

PVETE

L'Open Data en pratique

*Projet de veille technologique
Encadré par Guillaume MOREAU*

Sébastien COHENDET;Maxime
Maxime TERRADE



Table des matières

Avant-propos.....	4
Introduction : présentation de l'Open Data.....	5
I. Contexte de l'Open Data.....	7
1. La nature des données.....	7
a. Les lacunes actuelles.....	7
b. La place du citoyen dans l'Open Data.....	7
c. Une volonté démocratique qui se met doucement en place.....	8
d. Licences.....	9
2. L'aspect économique.....	10
a. Le cas du gouvernement.....	10
b. Le cas des entreprises.....	10
c. Devenir producteur d'informations : se poser les bonnes questions.....	11
d. Un modèle économique à trouver.....	12
II. Les utilisations de l'Open Data.....	13
1. Les applications web et mobiles.....	13
a. Les applications de parking.....	13
b. Les applications de transports.....	14
2. De nouvelles utilisations.....	16
a. We Are Data.....	16
b. City Dashboard.....	17
c. La montre intelligente « Muni Watch ».....	18
d. La boîte.....	18
e. La data-journalisme.....	19
III. Les limites de l'Open Data.....	22
a. Des données non triées.....	22
b. L'absence de normes.....	22
c. Des données ouvertes, mais à ne pas voir.....	23
d. Une ouverture partielle des données gouvernementales.....	23
e. Ouverture de données limitée par les entreprises.....	24
f. La faible quantité de données temps réel.....	24
g. Des données non permanentes.....	24

Conclusion	26
Sources	27
1. Introduction.....	27
2. Contexte de l'Open Data	27
3. Les utilisations de l'Open Data	28
4. Les limites des applications	29
5. Limites des données ouvertes.....	29
Table des figures	30

Avant-propos

Ce rapport s'inscrit dans le cadre d'un projet de veille technologique (abrégé en PVETE), réalisé par deux étudiants en informatique à l'Ecole Centrale de Nantes. Il est une transcription de nos différents billets de blog sur le sujet de l'Open Data en pratique, disponibles sur veille-techno.blogs.ec-nantes.fr. Tous nos articles ont été publiés sous le pseudonyme « opdapra13 ».

Tout au long de ce projet, nous avons été encadrés par Guillaume Moreau, enseignant-chercheur à l'Ecole Centrale de Nantes. Nous tenons à le remercier.

Nous remercions également nos lecteurs, qui ont fait preuves d'intérêt pour nos billets en les citant à maintes reprises sur leurs propres blogs.

A noter que ce document a été pensé à la fois pour une lecture informatique (comprendre que les liens sont cliquables) mais aussi comme un document papier (l'ensemble des liens sont disponibles en dur, dans les sources se rapportant à chaque partie).

Introduction : présentation de l'Open Data

L'Open Data, tout le monde en parle ces derniers temps mais savez-vous réellement ce que cela signifie ?

Il s'agit d'un principe selon lequel les données publiques (celles recueillies, maintenues et utilisées par les organismes publics) ou privées doivent être disponibles pour accès et réutilisation par les citoyens et entreprises. Nantes est une des premières villes, en France, à avoir adopté l'Open Data, dès février 2011.

Le mouvement de l'Open Data provient de la fondation américaine Sunlight Foundation qui a énoncé les [10 principes de l'Open Data](#) :

“Les données doivent être complètes, primaires, fraîches, accessibles électroniquement, lisible par une machine, accessible sans discrimination, respectant les standards ouverts, disponible sous une licence claire, accessibles de façon pérenne et gratuite.”

Dans le cadre nantais, cette démarche a pour but de dynamiser l'économie nantaise par l'utilisation de ces données et d'améliorer les services dédiés aux usagers et entreprises. C'est ainsi qu'est né le site : data.nantes.fr qui regroupe toutes sortes d'informations que nous nous ferons un plaisir de critiquer (positivement ou non).

Nantes n'est pas la seule ville à s'être lancée : Paris ou le gouvernement français sont d'autres intervenants, pour n'en citer que deux. Bien que nous nous attachons ici à la France, des données mondiales sont également disponibles (les réseaux de vélos en libre-service ont leur Open Data mondiale par exemple).

Des données, c'est utile, mais encore faut-il que des applications se basant sur ces données existent : nous avons analysé un nombre conséquent d'applications pour mieux comprendre ce phénomène.

Par exemple, vous êtes-vous déjà demandé si des applications réellement utiles basées sur l'Open Data existaient ? En effet, de nombreuses applications sont présentes sur toutes les plateformes, se vantant d'utiliser des données libres issues de l'Open Data. Mais combien sont d'une réelle utilité ?

Autre chose qu'il est être important d'aborder : les informations adjacentes aux données Open Data. Certaines sociétés se vantent en effet de publier des centaines et des centaines d'Open Data, mais sont-elles réellement intéressantes ? Ne serait-ce pas plutôt les informations adjacentes, gardées secrètes, qui seraient utiles au public ? Pour quelles raisons ces données sont cachées ? Secret industriel ? Cache-misère ? L'arbre qui cache la forêt ?

Enfin, nous aborderons les normes en lien avec l'Open Data. Comment est organisée l'Open Data actuellement ? Quelles sont les normes permettant d'organiser tout cela ? Quelles sont-elles et combien en existe-t-il ?

I. Contexte de l'Open Data

1. La nature des données

Si le mot Open Data permet d'accéder à de la *donnée libre*, il semble opportun de s'intéresser aux données en elles-mêmes. Existe-t-il réellement des myriades d'informations mises à disposition librement en ligne ? Sont-elles des initiatives individuelles, d'entreprises ou encore gouvernementales ? Enfin, qui dit données, dit protection de ces données : nous verrons également les licences se cachant derrière l'Open Data.

a. Les lacunes actuelles

Il faut savoir que le lancement du mouvement Open Data en France ne s'est pas fait sans anicroches. En effet, la démarche a révélé les lacunes des informations en France : que ce soit un manque d'informations, un mauvais archivage ou encore une mauvaise structuration des données. C'est donc tout l'enjeu actuel : résoudre ces lacunes pour permettre un véritable essor des données libres en France.

Quand bien même de la donnée est présente sur le net, la mouvance Open Data n'est que peu connue du grand public. Pire, elle est très compliquée à appréhender pour ces derniers. En effet, les données publiées sont brutes, sans aucun traitement et dans des formats divers et variés. Difficile pour Mme Michu de s'y retrouver dans les formats exotiques proposés : fin 2013, sur le site data.nantes.fr, on trouve du .csv, du json, xml, xls, etc.

