Smart city

Des pratiques imaginées de l'Open Data aux réalités numériques des spécialistes de la donnée

Isabelle Hare

Université Lyon 2 / Elico

Résumé : le sujet de l'Open Data et ses usages a nourri depuis une dizaine d'années maintenant une littérature assez abondante dans les médias généralistes et spécialisés, comme dans des champs scientifiques variés (informatique, sociologie, économie, gestion, marketing, etc.). Dans le cadre particulier des *smart cities*, ce chapitre analyse et confronte les discours scientifiques sur l'Open Data aux pratiques et représentations des professionnels de la donnée, en appuyant une partie de ses observations sur deux expérimentations menées dans le cadre de l'ANR OpenSensingCity. Plus largement, dans quelles mesures les discours d'escorte et les propositions d'outillages scientifiques de l'open data font-ils écho aux préoccupations et aux besoins des professionnels de la donnée numérique ?

Mots-clés : Open Data, Big Data, pratiques professionnelles, discours scientifiques, outils, représentations, *smart city*.

1 Introduction

Open Data, Big Data, *smart cities*, *data cities*... ces termes recouvrent des désirs et des réalités politiques, scientifiques, médiatiques et citoyennes bien différentes. Ce chapitre étudie les discours sur les pratiques professionnelles et les représentations autour de l'usage des données numériques en général et de l'Open Data en particulier, sans perdre de vue qu'aujourd'hui la thématique des *smart cities* agrège une bonne partie d'entre eux.

La littérature sur les *smart cities* est abondante et ses définitions également. Rob Kitchin en propose une synthèse intéressante dans son article « The real-time city ? Big Data and smart urbanism » (2014 a) au sein duquel il souligne très clairement l'arrière-plan libéral qui régit la gouvernance et le développement des *smart cities*, guidés par

les intérêts économiques (« *market-led* ») et les innovations technologiques. Les données (numériques) participent pleinement à la réalisation de ces villes intelligentes, connectées et proposant deux idéaux-types : à l'Open Data, la cité « *ascendante, distribuée et collaborative* » et au Big Data, la « *ville centralisée et panoptique* » (Peugeot, 2014 : 53). Valérie Peugeot observe avec justesse qu'opposer ces deux modèles n'illustre guère la réalité des *smart cities*, qui se construisent majoritairement autour de l'hybridation de ceux-ci.

Ces douze dernières années, Big et Open Data ont bousculé les imaginaires informatiques, économiques et démocratiques, générant les fantasmes les plus ravis (Open Data et citoyenneté éclairée) comme les plus inquiets (Big Data et protection de la vie privée) au sein de l'espace public.

Au-delà de ces considérations sur les *smart cities* et la gestion des données numériques (1) ce chapitre interroge la pertinence des attendus scientifiques et des discours projetés sur l'Open Data et ses usages (2), pour mesurer leur adéquation avec les pratiques et les représentations des professionnels de la donnée (3). Cette étude devrait nous permettre de donner quelques éléments de réponses aux questions suivantes : quelles sont les représentations de l'Open Data pour chacun de ces acteurs (scientifiques et praticiens) ? Au-delà des discours, les pratiques outillées permettent-elles de trouver des points de convergences, des passerelles informationnelles et technologiques entre ces différents acteurs de la donnée ?

Ce chapitre tente d'appréhender la correspondance entre les discours d'escorte et les propositions d'outillage(s) scientifiques de l'Open Data d'une part, et les préoccupations et besoins des professionnels de la donnée numérique d'autre part. Il y aurait, selon nous, et c'est ce qui fait problème ici, une forte inadéquation entre le monde projeté de l'Open Data, souvent imaginé au chevet de la *smart city* aux côtés du Big Data, et la réalité des pratiques professionnelles de la donnée.

Ce développement prendra notamment appui sur deux expérimentations menées dans le cadre de l'ANR OpenSensingCity [1], réalisées auprès de spécialistes de la donnée à quelques mois d'intervalle en mars et en décembre 2017. Nous avons conscience qu'on ne peut faire de généralités à partir des seuls entretiens menés suite à ces deux tests, il faudrait pour cela étendre le terrain d'étude à d'autres acteurs de la donnée. Néanmoins, ce premier matériau d'enquête, bien qu'imparfait et incomplet, nous semble constituer une base à la formulation d'hypothèses et de pistes d'analyse qui devront être ré-interrogées dans des travaux ultérieurs. Au-delà, ces réflexions s'appuient également sur un travail en cours de publication qui traite de l'impensé [2] de l'Open Data et a donné lieu à d'autres séries d'entretiens (Hare, à paraître).

1. Ces deux expérimentations, que nous détaillerons plus loin, ont donné lieu à un rapport de recherche dans le cadre de l'ANR : Dymytrova V., Hare I., Peyrelong M.-F., 2018, *Synthèse des résultats des expérimentations des outils OpenSensingCity au Tuba et à la Scop la Péniche*, Livrable n° 4, EA 4147 Elico. Accès : https ://hal.archives-ouvertes.fr/hal-01823479.

2. Ce chapitre fait partie du deuxième tome sur *L'impensé numérique*, dirigé par Pascal Robert (à paraître).

2 Le monde imaginé de la *smart city* au prisme l'Open et du Big Data

Nous l'évoquions supra, *smart city* et Data (Big et Open) sont étroitement liées, les discours et les initiatives politiques véhiculent cet imaginaire de la ville intelligente (Peugeot, 2014), qui voient dans la récolte et l'analyse des données numériques la possibilité d'une meilleure connaissance et d'une gouvernance plus efficiente du territoire urbain (Kitchin, 2014a).

Au-delà, le développement de ces *data driven city* a généré, comme le précise V. Peugeot (2014) un rapport au territoire et à l'alter-urbanité fondé sur la compétitivité et la concurrence.

2.1 *Smart cities* : du Big à l'Open Data

Open et Big Data sont régulièrement associées à la *smart city,* alors qu'ils s'organisent et se développent selon des logiques différentes. En premier lieu, comme l'explique Samuel Goëta : « *L'Open Data désigne la diffusion volontaire et proactive de données, essentiellement publiques, qui deviennent librement réutilisables alors que le Big Data regroupe des pratiques d'exploitation de données massives, à grande vitesse et en temps réel* » (2015 : 26). Ensuite, les politiques et les discours autour de l'Open Data se concentrent autour du triptyque transparence-innovation-modernisation, alors que le Big Data s'inscrit davantage dans une approche économique de la donnée, autour des notions de connaissance-gestionnarisation-performance (Hare, à paraître).

Enfin, l'Open Data est une donnée partielle (via les jeux de données) distribuée en accès ouvert dont la réutilisation est encouragée, alors que le Big Data est récolté massivement, le plus souvent en circuit fermé, ne bénéficiant souvent qu'à l'organisation qui les moissonne.

Open et Big Data affichent donc clairement leurs différences dès qu'on les confronte ; cependant le « *discours commun* » (Goëta, 2015 : 26) les associe fréquemment à la *smart city.* Dans les faits, les données urbaines sont captées en gros volume, destinées à être analysées et retraitées dans des formats propriétaires, et produisent donc rarement de l'Open Data.

Pourquoi cet amalgame ? Très probablement parce que leurs destinées se sont croisées dès 2007-2008 (Goëta, 2015) et qu'elles ont animé les discussions politiques au niveau national comme local, faisant se télescoper l'injonction étatique à ouvrir les données publiques et la nécessité territoriale à générer des Big Data. Dans des temporalités très proches, nous avons donc vu émerger au sein des collectivités les portails Open Data (Labelle, Le Corf, 2012 ; Dymytrova et Paquienséguy, 2017) et fleurir les outils de captation de données en temps réel et les API (*Application Programming Interface*), caractéristiques du Big Data urbain (par exemple, Lyon VeloV qui retranscrit les informations sur les stations de vélos en libre-service à Lyon). En 2019, l'Open Data participe matériellement encore assez peu à l'intelligence de la ville, alors que les discours politiques l'y associent très régulièrement. Dans les faits, il contribue bien davantage à repenser le dialogue entre le citoyen, l'action de l'Etat et les collectivités.

