partir des plateformes institutionnelles de quatre pays : Etats Unis, Royaume Uni, France et Italie. En utilisant une démarche généralisable à d'autres domaines, nous mettons en évidence des éléments quantitatifs et qualitatifs concernant les différents niveaux de maturité des portails, en termes de couverture territoriale, thématiques prioritaires, plateformes techniques, ouverture vers le Web sémantique et métadonnées. Nous identifions les principaux moteurs qui impulsent cette dynamique, ainsi que des freins stratégiques et organisationnels à une plus forte exploitation des DP dans des contextes élargis.

Nous allons poursuivre l'évaluation de la dynamique de réutilisation des DP, en nous focalisant sur les applications inter-régionales ou internationales au sein de l'Union Européenne. Deux initiatives d'ordre stratégique, "découvertes" lors de cette étude, devraient favoriser une progression dans la réutilisation des DP : la mise en place imminente d'un groupe de travail sur les métadonnées pour l'échange de données parmi les portails de l'UE et l'ouverture très récente du portail européen de *l'Open Data*. Elles devraient aussi permettre de systématiser les étapes de choix des plateformes et d'observation de notre démarche.

Remerciements

Cette recherche a bénéficié d'échanges avec Mme J. Boustany, Chercheur de l'Équipe de recherche DICEN du CNAM de Paris, qui traite le thème des données publiques du point de vue de l'accès aux documents publics et de leur réutilisation.

9. Bibliographie

- Akoka J., Comyn-Wattiau I. « Vers l'ingénierie des évolutions », *Ingénierie des Systèmes d'Information, RSTI, série ISI*, Vol. 14, n° 6/2009, pages 9-17.
- Boustany, J. « La politique d'accès aux documents publics : étude comparative entre les États-Unis, la France et le Royaume-Uni», In E. Broudoux & G. Chartron (Eds.), *Enjeux politiques du document numérique* (pp. 251–276). Paris: ADBS, 2010
- Davies T., « Supporting open data use through active engagement », *W3C Using Open Data Workshop*, June 2012, <http://goo.gl/27OIN>
- Ding L., Lebo T., Erickson J. S., DiFranzo D., Williams G. T. et al., « TWC LOGD: A Portal for Linked Open Government Data Ecosystems », *J. Web Semantics*, vol. 9, no. 3, Elsevier, 2011, pp. 325–333
- Ferro, E., Osella, M., « Modelli di Business nel Riuso dell'Informazione Pubblica », *Regional ICT, Observatory of Piedmont Region*, 2011 (in italiano), <http://goo.gl/4VvpB>
- Heath T., Bizer C., « Linked Data: Evolving the Web into a Global Data Space », *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1:1, 1-136. Morgan & Claypool, 2011
- Martinez R., Marx B. « La longue marche de l'information publique, de la liberté d'accès aux documents administratifs à la réutilisation commerciale des informations publiques », *Documentaliste-Sciences de l'Information* 3/2007 (Vol. 44), p. 218-227.
- Salzano G., « Vers un observatoire de recherche sur les données publiques en santé », *Colloque de l'Association Information et Management (AIM), Vers un Management Ethique et Responsable ? La Contribution des Systèmes d'Information*, 21-23 mai 2012, Bordeaux.
- Shadbolt, N., « From Data to Decisions: The Power of Information in the Age of the Web of Linked Data. » *Royal Signals Institution Annual Seminar*, HQS Wellington, London, 2010.
- Shadbolt N., Kieron O'H., Berners-Lee T., Gibbins N., Glaser H., Wendy H., Schraefel M.C., « Why Open Government Data? Lessons from data. gov. uk. », *IEEE INTELLIGENT SYSTEMS*, 27, Issue, 16 – 24, 2012. Retrieved from <http://eprints.soton.ac.uk/271648/>
- W3C, Report on Using Open Data : policy modeling, citizen empowerment, data journalism, 19 - 20 June 2012, The European Commission's Albert Borschette Conference Center, Brussel, <http://www.w3.org/2012/06/pmod/report>.
- Yin R. K. *Case Study Research. Design and Methods*, Fourth Edition, 240 pages, SAGE Publications, Inc ., 2009