On comprend ainsi que l'Open Data fournit la donnée brute qui sera alors interprétée par des applications développées pour le grand public.

b. La place du citoyen dans l'Open Data

On retrouve cette idée dans le schéma ci-dessous : les collectivités transmettent leurs données pour l'Open Data à des plateformes de données en ligne sur lesquelles on peut trouver l'information Open Data de façon brute. C'est alors que des personnes expérimentées prennent le relai et vont pouvoir utiliser ces données brutes pour les transformer en un produit fini à plus forte valeur ajoutée. Ainsi, les développeurs vont pouvoir créer des applications se basant sur les plateformes de données.

Une entreprise pourra enrichir ses services ou en créer de nouveaux : l'Open Data peut alors favoriser la création de nouvelles entreprises. De la recherche peut être effectuée

dans le domaine afin d'y faire des découvertes. Les designers peuvent enrichir l'expérience en créant de nouvelles façons de représenter cette donnée brute.

Enfin, les journalistes peuvent écrire des articles dans le cadre du data journalisme (journalisme s'appuyant sur les données libres, typiquement l'Open Data). Ces nouveaux produits vont ainsi pouvoir toucher le grand public mais aussi les collectivités elles-mêmes.



Figure 1 - La place du citoyen dans le mouvement Open Data

Ainsi, un cercle vertueux semble se mettre en place : les collectivités donnent leurs données libres et ces données pourront être utilisées par d'autres (voire elles-mêmes) pour faciliter leurs fonctionnements grâce aux travaux d'intermédiaires.

c. Une volonté démocratique qui se met doucement en place

Si l'Open Data peut aussi être synonyme de transparence, notamment au niveau du gouvernement et de ses actions, ce but est encore loin d'être atteint. Rares sont les plateformes de données donnant accès aux données budgétaires, des subventions ou encore des actions chiffrées des collectivités.

Cette volonté démocratique se met doucement en place, particulièrement sur la [plateforme du gouvernement](#). On y trouve les dépenses en matière de santé, les dépenses des administrations, de grandes villes (de plus de 10 000 habitants) sur différents domaines.

Là encore, ces données sont dans un format peu engageant pour le grand public (format Excel XLS, peu agréable à lire). De même, le format XLS est très médiocre pour le développement (l'accès à la donnée n'est pas simple car le « parsing » n'est pas évident...).

d. Licences

Pour terminer cette partie, amenons la question des licences sur la table en évoquant les deux licences principales de l'Open Data.

A l'heure actuelle, deux licences semblent s'imposer :

- la « [Licence Ouverte](#) » créée par la mission gouvernementale [Etalab](#)
- la licence « [Open Database License \(ODbL\)](#) » créée par [Open Data Commons](#)

Tandis que la licence ouverte impose seulement de citer la source (il s'agirait d'une licence Creative Commons BY) afin d'utiliser les données, commercialement y compris, la licence ODbL nécessite le partage des données enrichies suivant la même licence (c'est donc une licence Creative Commons BY-SA).

2. L'aspect économique

La notion d'Open Data est souvent associée à une dimension citoyenne/démocratique, qui connote par là une absence de profit. La réalité est pourtant tout autre. En effet, l'Open Data s'inscrit la plupart du temps dans un contexte économique, que ce soit à court ou à long terme. Mais quels sont les bénéfices directs apportés par les données libres ? Et quels sont les profits indirects qui sont actuellement tirés de celles-ci ?

a. Le cas du gouvernement

35 millions d'euros, c'est le montant qu'a rapporté la réutilisation des Open Data aux administrations françaises en 2012. Le magistrat à la Cour des comptes Mohammed Adnène Trojette évoque ce chiffre au travers son rapport « [Ouverture des données publiques : les exceptions au principe de gratuité sont-elles toutes légitimes ?](#) ». Ce rapport permet d'apporter des éléments de réponses sur la non-gratuité de certaines données libres. On apprend par exemple qu'« *une vingtaine de services publics administratifs ont institué ou maintenu des redevances de réutilisation* » de certaines données.

Ce réel frein à l'Open Data a des causes connues : il s'agit principalement de la peur que « *les informations ainsi communiquées ne soient réutilisées pour critiquer le service public* », explique le rapport.

Pour nuancer ce chiffre colossale de 35 millions d'euros, il faut toutefois savoir que la majorité des gains est perçue par trois organismes / ministères : l'INSEE et l'IGN perçoivent chacun près de 10 millions d'euros et le ministère de l'intérieur près de 4 millions d'euros. De plus, ces gains sont en baisse depuis plusieurs années (30% en deux ans).

b. Le cas des entreprises

Parlons maintenant des entreprises, en abordant une phrase de François Bancilhon, PDG de Data Publica : les entreprises « *peuvent valoriser leurs données en les ouvrant pour créer un écosystème vertueux autour d'elle* ». Cette citation montre bien que les entreprises cherchent elles-aussi à tirer profit de l'Open Data. Petit tour d'horizon des profits attendus par les entreprises.

Premièrement, les entreprises cherchent à améliorer la satisfaction de leurs clients grâce à l'ouverture des données, leur donnant ainsi une meilleure image de marque.

Deuxièmement, elles cherchent à améliorer sa faculté d'innovation : grâce à de tierces personnes (développeurs, indépendants, start-up...) réutilisant leurs données, de nouvelles idées vont alors apparaître, permettant ainsi aux entreprises de se développer.

Troisièmement, la libéralisation de leurs données permet de prévenir les réutilisations non autorisées. C'est le cas de la RATP qui a mis fin à ces pratiques en rendant publiques leurs données, celles-ci ayant ainsi permis la création de plusieurs applications pour les usagers de RATP dont une permettant de connaître les flux de voyageurs, utiles pour les futurs enseignes qui chercheraient la meilleure station de métro pour s'installer.

Un autre point concerne les interfaces de programmation (API). Ces dernières, ouvrant l'accès à des données de l'entreprise, peuvent permettre de réaliser indirectement du profit. Amazon en est un très bon exemple : ouvrant ses bases de données de livres via une API, des nombreux sites professionnels et des amateurs ont pu afficher sur leur site ou blog des informations sur les livres. Les internautes intéressés par un roman sont alors redirigés vers Amazon pour l'acheter. Le revenu est partagé entre Amazon et le site utilisant l'API.