Il y aurait donc une forme de paradoxe à penser la *smart city* au prisme de l'Open Data et une sorte d'injonction politique à le faire (Le Corf, 2015), alors que ce dernier se situe – pour l'instant – à la périphérie de la ville intelligente, enjoignant davantage à une citoyenneté éclairée et vigilante (Hare, Le Béchec, 2015). C'est néanmoins dans cet interstice qu'a choisi de se glisser l'ANR OpenSensingCity en visant à une meilleure compréhension des données ouvertes urbaines et à faciliter leurs usages auprès de la catégorie des développeurs d'API, au bénéfice du citoyen.

Les données numériques permettent aujourd'hui de prendre le pouls du territoire urbain, c'est une certitude : la ville intelligente s'appréhende à travers la donnée. Mais elle participe à produire bien d'autres choses comme l'explique notamment R. Kitchin (2014 a) et V. Peugeot (2014) avec quelques différences néanmoins. Pour R. Kitchin, la notion de *smart city* convoque une double approche : d'un côté, une ville conduite par l'innovation, la créativité, l'entreprenariat motivée par la connaissance économique d'un territoire ; de l'autre, une intelligence urbaine fondée sur la mesure très outillée des représentations et modélisations du territoire, qui permet une meilleure prédictibilité des événements à venir, et l'efficacité dans la planification des actions par un *monitoring* accru.

V. Peugeot offre une vision plus sociale de la *smart city* dans la mesure où les deux conceptions qu'elle oppose (voir supra) mettent la communication au cœur du processus des villes intelligentes, que ce soit par le contrôle de celles-ci (regard panoptique) ou leur caractère participatif (visée collaborative). Et c'est précisément cette « *ville ascendante, distribuée et collaborative* » (Peugeot, 2014 : 49) qui nous intéresse ici, puisqu'elle propose des « *lieux de partage et de co-construction des ressources* » (Peugeot, 2014 : 51), qui nous semblent proches des idéaux de l'Open Data.

Centralisée ou distribuée, la ville met la donnée numérique au cœur de l'intellection du territoire urbain que ses ambitions soient économiques, politiques ou illustrées à travers « *le pouvoir collectif des communautés* » (Peugeot, 2014 : 54). Mais il nous semble qu'il y a dans ces nombreux « *discours communs* » sur les *smart cities* – pour reprendre l'expression de S. Goëta (2015 : 2), une forme d'inexactitude qui conduit à un anachronisme : ce n'est pas la *smart city* qui produit le foisonnement des données numériques, mais c'est bien la prolifération des data qui a permis la naissance puis la croissance de ces villes intelligentes.

C'est d'ailleurs ce qu'explique R. Kitchin lorsqu'il évoque l'importance des data : « *Another vital conjoin between these two visions of a Smart City is the prioritization of data capture and analysis as a means for underpinning evidence informed policy development, enacting new modes of technocratic governance, empowering citizens through open, transparent information, and stimulating economic innovation and growth.* » (2014 a : 2).

Sans verser dans le déterminisme technologique, c'est bien la production et la captation quasi-exhaustive de données numériques en gros volumes émanant du secteur privé [3] (Goëta, 2015), qui a instauré un climat propice à la création d'analyses outillées

3. S. Goeta explique dans ce rapport que de grandes entreprises telles que IBM ou SAS ont fait du Big Data un slogan pour vendre leurs solutions informatiques.

et favorisé l'émergence d'une pensée politique (Peugeot, 2014 : 47), fondée sur l'auscultation minutieuse de la ville : « *Forgées par quelques acteurs majeurs des TIC, IBM en tête, mais aussi CISCO, Siemens, Schneider, Hitachi. . . , elles ont depuis essaimé à la fois dans différents milieux industriels, mais aussi dans le monde politique* ».

Pourquoi un tel engouement pour les data urbaines ? Selon R. Kitchin (2014a), elles sont perçues comme l'expression objective et neutre de ce qu'il se passe au sein de la cité, la mesure empirique qui délivre une vérité sociale et économique (une sorte d'instantané photographique), à partir de la laquelle les élus vont pouvoir élaborer leurs stratégies de développement. R. Kitchin (2014 a : 3) analyse cette situation en citant le travail de Mayer-Schonberger and Cukier (2013) : « *Such data are seen as providing objective, neutral measures that are free of political ideology as to what is occurring in a city, with the weight of data speaking an inherent truth about social and economic relations and thus providing robust empirical evidence for policy and practice.* » Ces formes automatisées de production de données participent certes à une meilleure appréhension de la ville en temps réel pour les citoyens comme pour les administrateurs, néanmoins, la prégnance grandissante de ces data dans la gouvernance urbaine fait notamment craindre la mainmise du secteur privé (pour la captation, le retraitement et l'analyse des données) dans la gestion quotidienne de la cité et du citoyen (Greenfield, 2013 ; Townsend, 2013).

Et c'est précisément à cet endroit que les données issues de l'Open Data ont un rôle à jouer au sein de la *smart city*, dans la mesure où non seulement elles sont en accès libre mais encouragent au partage et à la redistribution, participant ainsi à une forme beaucoup plus collaborative de la vie en commun et donc potentiellement à une urbanité partagée.

2.2 Les discours autour de l'Open Data urbain

Ces discours, qu'ils émanent de la société civile (Hare, 2016) ou de la sphère politique (Hare, à paraître) produisent une représentation idéalisée de l'Open Data, qui s'explique en partie par le positivisme latent qui accompagne le Big Data depuis le milieu des années 2000 (Kitchin, 2013 ; Ribes, Jackson, 2013). La croyance en la donnée comme facteur de développement économique et social n'est certes pas nouvelle (Goëta, 2015), la promotion de l'Open Data comme co-promoteur de la *smart city* (au côté du Big Data) nous semble l'être bien davantage.

En effet, les discours communs sur les *smart cities* semblent avoir progressivement fait glisser dans un même panier communicationnel le Big et l'Open Data, alors que le premier s'origine dans l'idée de performance et le second dans celle de transparence (*voir supra*).

Certes, dès 2011, la mission Etalab promeut la libération progressive de l'ensemble des données publiques avec, pour dessein, une plus grande transparence au sein des administrations françaises et, sous-jacente, la promesse d'un Open Data promoteur d'innovations (technologiques, sociales, . . .), susceptibles de constituer un important levier économique sur le territoire français. Selon nous, c'est dans ce second volet que le glissement sémantique va se faire progressivement de l'Open au Big Data dans l'imaginaire projeté de la *smart city*.

Sarah Labelle et Jean-Baptiste Le Corf (2012 : 64) observent que, *via* les portails Open Data, « *La collectivité, en proposant l'accès à ses données, autorise une nouvelle forme de relations avec les citoyens* ». Or, si une ville connectée en temps réel (places libres dans les parkings, horaires et retards des transports en commun, *etc.*) facilite indéniablement la déambulation du citoyen et son appréhension du territoire urbain, il faut néanmoins noter que beaucoup de ces données – accessibles le plus souvent *via* des API [4] – émanent davantage des Big Data que de l'Open Data. Nous verrons dans ce chapitre que les formats hétérogènes des données issues de l'Open Data ne facilitent pas leur manipulation par les développeurs. De fait, l'accès direct du citoyen à la *data collectivité* (l'accès à la collectivité par les données publiques et/ou privées) reste rare, et s'avère le plus souvent médiaté par le travail de développeurs et d'entreprises privées (Turki, Falonneau, 2015).

Nous avons par ailleurs observé (Hare, à paraître) que la donnée, une fois mise en jeu sur les portails Open Data des métropoles, perdait de sa substance politique (incarnation de la transparence) pour devenir « *un cadre d'action qui définit les conditions d'existence du projet 'Open Data'* » (Labelle, Le Corf, 2012 : 64). Cet effacement du politique au profit d'une dynamique pragmatique et programmatique (Hare, à paraître), faciliterait un rapprochement symbolique entre Open et Big Data sur le terrain des *smart cities*, représentations présentes notamment dans certains discours d'élus de collectivités locales (Hare, à paraître). Cependant, seuls le volume, la vélocité et la variété du Big Data permettent un traitement des données urbaines en temps réel ; l'Open Data est, quant à lui, libéré par les collectivités selon une temporalité beaucoup plus aléatoire et dans des formats hétérogènes (Larroche, Dymytrova, 2017). Le développement des portails métropolitains depuis 2011 [5] tend néanmoins à rapprocher l'Open Data de la *smart city* par la thématique du territoire et de l'espace urbain. Ainsi, Valentyna Dymytrova et Françoise Paquienséguy montrent-elles que le portail de la métropole lyonnaise propose deux lignes d'objectifs :

> D'une part, celle de l'innovation, de la création de services, porteuse de valeur ajoutée et de développement économique du territoire, pensée aux mains des professionnels de la data, du territoire, des acteurs économiques locaux ; de l'autre, celle de la citoyenneté et du partage collaboratif, porteuse de convivialité, de liens sociaux, proche d'une urbanité numérique qui rendrait la ville plus ou mieux habitable. (Dymytrova, Paquienséguy, 2017 : 117)

Nous retrouvons ici illustrées les deux ontologies de la *smart city* présentées par V. Peugeot (2014) et c'est donc bien par les territoires géographiques, politiques et sociaux que se tissent le lien entre *smart city* et Open Data.

Jusqu'où ces données peuvent-elles et doivent-elles être partagées ? Car il y a quand même un monde numérique entre les jeux de données ouvertes qu'offrent les municipalités et les flux continus de données des capteurs urbains souvent exploités par des entreprises privées et revendus aux communes. A qui profitent les Big Data ?

4. Nous faisons référence ici à des API comme TCL à Lyon ou encore Moovit à Nantes, ou encore au service de cartographie (en Open Data) proposé par OpenStreetMap

5. Rennes a été précurseur, puisqu'elle a créé son portail Open Data en 2009.

Aux acteurs privés et aux collectivités assurément, aux citoyens c'est peu probable. N'y a-t'il donc pas là « *une forme d'enclosure de l'information urbaine* » (Peugeot, 2014), qui ancrerait la *smart city* dans une sorte de *dataïsation* à deux vitesses : au secteur privé et aux collectivités, une donnée vive, volumineuse, enrichie et monnayée pour une gestionnarisation contrôlée de la ville ; aux citoyens, une donnée ouverte déconnectée du temps réel dont la fonction première serait un accès plus transparent à la gestion du territoire ainsi que le rétablissement de la confiance citoyenne dans ses institutions.

3 Le monde projeté de l'Open Data dans l'outillage scientifique de la donnée

Cette seconde partie interroge la façon dont les imaginaires qui traversent l'Open Data (évoqués dans la section précédente) sont en partie incarnés dans la volonté d'outiller les médiateurs des données ouvertes que sont par exemple les développeurs.

3.1 Des données brutes et hétérogènes

Nous avons mentionné supra que les données ouvertes sur les portails Open Data, l'intérêt de leur contenu et les formats proposés, les rendaient peu accessibles aux citoyens. J.-B. Le Corf (2015) affirme à ce propos qu'il y aurait une sorte de contradiction entre l'incitation faite aux citoyens d'utiliser les données ouvertes et les formats proposés qui s'adressent essentiellement aux développeurs. Selon lui, « *la communication publique de l'administration* via *les portails relative à la mise à disposition de nouveaux jeux de donnée (servant une communication politique) permet autant de montrer le travail de l'administration et de valoriser l'action politique de la ville*, *que de faire connaître ces nouvelles données disponibles (objectif relevant de la communication publique)* » (Le Corf, 2015 : 104-105). L'annonce de nouveaux jeux de données – ou leur dernière mise à jour, caractériserait davantage le volontarisme politique et le fait d'être dans le coup de la smart data, et s'adresserait donc à l'ensemble de la collectivité urbaine.

Cette publicisation de l'action politique en faveur du citoyen se double de la volonté parfois contradictoire de vouloir séduire le public des développeurs, parfois réfractaire à utiliser les données libérées sur les portails Open Data soit parce que les formats sont hétérogènes, soit parce qu'ils ne sont pas assez bruts (Denis, Goëta, 2016). Comme le précise les S. Goëta et J. Denis, ce caractère « brut » des données ouvertes apparaît dès sa définition originelle [6] à l'occasion de la réunion de Sébastopol en 2007 : « *Data Must Be Primary : Data are published as collected at the source, with the finest possible level of granularity, not in aggregate or modified forms.* »

Ce point est extrêmement important, car il préfigure l'une des caractéristiques mais aussi l'une des ambiguïtés des données Open Data car toute donnée libérée sur un portail nécessite en amont de sa diffusion, un travail, de reconfiguration, ne serait-ce que par la sélection et le nettoyage des jeux de données (Denis, Goëta, 2016). Il y

6. Nous nous référons ici aux huit principes de l' Open Data, définis suite à la réunion de Sébastopol. Ainsi la donnée ouverte doit être : « complete, primary, timely, accessible, machine processable, non-discriminatory, non-proprietary, licence-free ». Disponible depuis : https ://www.opengovdata.org.

aurait là une double injonction technicienne à produire des données « *brutifiées* » : d'une part, parce que l'Open Data s'origine dans les univers croisés du logiciel libre et de l'*open access* et que cette « *brutification* » (Denis, Goëta, 2016) est reprise dans les textes législatifs qui encadre l'Open Data [7]. D'autre part, parce que les publics visés (développeurs) par les éditeurs de portails Open Data attendraient des formats *premiers* afin de pouvoir les éditorialiser à leur guise. A défaut de pouvoir être véritablement « *brutes* » (Denis, Goëta, 2016), les données partagées sur les portails Open Data nécessitent une importante mise en œuvre pour les rendre intelligibles (par les métadonnées notamment mais aussi par des manipulations en amont). Selon J. Denis et S. Goëta (2016 : 26), il faut alors « *appréhender le travail d'ouverture des données comme un travail de 'brutification'* ».

Le problème résiderait principalement dans le fait que les politiques Open Data, au-delà de l'ouverture des données publiques, réclament de « *transfigurer des données aux usages et aux significations étroites, ciblées, en données universelles* » (Denis, Goëta, 2016 : 24). Cette volonté d'universalisation de la donnée ouverte constitue un problème technique de taille, puisque des univers professionnels différents (développeurs, *data scientists*, *etc.*) se confrontent ici, des langages et des formats dont l'utilisation dépasse largement le cadre purement technicien et renvoie à l'adhésion de son utilisateur à telle ou telle communauté informatique. Nous faisons référence ici à plusieurs témoignages recueillis lors d'entretiens auprès de scientifiques, dans le cadre de l'ANR OpenSensingCity, expliquant que telle communauté scientifique et professionnelle spécialiste du Web sémantique va utiliser de façon privilégiée un certain format (SPARQL ou RDF par exemple) alors que celui-ci est méconnu des industriels pour lesquels travaillent la communauté des développeurs.

3.2 La reconfiguration outillée des données ouvertes : des expériences riches d'enseignement sur les discours autour des pratiques de l'Open Data

Toute la complexité du travail à partir de données Open Data réside dans cette volonté d'universalisation, tant celles-ci peuvent-être hétérogènes, incomplètes et fluctuantes. Dans le cadre de l' ANR OpenSensingCity, des chercheurs de l'EMSE [8] -spécialistes du Web sémantique- ont essayé de résoudre ce problème, en collaboration avec des partenaires industriels (Antidot et HIKOB), en mettant en place des outils informatiques susceptibles de réduire cette hétérogénéité sur le terrain des *smart cities*.