Dernier point : certaines entreprises proposent des services concernant les « Open Data », par exemple un accès à des données personnalisés (triées et sélectionnées pour le client) ou à des données libres enrichies. Data Publica est une entreprise proposant ce type de services.

c. Devenir producteur d'informations : se poser les bonnes questions

On le comprend aisément, l'Open Data prend de la valeur lorsqu'elle circule ou lorsqu'on la transforme. C'est pourquoi il peut être intéressant de devenir producteur d'informations. Mais il faut commencer par se poser les bonnes questions, lesquelles se regroupent en trois principaux points :

- Le mode d'accès aux données : est-ce que nos données vont être accessibles via le site web de l'entreprise, via un portail spécifique, ou bien par le biais de plateforme mobile : tablettes, smartphones, ... Tout ceci sans oublier de présenter ces données de manière simple et pratique. Certains producteurs de données ont d'ailleurs du progrès à faire de ce côté-là.
- L'hébergement des données : est-il possible de stocker, gérer et maintenir ces données en interne ou mieux vaut-il faire appel à une autre entreprise, spécialisée dans ce domaine ?
- Le niveau de protection : quels vont être les administrateurs du projet ? Qui aura accès à quoi ? Et surtout la sécurité de l'entreprise : cette ouverture de données crée-t-elle une faille de sécurité qui permettrait d'infiltrer le réseau interne de l'entreprise ?

d. Un modèle économique à trouver ...

Dans un contexte de crise (plus ou moins terminée), on comprend bien pourquoi les administrations et entreprises utilisent l'Open Data à des fins économiques, tout au moins en partie. Toutefois, comme les spécialistes l'avaient déjà constaté lors de la semaine européenne de l'Open Data à Nantes en mai 2012, le modèle économique de l'« open data » reste encore à trouver.

Si on prend l'exemple des applications basées sur l'Open Data, on se rend compte qu'elles ne génèrent pas assez de revenus pour développeurs... Il est donc impossible que le monde de l'Open Data s'appuie sur les applications qui les utilisent.

Ceci s'explique par la faible portée des applications, majoritairement développées pour une certaine ville ou région et pour un domaine spécifique (transport en commun, vélo, tourisme, etc.). Sans oublier que ces applications sont souvent redondantes et concurrentes.

II. Les utilisations de l'Open Data

1. Les applications web et mobiles

a. Les applications de parking

Les données qu'on retrouve le plus souvent en ligne et qui sont aussi les plus utilisées par les développeurs sont les données liées aux parkings d'une ville. Rien que sur la [plateforme d'applications Open Data](#) tenue par Nantes Métropole, on trouve pas moins de 4 applications sur les parkings sur les 13 applications disponibles. Près d'un tiers donc des applications.

Il faut dire que l'information est simple à mettre en ligne car déjà présentes dans les systèmes informatiques (pour les panneaux annonçant le nombre de places disponibles devant un parking). La tâche est également aisée pour le développeur qui n'a qu'un seul type d'informations à afficher.

Il est d'ailleurs malheureux que les applications actuelles ne fassent qu'afficher de la donnée brute, sans réel traitement derrière. On peut citer [Parking Direct](#), disponible sur Android, qui ne fait qu'inventorier les parkings et le nombre de places disponibles, avec lancement de Google Maps pour guider l'utilisateur au parking. Utile mais peu original.

On pourrait imaginer une application qui d'un point A à un point B recoupe les informations de trafic, de transports en commun et de parking pour pousser l'utilisateur à lâcher sa voiture (concept imaginé il y a 2 ans lors d'un Projet d'Etudes Industrielles entre Transway et l'Ecole Centrale de Nantes, toujours au stade de concept).

A savoir qu'Orange a déployé sur Nice une flotte de capteurs pour détecter si des emplacements de stationnement sont libres ou pas, et ce afin de guider les conducteurs vers la place libre la plus proche (ou communiquer l'information aux agents de polices municipaux...). Encore en phase de test, nous ne pouvons savoir si les données seront à vocation libres...

Toutefois, si vous ne voulez pas vous servir de toutes ces applications de parking, vous avez encore la possibilité de vous garer comme bon vous semble, sur un trottoir ou un emplacement non autorisé. Vous vous exposez alors à la fourrière !

Une application smartphone étonnante a été créée à cet effet : il s'agit de « *Pref Police* » qui, comme on peut le deviner, a été développée par la préfecture de Police de Paris. Grâce à *Pref Police*, vous pouvez savoir instantanément dans quelle fourrière votre voiture se trouve, en renseignant simplement sa plaque d'immatriculation. Vous pouvez par

ailleurs découvrir (ou redécouvrir peut-être) quels sont les documents nécessaires pour récupérer votre voiture.

Malgré l'effort de la préfecture de Police, peut-on réellement parler d'Open Data dans ce cas ? En effet, la préfecture de Police parisienne ne publie pas la liste des véhicules enlevés, contrairement à ce qui est fait à [Chicago](#) par exemple. Ces données pourraient intéresser bon nombre de citoyens : on peut imaginer une application qui informerait les utilisateurs du nombre de véhicules enlevés à un endroit précis. Si dix véhicules ont été mis en fourrière dernièrement dans la même rue, un automobiliste ne s'y risquera sûrement pas !

b. Les applications de transports

Parlons maintenant de l'application Tan, mise en ligne le 12 mars 2012. L'application permet en particulier d'obtenir les horaires des prochains tramways/bus à proximité, de rechercher des itinéraires et connaître l'état du trafic en temps réel (informations concernant les mouvements sociaux, les manifestations culturelles ou sportives sur les lignes, les désagréments dus aux intempéries, ...).

L'application se base sur trois API, ouvertes à tous les développeurs. Au début de l'année 2012, la Tan voyait les choses en grand. Elle annonçait trois nouvelles API disponible fin mars 2012 : la liste des arrêts, le temps d'attente aux arrêts et les horaires aux arrêts. Cette dernière API devait, soi-disant, être en temps réel pour le tramway et le busway dès le lancement (fin mars 2012), et d'ici à début 2013 pour l'ensemble des bus du réseau.

La réalité est tout autre, comme on peut le lire dans le témoignage suivant. En effet, fin septembre 2012, un développeur agacé s'exprimait : "*Les données sont aujourd'hui statiques et théoriques, ce qui ne nous permet pas de garantir un service de qualité en temps réel. [...] Cela ne constitue pas une ouverture de leurs données en temps réel comme cela devrait être le cas lorsque l'on parle d'Open Data*".

Aujourd'hui, la Tan affiche sur son site : « *les horaires sont disponibles en temps réel pour le Tramway et le Busway* », soit 4 lignes parmi les 47 lignes de la Tan. Même si les données temps réel commencent à apparaître, un gros progrès reste donc à faire. D'autant plus que ces données existent, puisque qu'elles sont disponibles sur des panneaux d'affichage à de nombreux arrêts. On peut d'ailleurs se demander si la Tan nous cache ces informations délibérément, pour masquer les retards éventuels sur leurs lignes.

Prenons pour continuer l'exemple de l'application « *ToutNantesBouge* », une n-ième application pour les transports publics nantais, développée par la SSII « Gfi Informatique ».

Celle-ci, comme de plus en plus d'applications, est 100% web mobile. Comprenez par là que l'application s'ouvre sur n'importe quel navigateur mobile. Le principal avantage de

cette solution réside en un développement rapide et multiplateforme, contrairement aux applications natives.

Pour en revenir à l'Open Data, cette application se base sur les API de la Tan, ainsi que sur les données ouvertes de l'ensemble du réseau de transport de la ville. On peut citer la possibilité de connaître la disponibilité des places de parking en temps réel ou les perturbations (bouchons, travaux, ...) sur un itinéraire.

Sachez que « *ToutNantesBouge* » est né dans le cadre d'un concours Open Data, proposé par la ville de Nantes. En effet, pour essayer de lancer des projets innovants et surtout différents, de nombreuses métropoles ont lancé, au travers de leur portail Open Data, des concours pour les développeurs les incitant à créer de nouveaux concepts.

Ce fut le cas à Nantes, en début d'année dernière où un appel à projet a été effectué. Le 1er prix a été décerné à ... une [application d'affichages des places libres](#) de parkings aux alentours. Décidément !

Strasbourg a également lancé un tel concours mais avec, cette fois, une rubrique "*Jamais vu*" visant à créer, enfin, quelque chose de novateur. On peut retrouver les candidats sur [cette page](#).

Même si ces concours sont une bonne chose, on croise trop souvent les mêmes concepts, désormais usés. A quand une révolution en matière de traitement des données Open Data ?

2. De nouvelles utilisations

Avec des phénomènes tels que le big data ou l'Open Data, un des enjeux de demain sera l'utilisation des données qui nous entourent. Si l'exploitation des données Open Data passe la plupart du temps par des applications mobiles, elle peut aussi passer par une visualisation graphique de ces données.

Un nouveau mouvement journaliste est même né : le data-journalisme. Devant l'abondance de données, un certain esprit de synthèse semble primordial.

Nous verrons dans cette partie, différentes initiatives de visualisation de données ainsi qu'un approfondissement sur le data-journalisme.

a. We Are Data

Commençons par certainement la plus originale mais pas forcément la plus utile, le site internet [We Are Data](#), une campagne de publicité virale pour le jeu vidéo *Watch Dogs*, d'Ubisoft.

Le jeu se base sur la manipulation des données nous entourant : caméra de surveillance, téléphones portables et autres appareils connectés. Afin de démontrer que le monde décrit dans le jeu n'est pas si irréel que cela, l'équipe Marketing du jeu a décidé d'agréger les contenus Open Data de 3 capitales (Paris, Londres et Berlin).

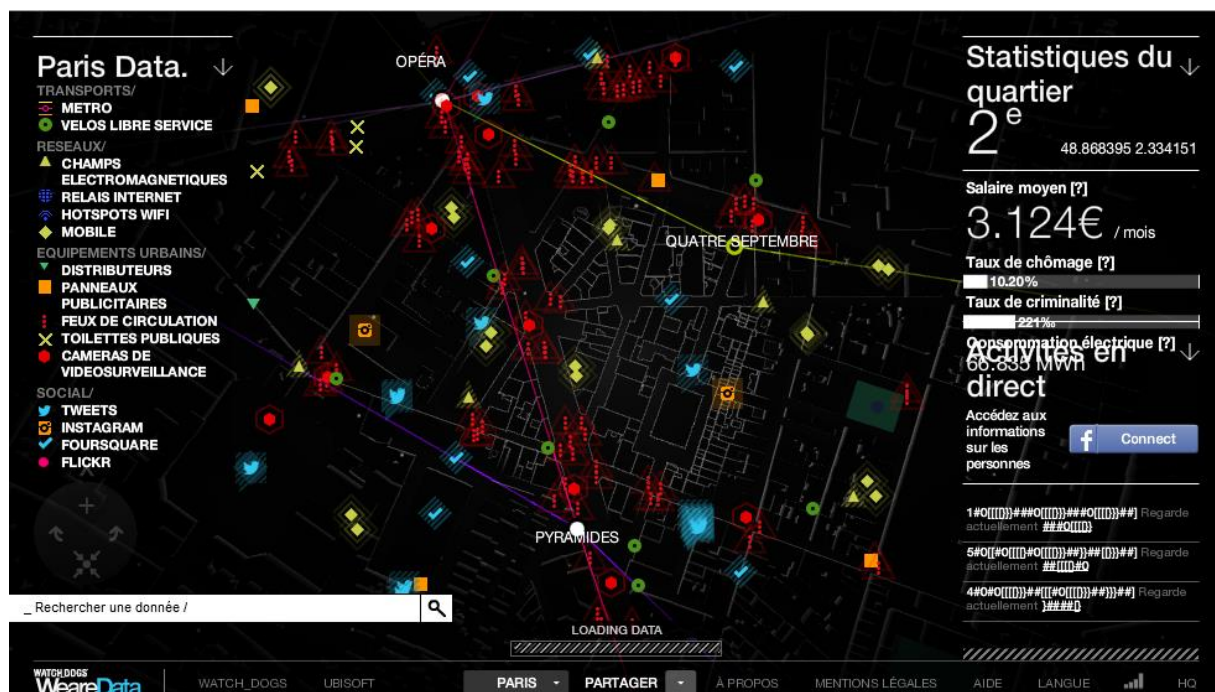


Figure 2 - Interface du site "We Are Data" pour la ville de Paris

On découvre alors une autre facette des données, celle de l'information graphique. Cela passe par une vue de nombreuses informations dans une carte interactive

des villes (sous forme d'un player flash) : l'affichage des stations de vélos et du nombre de vélos disponibles (via les données libres de JCDecaux), des stations de métro avec positionnement de la rame en temps réel (en réalité la RATP ne donnant que les horaires de passage théorique des rames, l'application ne fait qu'une interpolation suivant l'heure actuelle), les positions des antennes téléphoniques, etc.

Le plus intéressant dans cette expérience, c'est la synthèse de multiples informations, qu'elles soient ponctuelles (stations de vélo par exemple) ou étendues (les informations économiques sur les arrondissements de Paris). On peut retrouver la liste des données utilisées par l'application web dans les mentions légales du site.

Il s'agit là d'une belle représentation graphique des données Open Data et une démonstration de leur diversité.

b. City Dashboard

Revenons maintenant sur l' "Open Data Transport", dont on a beaucoup parlé dans la partie « applications web et mobiles ». Pour mémoire, de très nombreuses applications mobiles s'appuient sur des données libres orientées transport (routier, en commun, ...). Mais il existe d'autres outils permettant de consulter ces données, et ce de manière plus originales ! Certains vous paraîtront d'ailleurs futuristes, mais tout ceci est bien réel.