Ces outils ont été testés à l'occasion de deux expérimentations réalisées en 2017 et les analyses en résultant ont été réalisées à partir de la chaîne de traitement des données identifiée dans le livrable n° 3 de l'ANR OpenSensingCity. Nous avions alors observé qu'« *au-delà des caractéristiques professionnelles, tous les réutilisateurs des données ouvertes (étaient) confrontés à la chaîne de traitement comprenant la collecte*

7. Pour l'Europe : la Directive PSI 2003/98/EC du Parlement européen et du Conseil du 17 novembre 2003 vise la réutilisation des informations du secteur public, qui sera ensuite reprécisée dans la Directive 2013/37/UE du 26 juin 2013, concernant les domaines d'exercice des données libérées et leur facturation potentielle. La Directive INSPIRE 2007/2/CE du Parlement européen et du Conseil du 14 mars 2007 a participé à l'harmonisation des données géographiques européennes et à leur ouverture dans chacun des états membres, dont la France. La loi Macron, les lois Walter et NOTRe en 2015, puis la loi pour une République, dite Loi Lemaire en 2016, encadrent en France l'ouverture et l'utilisation des données.

8. Ecole des Mines de Saint-Etienne, Institut Henri Fayol.

et l'exploration, la compréhension et l'analyse des données, la transformation et enfin, l'exploitation ou l'implémentation des données à travers un développement ou une modélisation » (Dymytrova V., Hare I., Peyrelong M.-F., 2018 : 15). Suite à ces deux événements, des entretiens ont été menés par les chercheuses d'ELICO [9] avec certains des participants aux expérimentations, prenant justement en compte cette chaîne de traitement.

3.2.1 Dialogues enrichis autour de la donnée ouverte et des formats (TUBÀ)

La première expérience s'est déroulée le 17 mars 2017, dans les locaux de TUBÀ, le *living lab* de la Métropole de Lyon. Elle avait pour objectif premier la confrontation des résultats à l'issue de l'expérimentation ; il s'agissait également de faire dialoguer autour des outils et de leurs usages les différents acteurs présents [10] :

> 1/ quant à la pertinence de scénarios d'usages de réutilisation des données ouvertes par des professionnels de la donnée ou d'applications ; 2/ quant aux problèmes rencontrés, ou résolus, que ceux-ci rencontrent en situation de réutilisation de données ouvertes ; 3/ quant aux outils OpenSensingCity développés par l'EMSE. [...] L'hypothèse principale de la journée consistait à vérifier si l'usage d'un modèle commun de présentation de données ouvertes (RDF) et la possibilité d'un unique point d'accès aux données (portail Antidot OSC Search Box [11]) faciliteraient l'exploitation des données ouvertes analogues issues de différents territoires. (Dymytrova, Hare, Peyrelong, 2018 : 4-5)

Les participants avaient pour consigne de travailler à partir de quatre types de données (météo, disponibilités des parkings, pollution, disponibilités des vélos) provenant des portails Open Data de cinq métropoles françaises : Paris, Lyon, Strasbourg, Nantes et Rennes. Leur exploitation devait permettre de mettre en place une application ou une data visualisation afin d'aider au stationnement en milieu urbain [12].

Si nous reprenons les différentes étapes de la chaîne de traitement, nous observons plusieurs phénomènes : au moment de la recherche et de la récupération des données, le format (JSON-LD) mis à disposition grâce à l'outil proposé par l'EMSE (SPARQL Generate) et les données Open Data présentes agrégées en modèle RDF sur le portail d'Andidot ont beaucoup gêné certains participants (*data analysts* et *data scientists*) peu habitués à manipuler ces formats, préférant les packages disponibles dans des bibliothèques logicielles. Suite à cette première difficulté, ils ont pour la plupart choisi une stratégie de contournement au moment de la phase de transformation, soit en convertissant la donnée initiale dans un format familier [13], soit en allant récupérer les données directement sur les portails des agglomérations. Seul le public des développeurs semblait à l'aise avec le format JSON-LD. Ces premières

9. Equipe de recherche lyonnaise en Sciences de l'Information et de la Communication (EA4147).

10. 11 personnes étaient présentes, outre les chercheurs en informatique et en sciences humaines : développeurs, *data scientists et data anaylists.*

11. Portail d'indexation de jeux de données ouvertes crée dans le cadre des deux expérimentations.

12. Pour plus de détail quant au déroulé de l'expérimentation, nous renvoyons au Livrable n° 4 de l'ANR.

13. Ils ont donc passé beaucoup de temps à transformer les données dans les formats compatibles avec leurs solutions professionnelles (logiciel R par exemple).

observations témoignent du caractère rémanent des *habitus* professionnels, dès qu'il s'agit de manipuler des formats inhabituels à son univers technologique et informatique. Les langages du Web sémantique sont très spécifiques et nécessitent pour être compris et efficients de passer du temps sur leur modèle de structuration, ce que n'ont pas nécessairement ces professionnels de la donnée qui doivent répondre rapidement aux demandes des clients. Néanmoins, certains des participants à l'expérimentation (les développeurs) ont reconnu l'utilité de formats (comme JSON-LD) et du modèle RDF qui, s'ils devenaient standards sur l'ensemble des portails Open Data, permettrait de relier les données entre elles et de produire des applications transversales à l'ensemble des territoires. Cette vision d'une donnée enrichie car devenue homogène et exploitable *universellement* renvoie ici très fortement aux communautés de l'*open source* et du logiciel libre.

L'expérimentation au TUBÀ a permis de constater que la formation et les métiers des participants tiennent une place prépondérante dans la manière dont les outils sont manipulés et les données appréhendées : « *Les uns ont en effet privilégié la satisfaction rapide de la commande (l'application) quitte à revenir ensuite sur les données, alors que d'autres ont choisi d'abord d'étudier l'architecture des outils proposés et des données pour ensuite pouvoir les enrichir de façon plus efficace en amont même de la création d'une application* » (Dymytrova, Hare, Peyrelong, 2018 : 26). L'utilisation de données ouvertes n'est jamais apparue comme un frein à la réalisation de l'exercice ; en revanche, leur accessibilité sur le portail Antidot, ainsi que la spécificité du format JSN-LD et du langage RDF ont posé problème au moment de la récupération et de la transformation de ces données, générant une importante glose autour des standards informatiques. Néanmoins, les discussions initiées au moment de l'expérimentation au TUBÀ et dans les entretiens qui ont suivi, ont montré une certaine curiosité de ce public pour ces formats et pour des outils issus du Web sémantique, bien qu'exogènes à leurs pratiques professionnelles. Cependant, pour que cette curiosité se transforme en usage, il faudrait que les outils-métiers qu'utilisent quotidiennement ces professionnels de la data intègrent ces nouveaux formats et surtout y voient un intérêt mercatique à le faire.

3.2.2 Lorsque l'outil passe à côté de son public (la Scop la Péniche)

La seconde expérimentation a eu lieu le 13 décembre 2017 à Grenoble, en partenariat avec la Scop la Péniche. Il s'agissait cette fois-ci pour les chercheur.se.s d'observer si les outils développés pour l'événement au TUBÀ – produits cette fois-ci dans des versions augmentées – facilitaient l'exploitation de données ouvertes urbaines issues de multiples portails. La commande n'était plus celle d'une application Web, mais d'une extension des fonctionnalités d'une application déjà existante, Parking-Loader :

> Les chercheurs de l'EMSE ont conçu une application Web Parking-loader qui permet de trouver des parkings proches avec leur disponibilité en nombre de places libres, à la manière de l'application Gre.park. En revanche, si l'application Gre.park ne fonctionne que pour les parkings de Grenoble, l'application proposée peut s'adapter à tout territoire dès lors qu'un jeu de données ouvertes est identifiable localement et est chargé

> dans le bon format. En effet, l'application fonctionne uniquement avec les données en RDF (JSON-LD). L'application permet de croiser les jeux de données statiques (adresse parking, information sur le parking) avec les jeux de données dynamiques (disponibilité en temps réel). (Dymytrova, Hare, Peyrelong, 2018 : 12)

Sept personnes aux profils variés ont participé à cette seconde journée : deux ingénieurs R&D spécialisés en édition de logiciels, un ingénieur de recherche, un chargé de mission « Données ouvertes », un développeur et deux étudiants en informatique à l'université.