Le premier, le [City Dashboard](#), est tout simplement une page web, sous forme d'un tableau de bord. Il s'agit d'un unique écran présentant différentes données pour la ville de Londres.

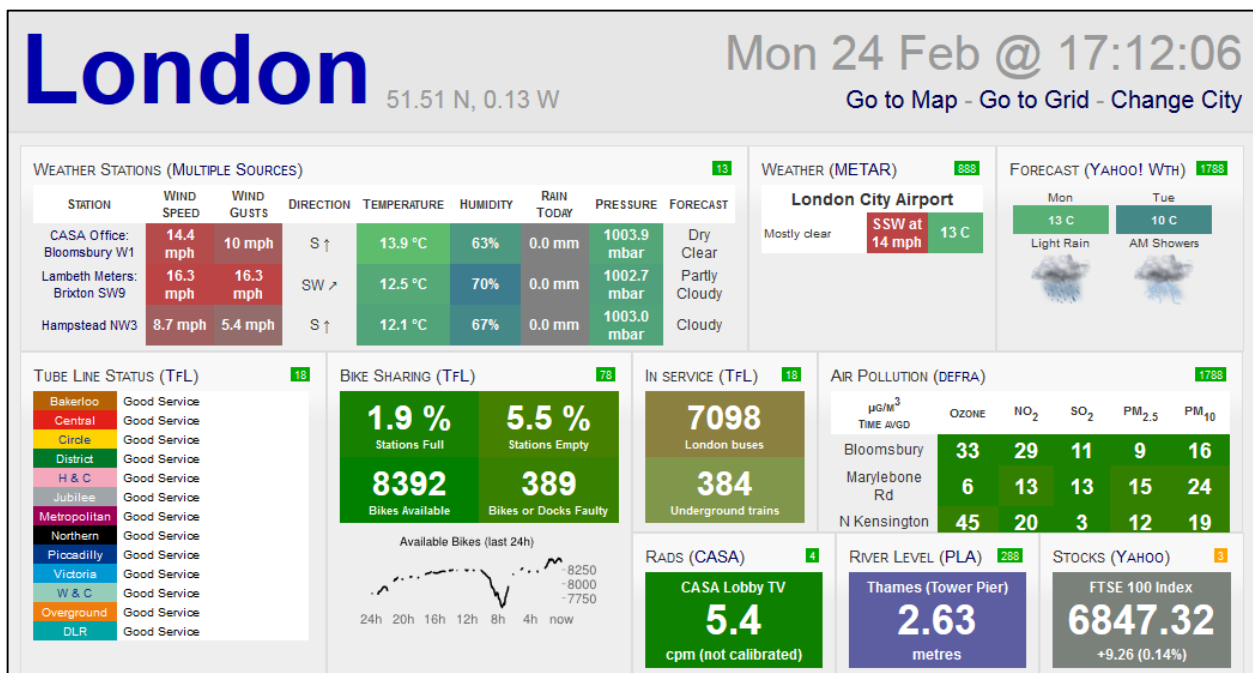


Figure 3 - Interface du City Dashboard

Ce tableau de bord a été développé par des universitaires londoniens, en s'appuyant sur de nombreuses données ouvertes, toutes en temps réel.

Parmi les données affichées, on peut citer le nombre de trains et bus en service, les perturbations dans le métro, le nombre de vélos disponibles, ... mais aussi des données non inhérentes au transport, telles que la pollution de l'air, le temps et la température, le niveau de l'eau, les nouvelles journalistiques, ...

La fréquence de mise à jour dépend du type de données, variant ainsi de quelques secondes (pour les photos du trafic routier ou la météo) à 30 minutes (pour les prévisions météo ou la pollution de l'air).

Sachez que tous les éléments de ce tableau de bord sont déplaçables à souhait (tout au moins sur un ordinateur, l'opération n'étant pas possible sur navigateur mobile). Les Londoniens ont ainsi toutes leurs données préférées sous la main, et ce très rapidement. Sans oublier qu'elles sont tous temps réel. Il est par ailleurs possible d'afficher les données voulues dans n'importe quel lieu public, via un simple écran de PC.

c. La montre intelligente « Muni Watch »

Un deuxième moyen d'avoir accès à des données de transport : une montre qui informe son propriétaire sur les horaires de bus ! "[Muni Watch](#)" (tirant son nom du réseau de transport de San Francisco) indique en effet les trois prochains horaires de passage pour l'arrêt de bus le plus proche.

Pour cela, elle communique en Bluetooth avec un smartphone Android, lequel fait tourner le code permettant de récupérer l'arrêt le plus proche et les horaires associés.

Petit bémol cependant, cette montre est un exemplaire unique : émanant d'un hobby personnel, Joe Hughes l'a développée pour lui et lui seul. Ce projet datant de 2009, il est d'ailleurs étonnant qu'aucune montre intelligente n'ait été mise en vente pour le grand public... Est-ce dû à une question économique ? Ou bien à un intérêt relativement limité, qui n'attire pas le grand public ?

d. La boîte

Dans un autre genre, Baptiste Gaultier travaille sur le projet "[La boîte](#)" depuis 1 an et demi, à ses heures perdues. Il s'agit en fait d'une petite boîte qui donne l'heure, la météo de demain, le nombre de minutes avant le prochain bus, le nombre d'emails non lus, les tweets, ... et ce n'est que la version 0.2 !

Du point de vue fonctionnement technique, la boîte est connectée à un serveur contenant une application web, celle-ci permettant de mettre à jour les boîtes et d'y installer de nombreuses applications. Ainsi, il est possible d'interagir avec le serveur grâce à son ordinateur ou smartphone, afin de modifier les informations affichées par chaque boîte connectée au serveur.



Figure 4 - « La Boite » de Baptiste Gaultier

Ce projet Open-Source est en ligne sur laboite.cc. Si vous voulez faire un test ou bien aider Baptiste Gaultier, ingénieur R&D rennais, dans son projet, allez y jeter un œil !

e. La data-journalisme

Face à l'émergence des données, les journalistes se sont adaptés et certains se proclament comme étant des data-journalistes. Ce mouvement, particulièrement présent en Angleterre, se trouve une caisse de résonance par le magazine *The Guardian* qui dispose d'une rubrique dédiée à la mouvance du data-journalisme.

Un de ses journalistes, [Simon Rogers, est intervenu lors d'un TEDx](#) au Panthéon-Sorbonne en déclarant "Les data-journalistes sont les nouveaux punks". Il argumente son discours en expliquant qu'être punk, tout monde pouvait le faire, tout comme faire le boulot de data-journaliste. En effet, il a présenté les outils que lui et ses collègues utilisaient pour leur travaux : de simples tableurs à des outils Google (ex : *Google Fusion Table*), les outils sont à portée de main du grand public.

Traiter un tel ensemble de données requiert de prendre du recul mais permet aussi de mieux comprendre le monde qui nous entoure. Simon a ainsi produit une carte des dépenses de toutes les différentes entités du gouvernement anglais. Une information utile mais pourtant bien cachée derrière une myriade de données.

Le data-journalisme, c'est donner du sens aux données. Dans le même domaine, David McCandless, un journaliste designer, est un des pionniers dans la représentation des données façon graphique. Il a notamment écrit deux livres sur le sujet, regorgeant de pas mal de ses créations (*Information is Beautiful* est son dernier).

En faisant un tour sur son site, on peut s'amuser avec quelques créations graphiques interactives. On trouve, par exemple, une carte sur les [plus grosses fuites de données utilisateurs](#) :



Figure 5 - « Cartographie des fuites de données utilisateurs »

D'un point de vue plus global encore, on trouve déjà des sociétés de design spécialisées dans la synthèse de données massives. C'est le cas de l'agence, [DatavEyes](#). Que cela soit par la mise en place d'un design interactif (ou non), voire de motion design (graphismes animées en vidéos), la société surfe sur le phénomène et a ainsi pu créer une [carte du métro parisien](#) "permettant d'interroger les temps de transports en métro à Paris, et de montrer l'affluence dans les stations de métro parisien." Vraiment bien faite et qui donne envie de s'y plonger ! D'ailleurs, en parlant de motion design, l'agence s'est faite une [vidéo promotionnelle](#) très sympa qui explique bien l'enjeu de l'exploitation des données.

Enfin, si le sujet du data-journalisme vous intéresse, nous ne pouvons que vous recommander chaudement ce blog de veille : <http://jeanabbiateci.fr/blog/> où de nombreux autres exemples sont présentés (comme un sur le match des étoiles sur [Allocine](#)).

III. Les limites de l'Open Data

L'Open Data prend de plus en plus d'ampleur depuis quelques années, les données se multipliant, la communication y ayant trait se développant, ... Les sociétés ou entités gouvernementales publient chaque jour davantage de données pour montrer au monde qu'elles s'inscrivent dans une démarche Open Data. Cependant, il convient de s'interroger sur ces données et de se demander jusqu'où va la démarche Open Data. Autrement dit, quelles sont les limites des données ouvertes ? Nous verrons que là encore, du progrès est à faire ...

a. Des données non triées

La première critique que l'on peut formuler sur l'Open Data est le côté fouillis qui l'entoure. En effet, les plates-formes de données sont souvent peu structurées, à l'exception récente de data.gouv.fr qui a beaucoup progressé depuis 1 an. Ce site dispose, à l'inverse des initiatives locales/régionales, d'un aspect communautaire, ce qui permet de noter l'utilité des jeux de données. De quoi défricher un peu la jungle que sont les plates-formes de données...

Le principal reproche que l'on peut faire à l'Open Data est l'amoncellement de données non cohérentes entre elles. Certes, c'est pour la bonne cause, pour la transparence et pour la diversité mais cela requiert un tri possible et efficace pour l'internaute. Ce n'était pas le cas pour la plate-forme Nantaise.

Désormais un tri est possible mais pas très conviviale. Et oui, l'interface des portails Open Data est très souvent austère : un néophyte ne voudra jamais se plonger là-dedans. Ce n'est peut-être pas anodin... Les éditeurs de données mettraient alors en avant des données pour empêcher les visiteurs de tomber sur des données plus critiques, mises en ligne par obligation (principalement dans le secteur public).

b. L'absence de normes

L'abondance des données semble cacher encore d'autres lacunes. Celle qui revient le plus souvent est l'absence de normes et l'utilisation de formats de fichiers peu adaptés. Pour mieux illustrer ce propos, prenons un exemple. Sur le site data.gouv.fr, on peut trouver la [liste complète des gouvernements de la Vème république](#). Deux choses clochent :

- le nom du fichier est tout sauf descriptif du contenu "Ressource sans nom"
- le format du fichier : en .rtf. Un tel format ne permet pas de faire tourner un algorithme : il n'est pas structuré ni formalisé.

L'imposition d'une norme pour le partage des informations de type Open Data permettrait de faire évoluer les choses dans le bon sens...

c. Des données ouvertes, mais à ne pas voir ...

Une autre critique concerne les données qui sont ouvertes (par obligation), mais qu'on cherche à cacher le plus possible.

C'est le cas des données sur l'insécurité rendues publiques par le gouvernement, qui manifestement ne tient pas à ce qu'on consulte ou analyse ces données. En effet, les data-journalistes qui ont publié le baromètre, ville par ville, de la délinquance en France, ont effectué un véritable partout du combattant !

Premièrement, le ministère de l'Intérieur ne publie lui-même aucun chiffre. Quant à l'ONDRP (Observatoire National de la Délinquance et des Réponses Pénales), il ne publie que des statistiques départementales, et non ville par ville ...

Les data-journalistes ont alors dû se plonger dans un rapport annuel publié par DCPJ (direction centrale de la police judiciaire). Parcourant ces 420 pages uniquement accessibles sous format PDF, les data-journalistes ont recopié une à une les données dans un document Excel... Tâche laborieuse qui leur a pris plusieurs jours ! La DCPJ pourrait évidemment transmettre ces données aux médias sous la forme d'un document Excel mais elle refuse catégoriquement de le faire. On se demande pourquoi...

d. Une ouverture partielle des données gouvernementales

Encore une déception : l'ouverture des données du gouvernement n'est que partielle. Certains types de données sont encore cadenassés, malgré la démarche d'ouverture du gouvernement.

C'est le cas des données culturelles, qui font partie des exceptions au cadre de l'Open Data fixées par la loi [Cada](#) de 1978. Par données culturelles, entendez les données produites par "*des établissements et institutions d'enseignement ou de recherche*" ou par des "*établissements, organismes ou services culturels*".

Le régime particulier d' "*exception culturelle*" permet à ces établissements de fixer les conditions de la réutilisation de leurs données, tandis que les autres administrations relèvent du régime général de cette loi, qui instaure un droit à la réutilisation des informations publiques au profit des citoyens.

Très peu d'administrations tendent donc à publier leurs informations culturelles. Contre-exemple tout de même : on peut féliciter la ville de Toulouse qui a rendu public ses

archives municipales, en publiant [un nouveau règlement relatif à la réutilisation de leurs données](#).

e. Ouverture de données limitée par les entreprises

Il convient maintenant de se demander pourquoi les entreprises n'ouvrent pas toutes leurs données. Une réponse possible est la perte de clients en cas d'ouverture des données.