Comme précédemment, selon un scénario prédéfini et munis d'une abondante documentation sur l'utilisation des outils Antidot Search Box, de SPARQL-Generate et sur le Web sémantique, les participants répondaient à une commande dans un temps limité : « *Le scénario proposé supposait la recherche des jeux de données* via *le moteur de recherche de l'outil Antidot et le service Web parking et trafic de la société HIKOB, l'exploration et la transformation des données dans un format pivot grâce à SPARQL-Generate et l'augmentation des fonctionnalités d'une application mise à disposition* » (Dymytrova, Hare, Peyrelong, 2018 :11).

Chacun des trois groupes devait suivre la chaîne de traitement des données [14] : collecte et analyse, transformation et implémentation, en essayant de croiser et de réconcilier des données ouvertes hétérogènes (statiques comme les noms de rue et dynamiques comme les places de parking en temps réel). Au moment de la restitution de leur travail à la fin de la journée, et comme au TUBÀ, ils devaient préciser leur mode opératoire et les éventuels obstacles rencontrés.

Contrairement à la première expérimentation, les participants ont consacré un temps restreint à la récupération des données (outil Antidot), ne choisissant délibérément que quelques jeux de données pour ensuite essayer de les transformer (de les réconcilier en quelque sorte) avec l'outil SPARQL-Generate. D'autres, comme au TUBÀ, ont préféré s'affranchir de l'outillage proposé et aller piocher directement les données sur les portails Open Data des métropoles. Point intéressant, au moment des entretiens post-expérimentation, le chargé de mission données ouvertes au sein d'une métropole a certes souligné l'intérêt d'un portail qui localiserait les données en format ouvert, mais a semblé craindre la somme de travail nécessaire pour arriver à ce genre de standardisation (de lieu et de format). Cette remarque renvoie à ce que souligne J. Denis et S. Goëta (2016) lorsqu'ils évoquent le long travail de mise en configuration des données ouvertes sur les portails Open Data.

L'étape de la transformation des data a été plus longue à la Scop la Péniche qu'au TUBÀ, car chaque groupe, tutoriel en main, a essayé de comprendre puis de transformer les données qu'ils avaient. A la différence de l'expérimentation au TUBÀ, aucun d'eux n'a contourné la difficulté en faisant migrer les données vers des formats connus et donc maîtrisés. Enfin, aucun groupe n'est arrivé au stade de l'implémentation des données réconciliées pour l'extension de l'application Parking Loader.

14. Pour plus de précisions sur cette deuxième expérimentation, nous renvoyons également au livrable n° 4. (Rapport de recherche ANR OpenSensingCity 14-CE24-0029). EA 4147 Elico. Repéré sur : https ://hal.archives-ouvertes.fr/hal-01823479

Le résultat mitigé de l'expérimentation de la Scop la Péniche – puisqu'aucun participant n'est arrivé au bout du scénario et qu'un seul outil sur les trois proposés a été véritablement utilisé – amène à trois constats au moins : tout d'abord, exploiter des données en Open Data et multiplier les lieux de récolte de celles-ci sur différents portails métropolitains semblent constituer un facteur d'entropie important pour certains professionnels de la donnée. Ensuite, manipuler de nouveaux outils présentant des langages informatiques non familiers et/ou non standardisés, induit parfois des attitudes d'enfermement [15].

Enfin, travailler les données (Open Data ou pas) selon le formalisme du Web sémantique nécessite un apprentissage long, souvent peu compatible avec certains impératifs métiers au sein desquels il s'agit de produire rapidement des applications à partir de données déjà réconciliées et nettoyées.

3.3 Outiller pour de micro-univers professionnels ?

Ces différentes expériences ont permis *in fine* la mise en perspective d'outils facilitant la réutilisation de l'Open Data, dont le but premier réside dans une forme de standardisation et d'homogénéisation des jeux de données et des métadonnées les accompagnant (plateforme proposée par la société Antidot). La seconde promesse, celle de SPARQL-Generate, présente une même structuration et un même format des données issues des différents métiers (Dymytrova, Hare, Peyrelong, 2018 : 28). Au-delà, cet outil promeut également la philosophie du Web sémantique qui contribue à lier et structurer l'information sur Internet.

Ces deux outils participent donc pleinement à la mise en conformité avec les idéaux d'universalité et de mise en commun de l'Open Data mais aussi plus largement du Web des données.

Suite aux deux expérimentations, il s'est avéré que peu de professionnels des data sont familiers du Web sémantique, ce qui constitue un obstacle important au développement de tels outils. De leurs propres avis, le temps d'acculturation à ce formalisme particulier n'est pas compatible avec l'injonction professionnelle de produire des solutions rapides et rentables ni pour l'entreprise qui les emploie, ni pour leur client. Par ailleurs, la question de l'accessibilité du format Json-LD se pose également et demeure problématique pour certains participants en raison de son incompatibilité avec des solutions professionnelles existantes.

Un principe de réalité professionnelle (être productif en utilisant les standards clivant d'un métier) entre donc en collision avec la mise en place d'outillage issus d'autres univers professionnels, ancrés autour de la sémantisation de la donnée et de son universalisabilité. Dans un entretien post-expérimentation, un *data scientist* (janvier 2018) souligne ainsi de façon pertinente : « *Ce qui me préoccupe c'est d'arriver à la généralité de cet outil... SPARQL a déjà une petite communauté, mais pas comme SQL qui est plus diffusé et populaire. Le risque c'est d'arriver à rendre la donnée générique, universelle mais par un outil qui n'est ni générique ni universel* ».

15. Nous employons le terme d'enfermement car, pendant la deuxième expérimentation, nous avons observé que certains groupes s'enfermaient dans la documentation proposée en oubliant la commande initiale.

La généricité de l'outil semble donc prendre le pas sur la sémantisation de la donnée en amont même de la réutilisation ; au-delà de l'efficience de l'outil, c'est sa capacité à réunir une communauté d'usages, et donc à intégrer des standards qui prime. Faut-il donc outiller envers et contre tous l'Open Data ? C'est à cette question que se propose de répondre la dernière partie de ce chapitre.

4 Le monde réel des praticiens de la donnée confronté à l'idéologie de micro-univers scientifique(s)

Cette dernière section vise à confronter deux freins importants à la généralisation de l'utilisation de l'Open Data par les professionnels de la donnée. En effet, suite aux différentes expérimentations et entretiens menés collectivement ou individuellement, nous avons observé qu'il y avait une sorte de consensus implicite autour des raisons de la frilosité des acteurs interrogés vis à vis de l'Open Data. Ainsi, selon eux, plusieurs facteurs seraient susceptibles de ralentir l'expansion de l'usage de l'Open Data *:* l'hétérogénéité de ses formats, l'irrégularité de ses mises à jour sur les plateformes, son caractère anarchique (chaque diffuseur décide de mettre sur les portails les données qu'il souhaite), leur incomplétude, *etc.* Une fois mises sur les portails Open Data, ces données ne sont ni réconciliées ni nettoyées, puisque que justement elles doivent apparaître comme « *brutes* » (Denis, Goëta, 2016). Or, la plupart des professionnels de la donnée ont l'habitude de travailler à partir de formats homogènes, standards et de données complètes.

Il y aurait donc là, une première aporie autour de l'usage de l'Open Data. En effet, seuls les développeurs semblent trouver un intérêt professionnel à ce type de données. Deuxième aporie : le Web sémantique pourrait être l'une des solutions à ces problèmes, tout comme la mise en place d'une plateforme (comme celle d'Andidot) qui génèrerait une forme de standardisation et d'homogénéisation des jeux de données. Or le premier est un formalisme réservé encore à des initiés – aux *happy few,* et la seconde, pour exister et être performante, générerait un tel travail de centralisation, que pour l'instant elle n'est qu'au stade de projet. Par ailleurs, ce travail de regroupement et d'hébergement de données Open Data (sans homogénéisation) est partiellement fait par le gouvernement français sur la plateforme datagouv.fr ou par des sociétés de service de partage de données (privées et publiques) comme Opendatasoft_._

4.1 Des univers professionnels difficilement conciliables avec les discours scientifiques

Les constats énoncés ci-dessus ont été corroborés par une série d'entretiens menés auprès des chercheurs de l'EMSE et d'Antidot. Les uns comme les autres soulignent le caractère clivant des formats et des langages informatiques, et plus particulièrement du Web sémantique et de la communauté du monde de la donnée. Pourtant, l'un d'entre eux met en avant le fait que les technologies sémantiques ont beaucoup à apporter aux *smart cities*, notamment du côté de l' Internet des objets. Ainsi, pour être exploitées efficacement, les données issues des capteurs installés dans les centres-villes devraient pouvoir être *lues* avec leurs métadonnées, être enrichies et reliées au niveau de la production comme au niveau de la captation.

4.1.1 Des mondes à part ?

Les technologies du Web sémantique, nous l'avons vu, nécessitent un travail d'apprentissage conséquent pour être manipulées efficacement. Cette difficile acculturation au formalisme du langage RDF et du format Json-LD entre autres, constituent certes un frein important mais ce n'est pas le seul. En effet, la question du coût matériel en termes de stockage, de *processing*, de mémoire pour héberger et traiter les Big Data freine les entrepreneurs : « *C'est sûr que si les industriels l'avaient adopté (RDF), les développeurs auraient compris. Une technologie a beau être formidable, si l'industrie ne l'adopte pas, les développeurs ne l'adopteront pas* » (un docteur en informatique de l'EMSE, juin 2018). Pourtant, ces formats qui gravitent dans l'univers du Web sémantique (SPARKL, Json-LD) sont considérés par les chercheurs de l'EMSE comme des standards. Ainsi Json-LD est-il utilisé par de grands groupes, notamment certains GAFAM, comme Google.

Suite à ces différents entretiens et à nos propres observations sur le terrain, nous formulons l'hypothèse qu'il y aurait donc des communautés d'usage des outils et des formats issus du Web sémantique très circonscrites, pour des raisons à la fois matérielles, financières et philosophiques, qui constitueraient aujourd'hui une limite forte à l'utilisation de ces technologies. La communauté scientifique les connaît mais ne semble pas s'en émouvoir, considérant que les discours et les pratiques clivantes, l'incommunication et l'inconnaissance des formats en circulation dans les autres mondes informatiques sont *monnaie courante* dans le champ de l'informatique et du numérique : « *Historiquement, il y a une communauté des logiciens, de ceux qui ont développé le Web et la communauté de ceux qui se sont intéressés aux jeux de données* » (un chercheur en informatique de l'EMSE, juin 2018).

4.1.2 Réconcilier les données ouvertes : une gageure ?

Intégrer des données depuis plusieurs sources constitue le travail quotidien d'un développeur, le faire avec de l'Open Data est moins habituel et uniformiser des volumes conséquents de données ouvertes issues de plusieurs portails métropolitains devient rapidement un casse-tête. C'est à ce problème que se proposaient de répondre, même de façon partielle et ponctuelle, les outils proposés par Antidot et l'EMSE : d'un côté, il s'agissait de permettre l'accès aux données et de l'autre, de faciliter leur réconciliation. En effet, l'un des freins à la mobilisation de ressources Open Data par les professionnels réside dans l'hébergement de ses données sur une grande diversité de portails, selon des formats et des classifications différentes. Pour que les données soient réconciliables, il faut réaliser un travail de taxinomie conséquent afin de créer une ontologie standard à tous les portails :

> C'est très important parce que savoir que la pollution à Grenoble, ça s'appelle 'Pollution' et à Lyon 'Trafic routier', que c'est identifié avec des mots différents, mais que ça veut dire la même chose [...]. Donc avoir une ontologie qui réunit les bons concepts et qui identifie leurs synonymes ou leur façon d'être classés localement c'est très important. On est sûr que si on cherche des jeux de données sur la pollution [...] on aura ces jeux sur Lyon, Grenoble, Paris, parce qu'il y aura eu un ontologue qui aura fait

> ce travail de réconciliation des données. (un informaticien d'Antidot, juin 2018)

L'indexation hiérarchisée, l'interprétation et la réconciliation des données ouvertes entrent pleinement dans le champ de compétences du Web sémantique, mais celui-ci n'a pas encore su séduire – ou trop peu –les premiers réutilisateurs de l'Open Data, à savoir la communauté des développeurs. En effet, même si ceux-ci, lors des deux expérimentations effectuées dans le cadre de l'ANR OpenSensingCit, ont montré un certain intérêt pour les technologies du Web sémantique, tous ou presque ont souligné que ses formats (RDF, Json-LD) n'étaient pas aujourd'hui adaptés à leurs pratiques métiers. En d'autres termes, le langage RDF permettrait certes l'exploitation des données, mais ne conviendrait pas *a priori* pour la réalisation d'une application *end user*[16] (un docteur en informatique de l'EMSE, juin 2018).

Un autre chercheur en informatique de l' EMSE (juin 2018) revient également sur ce point mais, selon lui, la question de l'incompatibilité d'un format pour une pratique métier ne serait pas en soi un obstacle à son développement, dans la mesure où avoir un modèle commun intermédiaire peut faciliter le « *dialogue de données* » entre différents métiers. Ces propos renvoient à nouveau à la question du (format) standard, du *langage commun* facilitant l'échange (voire la redistribution) propre à l' Open Data.

4.2 Open Data et *smart city* pourquoi pas, mais au prix de la productivité informationelle ?

Au final, les réutilisateurs interrogés dans le cadre de l'ANR OpenSensingCity ne rejettent pas l'Open Data en bloc, leur frilosité vis à vis des données ouvertes dans leur pratique métier est plutôt le fait d'*habitus* et d'une forme de communautarisme, même si la qualité des données exploitables entre en jeu également.

4.2.1 Les limites techniques de l'Open Data

En effet, un chercheur en informatique de l'EMSE (juin 2018) le constate : la qualité des jeux de données ouvertes en ligne ainsi que leur disponibilité constituent la première limite à leur exploitation par des professionnels. Selon lui, les communes ont deux possibilités pour libérer leurs données sur les portails : soit en « *format brut* », soit sous la forme d'API. Cette seconde approche semble constituer le choix privilégié des communes mais aussi d'une partie des professionnels de la donnée [17], car elle leur permet « *de construire des URL de manière intelligente avec des paramètres dans l'URL ou dans la requête de l'URL* » (un chercheur en informatique de l'EMSE, juin 2018).

Néanmoins, le fait de pouvoir accéder aux données « *brutes* » ne semble pas séduire outre mesure ces praticiens, dans la mesure où « *ils ont en fait plus l'habitude de travailler sur des API. Cela veut dire aussi qu'ils sont consommateurs d'un service qui*

16. Utilisateurs finaux, anglicisme employé par l'interviewé.

17. Nous nous référons ici aux différents praticiens interrogés dans le cadre de l'ANR OpenSensingCity, qui représentent différents corps de métiers en informatique.

a déjà été pensé et conçu par des développeurs avant » (un chercheur en informatique de l'EMSE, juin 2018).