Petit exemple : supposons que la SNCF rende public ses données concernant les horaires des trains. Plusieurs applications se basant sur ces données seraient très certainement créées. *Google Maps* pourraient par exemple proposer un outil de recherche des trains à prendre pour effectuer tel ou tel itinéraire. Ces nouvelles applications vont alors capter le public, et faire perdre des opportunités à la SNCF...

f. La faible quantité de données temps réel

Dans le domaine des données manquantes, passons au cas des données temps réel, toujours très peu présentes, malgré l'omniprésence de l'Open Data dans les transports. Intéressons-nous à la ville de Nantes, où l'estimation du temps restant avant le prochain tram est désespérément absente du site d'Open Data nantais. Pourtant cette donnée existe car elle est affichée sur les panneaux le long des arrêts des tramways. Le plus étonnant est que cette information est aussi absente de l'application *Tan* pour smartphones, qui se base sur les données libres mises à disposition par l'entreprise. Est-ce de la paresse ou une indisposition technique qui empêche une telle donnée cruciale d'être disponible ? Nul ne sait.

Dans la même optique, la *Muni Watch*, dont nous évoquons son existence dans la partie précédente, s'est également frottée à l'épineux problème des données temps réelles. Depuis son [blog](#), l'inventeur de cette montre a expliqué son combat avec la société de transport en commun de Los Angeles, *MTA*, pour obtenir le droit d'accès aux données temps réel du réseau. Il a réussi à obtenir pour lui mais aussi pour les autres développeurs l'accès à cette donnée en 2009. Si cette avancée a été possible, pourquoi pas dans d'autres villes ? Même à Paris, cette donnée n'existe pas (l'application *Watch Dogs* interpolant simplement les données à partir des horaires de passage théorique, ce n'est pas du temps réel).

g. Des données non permanentes

Pour finir, un dernier aspect sur la limite des données Open Data : la publication de certaines données mais non réutilisables ! Pire, des applications s'appuyant sur certaines données publiques se sont vues fermées... Deux exemples pour illustrer cela.

Le premier concerne la fermeture, un an et demi après son ouverture en mai 2011, du comparateur de tarifs médicaux *“fourmisanté”* qui vous donne la carte des médecins à proximité, leurs prix et ce qui reste à charge. Ce site utilise les données publiques présentes sur les sites d'organismes officiels : ordre des médecins, CNAM, mutuelles santé et ministère de la santé. Primée pour l'innovation de cette ouverture de données publiques, cela n'a pas empêché la CNAM de demander le retrait de ces données selon elle *« manipulées »*.

La ministre de la santé, Marisol Touraine, s'est exprimée sur ce sujet en expliquant *« qu'il revient aux pouvoirs publics de rassembler ces informations sur les hôpitaux, pour les rendre plus accessibles et plus transparentes »*. En d'autres termes, l'open data, c'est très bien, mais sous contrôle de l'administration...

Deuxième exemple, la fermeture totale et définitive du site Dentistedegarde.net. Le service, s'appuyant sur un partenariat avec l'ordre des chirurgiens-dentistes de Loire-Atlantique, permettait de trouver le dentiste de garde le plus proche. Le CHU de Nantes redirigeait vers ce service depuis son site internet et dentistedegarde.net a reçu plus de 18 000 visites en moins d'un an.

Fin 2012, l'ordre a mis fin à son partenariat, indiquant aux développeurs qu'il devait cesser de fournir la liste des gardes. Le remaniement dans la diffusion des gardes a été demandé par le ministère de la Santé afin d'organiser une redirection générale vers le Samu. C'est le 15 qui orientera les patients auprès des praticiens ou hôpitaux selon les besoins.

D'après Jérôme Mousseau, Président de l'ordre, cette volonté de remaniement répond à un manque d'informations sur le service de garde : *« Beaucoup de gens ne savent pas qu'il y a un service de garde tous les dimanches matins et tous les matins des jours fériés »*. Ce qui est tout à fait paradoxale : on peut en effet se demander en quoi supprimer l'information en ligne et la restreindre au 15 permettra au public de mieux connaître l'existence de ces services...

Se pose aussi la question de l'ouverture de données d'utilité publique (ayant par exemple trait à la santé, la sécurité, la culture ou les finances), qui restent jusqu'alors en marge de l'Open Data. C'est ce genre de données qui devraient être ouvertes, et non celles dont tout le monde se fichent...

Conclusion

Au premier abord, l'Open Data fait peur notamment avec sa farandole de formats mais aussi ses licences. Néanmoins, une fois passé ces barrières, on ne peut que saluer cette initiative qui veut donner plus de visibilité aux données pour les citoyens. Notons que la majeure partie des données est mise en ligne par les collectivités via leur plate-forme de données. Le principal objectif est la transformation de ces données, par des intermédiaires (développeurs, journalistes, etc.), pour en faire un service aux citoyens.

Cependant, il ne faut pas se leurrer, derrière ce partage de données se cache un aspect économique important, que cela soit pour l'Etat, les entreprises ou les entités réutilisant ces données. Preuve de bonne santé d'une entreprise, l'ouverture de ses données permet de créer un écosystème autour : des applications notamment qui auront été créées par des ingénieurs/développeurs. Une sorte de relance économique par l'IT.

Si la plupart des applications tournent autour des places de parking et des transports, il existe tout de même des utilisations originales comme le data-journalisme ou encore des synthèses d'informations diverses en un même lieu (citons le site *We Are Data* ou encore le *City Dashboard*).

Enfin, nous pourrions conclure avec un constat des limites de l'Open Data. C'est surtout le côté « jungle » dans les données (non triées, peu utiles) qui rebutent et qui nous a poussé à vérifier si ce surplus d'informations ne cachaient pas un manque de données cruciales.

Et cela s'est malheureusement révélé être le cas : les données en temps réel sont encore très peu présentes. A cela s'ajoute l'absence de normes et une ouverture pas si franche des entreprises et de l'Etat.

Tout n'est pas si noir et des efforts ont été faits pour améliorer les plates-formes de données, en les rendant plus agréable à la navigation. Est-ce pour mieux cacher le manque de données ... ?