Il ressort, par ailleurs, de l'ensemble des entretiens menés suite aux deux expérimentions (TUBÀ et le Scop la Péniche) que l'Open Data pour être plébiscité par les professionnels de la donnée doit proposer des formats plus homogènes mais aussi des métadonnées plus uniformes. A cet égard, le Web sémantique constitue sans nul doute l'une des solutions à ces limites, en ne proposant pas seulement une standardisation mais en créant du sens et du lien, là où sont juxtaposées des données ouvertes hétérogènes et éparses.

Ces différents constats amènent une question : les nombreux obstacles à une utilisation élargie de l'Open Data par les professionnels s'avèrent-ils irréconciliables avec le rôle que les politiques attribuent au sein de la *smart city* ?

4.2.2 Open Data et smart city, un mariage forcé ?

Revenons un instant sur la définition que donne S. Goëta du Big et de l'Open Data : « *L'Open Data désigne la diffusion volontaire et proactive de données, essentiellement publiques, qui deviennent librement réutilisables alors que le Big Data regroupe des pratiques d'exploitation de données massives, à grande vitesse et en temps réel* » (Goëta, 2015 : 26) ; celle-ci montre deux choses au moins : d'une part, Big et Open Data ont un rapport à la production des données très différents. D'autre part, l'exploitation de ces données divergent comme leurs bénéficiaires : celle des Big Data vise à « *l'efficience et la prédictibilité* » alors que derrière celle de l'Open Data, se joue une forme *« d'injonction à l'innovation et la transparence »* (Goëta, 2015 : 27).

Enfin, si l'exploitation des Big Data permet aux organisations une meilleure connaissance de leur environnement et donc renforce leur capacité à agir ou réagir selon les cas de figures qui se présentent à elles, l'Open Data vise (selon les textes législatifs) la modernisation de l'Etat et l'*empowerment* des citoyens. Au regard de ces différents points, il apparaît clairement que l'exploitation des Big Data bénéficie aux communes – organisations à part entière – et les conduit vers une nouvelle forme d'intelligence (Smart City), alors que l'ouverture des données publiques constitue davantage une contrainte pour ces mêmes communes, sans leur apporter de véritables bénéfices informationnels en retour (Hare, 2016 ; Paquienseguy, 2016 a).

Pourquoi alors les discours d'accompagnement de l'Open Data associent-ils fréquemment données ouvertes et *smart city* ? L'ouverture des données rendrait-elle la ville plus intelligente ? Probablement pas car, en l'état, les données produites de façon massive (les Big Data) sont rarement partagées et donc ouvertes. D'une part, parce que ces données sont souvent produites par des opérateurs privés puis revendues aux communes et, d'autre part, parce que leur exploitation nécessite des données nettoyées et reliées. L'Open Data n'en libère pas et sa production est trop aléatoire pour pouvoir concurrencer aujourd'hui le Big Data_._

Néanmoins, certaines initiatives contredisent ce point puisque des « *données publiques sont aussi traitées et agrégées par des acteurs privés et associatifs pour constituer des bases massives et partagées* » (Goëta, 2015 : 32). C'est le cas d'initiatives privées

comme celles de Regards citoyens qui donne accès – *via* les sites Nosdeputes.fr et Nossenateurs.fr – aux statistiques de de l' Assemblée nationale et du Sénat, permettant ainsi une surveillance citoyenne (Hare, Le Béchec, 2015). De son côté, OpenStreetMap produit une cartographie mondiale et collaborative des données géographiques des pays. Ces différents projets participent au développement d'une société plus *éclairée* (au sens des Lumières) et d'une connaissance enrichie du territoire (physique et/ou symbolique, national ou local), contribuant d'une certaine manière à produire de nouvelles formes d'intelligence, qui s'inscriraient davantage dans l'idéal d'un espace public éclairé par l'Open data que dans les cadres définitoires des *smart cities*. Toutefois, la production de données ouvertes massives et leur exploitation au profit du territoire urbain restent encore rares.

L'association fréquente des termes d'Open Data et de *smart city* alors que, dans les faits, c'est avant tout l'exploitation des Big Data qui permet une meilleure connaissance de la ville s'expliquerait selon S. Goëta de la manière suivante :

> La proximité sémantique entre Big Data et Open Data prête à confusion, tout comme leur émergence simultanée. Les deux mouvements ont suscité des espoirs en matière de création de savoir et d'efficience des organisations mais aussi des craintes portant essentiellement sur les atteintes à la vie privée qui pourraient survenir lors de l'exploitation des données. (Goëta, 2015 : 32)

Il y aurait donc une forme d'amalgame entre les deux types de data, les uns et les autres étant porteurs d'une promesse d'*empowerment,* mais sur des plans différents _(_voir *supra)*. Pas de mariage forcé donc, mais des trajectoires parallèles qui peuvent prêter à confusion, et que nous retrouvons fréquemment dans les discours d'accompagnement présents sur les portails métropolitains (Labelle, Le Corf, 2012), dans les stratégies et les outils de communication déployés par les collectivités locales ou les institutions gouvernementales (Hare, à paraître).

Au final, même si cela constitue l'un des objectifs avancés par les promoteurs de l'Open Data au niveau local, il y a peu d'exemples d'exploitation massive et systématique de données publiques ouvertes. Effectivement, certains jeux de données présents sur les portails Open Data sont réutilisés par des développeurs pour produire des applications au bénéfice des citoyens et leur permettre une meilleure appréhension de la complexité et de la labilité du territoire urbain, mais cela reste marginal (Turki, Foulonneau, 2015) au regard de l'ampleur de l'exploitation des Big Data.

L'Open Data contribue, d'une certaine manière, au développement de la *smart city*, mais davantage du côté du citoyen que de l'organisation. Quoi qu'il en soit, l'une des hypothèses fortes qui ressort de l'ensemble des entretiens menés dans le cadre de l'ANR OpenSensingCity résiderait dans le fait que les professionnels de la donnée ne semblent pas ou peu enclins à céder, en l'état, aux sirènes de l'Open Data tant ces données ouvertes leur semblent imparfaites, désordonnées et difficilement réconciliables, en tout cas peu conciliables avec les impératifs professionnels auxquels ils sont confrontés quotidiennement en matière de productivité informationnelle. Cette

hypothèse ne pourra bien évidemment être pleinement validée (ou invalidée) qu'à la suite de nouveaux terrains et entretiens.

5 Conclusion

Pour répondre à la problématique initiale qui interrogeait la correspondance entre les discours d'escorte et les propositions d'outillage scientifique de l'Open Data d'une part, et les besoins des professionnels de la donnée d'autre part, nous observons trois choses : tout d'abord, les représentations et les discours sur la *smart city* participent grandement à la confusion entre open et Big Data, tant en termes de formats, de formes que d'objectifs.

Ensuite, si la pertinence du Web sémantique pour réconcilier les données ouvertes a été démontrée, cela ne suffit pas à dépasser les clivages de formats et d'*habitus* présents entre les pratiques métiers. Par ailleurs, la mise en œuvre de cette réconciliation représente à l'heure actuelle un coût humain beaucoup trop important pour être envisagée non seulement au niveau des diffuseurs de l'Open Data, mais aussi de celui de ses utilisateurs qui doivent répondre à une injonction d'efficacité et de productivité de leur employeur.

Enfin, sur le terrain des villes, les données publiques ouvertes participent encore assez peu à la construction de la *smart city*, en dépit des imaginaires affichés par les acteurs politiques.

Ces trois éléments concourent chacun, à des degrés divers, à nourrir les imaginaires de l'Open Data, que les données publiques ouvertes soient étroitement liées au développement de la *smart city* ou qu'elles caractérisent une forme de tropisme informationnel, générant des communautés de praticiens informatiques (*data analysts*, *data scientists*), souvent peu enclines à en faire usage dans le cadre de leur métier. Les raisons de cette frilosité à utiliser l'Open Data sont nombreuses et les façons d'y remédier peu efficientes pour l'instant.

Afin de sortir de cette impasse à la fois ontologique (confusion entre Big et Open Data), technologique (données Open Data incomplètes et non réconciliées) et pratique (communautés informatique et professionnelle clivantes), V. Peugeot propose une alternative au développement de la *smart city* par les data qui serait d'« *imaginer des régimes mixtes, avec des données froides gratuites et des données en temps réel payantes, ou des systèmes de seuil* » (2014, 22). Certaines agglomérations, comme le Grand Lyon, ont déjà tenté l'expérience avec le principe de licences échelonnées (ouverte, engagée et associée) selon leur utilisation sur la plateforme data.grand.lyon.

Pour conclure, et au-delà de la question de la participation des data (Open et Big) à la ville intelligente, nous souhaitons emprunter à Marc Chopplet l'idée que : « *L'horizon de la conception de l'espace-temps de la smart city, comme anticipation de l'agir possible, est naturellement technologique mais surtout politique* » (2018 : 81).

Bibliographie

Chopplet, M. (2018). Smart City : quelle intelligence pour quelle action ? Les concepts de John Dewey, scalpels de la ville intelligente. *Quaderni*, 96, p. 71-86. doi : 10.4000/quaderni.1179

Denis, J. et Goëta, S. (2016). « Brutification » et instauration des données. La fabrique attentionnée de l'Open Data, *i3 Working Papers Series*, 16-CSI-01.

Dymytrova, V. et Paquienséguy, F. (2017). La réutilisation et les réutilisateurs des données ouvertes en France : une approche centrée sur les usagers ». *Revue Internationale des Gouvernements Ouverts*, 5, p. 117-132. Repéré sur : http ://ojs.imodev.org/index.php/RIGO/article/view/204/338

Dymytrova, V., Hare, I., Peyrelong, M.-F., Larroche, V. et Paquienséguy, F. (2018). *Synthèse des résultats des expérimentations des outils OSC au Tubà et à la Péniche*. Livrable 4 (Rapport de recherche ANR OpenSensingCity 14-CE24-0029). EA 4147 Elico. Repéré sur : https ://hal.archives-ouvertes.fr/hal-01823479

Dymytrova, V., Larroche, V. et Paquienséguy, F. (2018). Cadres d'usage des données par des ANR OpenSensingCity 14-CE24-0029). EA 4147. Elico. Repéré sur : https ://hal.archives-ouvertes.fr/hal-01730820v1

Goëta, S. (2015). Un air de famille : les trajectoires parallèles de l'Open Data et du Big Data. *Informations sociales*, 5 (191), p 26-34. doi : 10.3917/inso.191.0026.

Goëta, S. et Mabi, C. (2014). L'Open Data peut-il (encore) servir les citoyens ? *Mouvements*, 3 (79), p. 81-91. doi : 10.3917/mouv.079.0081.

Greenfield, A. (2013). *Against the Smart City (The City is Here for You to Use)*. New York : Do Projects.

Hare, I. et Le Béchec, M. (2015). L'Open Data : un acteur au service de l'espace public ? Dans F. Lienard, S. Zlitini (dir.). *La communication électronique : enjeux, stratégies, opportunités,* Limoges : Lambert-Lucas, p. 185-198.

Hare, I. (2016). Les stratégies des diffuseurs de données ouvertes en Région. Dans Paquienséguy, F. (dir.). *Open Data : Accès, collectivités territoriales et citoyenneté : des problématiques communicationnelles* (p. 71-88). Paris : Éditions des archives contemporaines. doi : 10.17184/eac.9782813001856.

Hare, I. (à paraître en 2020), L'Open Data au prisme de l'impensé numérique. Dans P. Robert (dir.). *L'impensé numérique*, 2. Paris : Éditions des archives contemporaines,

Ibekwé-Sanjuan, F. et Paquienséguy, F. (2015). Open, Big, Collaboration : trois utopies de l'innovation au XXIe siècle. Dans E. Broudoux et G. Chartron (dir.). *Big data- Open data. Quelles valeurs ? Quels Enjeux ?* (p. 15-29). Louvain-la-Neuve : De Boeck Supérieur. doi : 10.3917/dbu.chron.2015.01.0015.

Kitchin, R. (2013). Big Data and human geography : Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), p. 262-267. doi : 10.1177/2043820613513388.

Kitchin, R. (2014 a). The real-time city ? Big Data and smart urbanism. *GeoJournal*, 79(1), p. 1-14. doi : 10.1007/s10708-013-9516-8.

Kitchin, R. (2014 b). *The Data Revolution : Big Data, Open Data, data infrastructures and their consequences*. Londres : Sage.

Labelle, S. et Le Corf, J.-B. (2012). Modalités de diffusion et processus documentaires, conditions du "détachement" des informations publiques. Analyse des discours législatifs et des portails Open Data territoriaux. *Les Enjeux de l'Information et de la Communication*, 13(2), GRESEC-Université Grenoble Alpes, p. 59-71. doi : 10.3917/enic.013.0059

Larroche, V. et Dymytrova, V. (2017). Le Web sémantique, un moyen de visibilité et d'exploitation des Open Data auprès des communautés de réutilisateurs professionnels ? *Colloque international pluridisciplinaire Big Data et visibilité en ligne : un enjeu pluridisciplinaire de l'économie numérique*, Novembre 2017, Université des Antilles, Fort-de-France, Martinique. Repéré sur : https ://hal.archives-ouvertes.fr/hal-01718068

Larroche, V. et Dymytrova, V. (2018). L'apport des SIC aux projets Open Data pour une meilleure prise en compte des réutilisateurs dans la conception des outils de traitement des data. *Communication & Organisation*, 2(54), Presses Universitaires de Bordeaux, France, p. 69-80.

Le Corf, J.-B. (2015). Management de l'information publique et innovation numérique de services urbains : l'intelligence territoriale en perspective. Dans E. Broudoux et G. Chartron (dir.). *Big Data – Open Data : Quelles valeurs ? Quels enjeux ?* (p. 97-112). De Boeck Supérieur, Louvain-la-Neuve. doi˜:10.3917/dbu.chron.2015.01.0097.

Mayer-Schonberger, V. et Cukier, K. (2013). *Big Data : A Revolution that will Transform How We Live, Work and Think*. Boston : Houghton Mifflin Harcourt.

Paquienséguy, F. (2016 a). Smart city & Open Data : à qui profitent les données ouvertes ? *Colloque international des Sciences du Territoires 2016*, Proceedings « En quête de territoire (s), Looking for territories », Grenoble, 17-18 mars 2016, p. 351-357.

Paquienséguy, F. (dir.) (2016 b). *Open Data : Accès, collectivités territoriales et citoyenneté : des problématiques communicationnelles*. Paris : Éditions des Archives contemporaines. doi:10.17184/eac.9782813001856.

Peugeot, V. (2014). Collaborative ou intelligente, la ville entre deux imaginaires. Dans M. Carmes et J.-M. Noyer (dir.). *Devenirs Urbains*, p. 43-64. doi : 10.4000/books.pressesmines.2368.

Ribes, D. et Jackson, S. J. (2013). Data bite man : The work of sustaining a long-term study. Dans L. Gitelman *et al.*(dir.). *Raw data is an oxymoron* (p. 147-166). Cambridge : The MIT Press. doi : 10.7551/mitpress/9302.001.0001.

Townsend, A. M. (2013). Smart Cities : Big Data, civic hackers, and the quest for a new utopia. New York : W.W. Norton.

Turki, S. et Foulonneau, M. (2015). Valorisation des données ouvertes : acteurs, enjeux et modèles d'affaires », Dans E. Broudoux et G. Chartron (dir.). Big Data – Open Data : Quelles valeurs ? Quels enjeux ? (p. 113-125). *Louvain-la-Neuve : De Boeck Supérieur.* doi : 10.3917/dbu.chron.2015.01.0113.