Sources

1. Introduction

- [10 principes de l'open Data](http://cursus.edu/institutions-formations-ressources/formation/17814/lexique-open-data/) : <http://cursus.edu/institutions-formations-ressources/formation/17814/lexique-open-data/>
(site visité le 04/10/2013)

2. Contexte de l'Open Data

a. La nature des données

- [Plateforme open Data du gouvernement](http://www.data.gouv.fr/) : <http://www.data.gouv.fr/>
- [Licence Ouverte](http://www.etalab.gouv.fr/pages/licence-ouverte-open-licence-5899923.html) : <http://www.etalab.gouv.fr/pages/licence-ouverte-open-licence-5899923.html>
- [Etalab](http://www.etalab.gouv.fr/) : <http://www.etalab.gouv.fr/>
- [Open Database License \(ODbL\)](http://opendatacommons.org/licenses/odbl/) : <http://opendatacommons.org/licenses/odbl/>
- [Open Data Commons](http://opendatacommons.org/) : <http://opendatacommons.org/>
- [Blog liberTIC](http://libertic.wordpress.com/) : <http://libertic.wordpress.com/>
- [Blog Open Data Touraine](http://opendata-touraine.fr/de-lopen-data-au-bien-commun-une-question-de-licence/) : <http://opendata-touraine.fr/de-lopen-data-au-bien-commun-une-question-de-licence/>
- [Le citoyen a-t-il une place dans l'open data ? \(figure 1\)](http://libertic.wordpress.com/2012/02/17/le-citoyen-a-t-il-une-place-dans-lopen-data/) : <http://libertic.wordpress.com/2012/02/17/le-citoyen-a-t-il-une-place-dans-lopen-data/>
(sites visités le 25/10/2013)

b. L'aspect économique

- [Ouverture des données publiques : les exceptions au principe de gratuité sont-elles toutes légitimes ?](http://fr.scribd.com/doc/181747389/2013-08-26-rapport-annexes-pdf) : <http://fr.scribd.com/doc/181747389/2013-08-26-rapport-annexes-pdf>
- (sites visités le 21/11/2013)

3. Les utilisations de l'Open Data

a. Les applications web et mobiles

- <http://data.nantes.fr/applications/>
- http://data.nantes.fr/applications/detail/?tx_icsodappstore_pi4%5buid%5d=45
- <https://data.cityofchicago.org/Transportation/Relocated-Vehicles/5k2z-suxx>
- <http://www.lesechos.fr/01/06/2012/LesEchos/21196-025-ECH --open-data-----les-collectivites-sont-en-quete-d-un-modele-economique.htm>
- <http://pro.clubic.com/technologie-et-politique/article-560928-2-open-data-ouverture-ameliorer-societe.html>
- <http://statiophone.fr/>
- <http://donneesouvertes.info/category/articles/economie-et-strategie-open-data/>
- <http://pro.clubic.com/actualite-e-business/actualite-597650-donnees-publiques-etat-rapport.html>
- <http://www.eurooptimist.eu/concours-open-data>
- (sites visités le 20/12/2013)

b. De nouvelles applications

- <http://wearedata.watchdogs.com>
- [City Dashboard : http://livepage.apple.com/](http://livepage.apple.com/)
- [Muni Watch : http://livepage.apple.com/](http://livepage.apple.com/)
- [laboite.cc : http://livepage.apple.com/](http://livepage.apple.com/)
- [Intervention Simon Rogers \(TEDx\) : http://veille-techno.blogs.ec-nantes.fr/wp-admin/www.youtube.com/embed/h2zbvmXskSE](http://veille-techno.blogs.ec-nantes.fr/wp-admin/www.youtube.com/embed/h2zbvmXskSE)
- [Site David McCandless : http://www.davidmccandless.com/](http://www.davidmccandless.com/)
- [Agence DatavEyes : http://www.dataveyes.com/](http://www.dataveyes.com/)
- [Carte du métro parisien interactive : http://metropolitain.io/](http://metropolitain.io/)
- [Vidéo promotionnelle de DatavEyes : http://vimeo.com/30144927](http://vimeo.com/30144927)
- [Les plus grosses fuites de données utilisateur \(figure 5\) : http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/](http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/)
- <http://jeanabbiateci.fr/blog/>
- [Match des étoiles Allocine: rue89.nouvelobs.com/rue89-culture/2013/10/19/allocine-presse-public-saffrontent-a-coup-detoiles-246717](http://rue89.nouvelobs.com/rue89-culture/2013/10/19/allocine-presse-public-saffrontent-a-coup-detoiles-246717)
- (sites visités le 24/01/2014)

4. Les limites des applications

- <http://www.01net.com/editorial/599467/orange-investit-le-stationnement-urbain-intelligent-sur-la-voirie/>
- <http://123opendata.com/blog/open-data-service-retrouver-voiture-fourriere/>
- <http://www.entreprenantes.com/magazine/481-open-data-les-developpeurs-toujours-brides-par-la-tan.html/>
- http://www.gfi.fr/gfilabs/apps_nantes.php
- (sites visités le 20/12/2013)

5. Limites des données ouvertes

- <http://www.data.gouv.fr/fr/dataset/composition-des-gouvernements-de-la-ve-me-republique-1959-2013-30382998Liste%20complète%20des%20gouvernements%20de%20la%20Vème%20république:%20http://www.data.gouv.fr/fr/dataset/composition-des-gouvernements-de-la-ve-me-republique-1959-2013-30382998>
- <http://www.e-administration-collectivites.fr/glossaire/cadaCada:%20http://www.e-administration-collectivites.fr/glossaire/cada>
- Règlement relatif à la réutilisation de leurs données : [http://www.archives.mairie-toulouse.fr/index.php?id=40&tx_ttnews\[tt_news\]=299&cHash=2f3efc7c73](http://www.archives.mairie-toulouse.fr/index.php?id=40&tx_ttnews[tt_news]=299&cHash=2f3efc7c73)
- Open Data culturel : ouverture des archives toulousaines : <http://scinfolex.com/2013/09/29/open-data-culturel-enfin-des-archives-libres-sous-le-signes-du-partage-a-lidentique/>
- Insécurité, des données publiques bien cachées : http://www.lexpress.fr/actualite/societe/insecurite-des-donnees-publiques-mais-bien-cachees_1301248.html
- Couvrez ces données que l'on ne saurait voir : <http://libertic.wordpress.com/2012/11/03/couvrez-ces-donnees-que-lon-ne-saurait-voir/>
- Discours de la ministre de la santé, Marisol Touraine : <http://www.social-sante.gouv.fr/actualite-presse,42/discours,2333/congres-de-la-mutualite-francaise,15303.html>
- (sites visités le 14/02/2014)

Table des figures

Figure 1 - La place du citoyen dans le mouvement Open Data	8
Figure 2 - Interface du site "We Are Data" pour la ville de Paris.....	16
Figure 3 - Interface du City Dashboard	17
Figure 4 - « La Boite » de Baptiste Gaultier.....	19
Figure 5 - « Cartographie des fuites de données utilisateurs »	20