



HAL
open science

OPEN DATA EN TRANSPORT URBAIN : QUELLES SONT LES DONNEES MISES A DISPOSITION ? QUELS SONT LES STRATEGIES DES AUTORITES ORGANISATRICES ?

Catherine Bouteiller, Sybille Berjoan

► **To cite this version:**

Catherine Bouteiller, Sybille Berjoan. OPEN DATA EN TRANSPORT URBAIN : QUELLES SONT LES DONNEES MISES A DISPOSITION ? QUELS SONT LES STRATEGIES DES AUTORITES ORGANISATRICES ?. 2013. halshs-00838632

HAL Id: halshs-00838632

<https://shs.hal.science/halshs-00838632>

Preprint submitted on 26 Jun 2013

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

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

OPEN DATA EN TRANSPORT URBAIN : QUELLES SONT LES DONNEES MISES A DISPOSITION ? QUELS SONT LES STRATEGIES DES AUTORITES ORGANISATRICES ?

C. Bouteiller,

Consultant-chercheur, Mobilité & Recherches

Doctorante au Laboratoire d'Economie des Transports, Lyon

S. Berjoan,

Accenture, Directrice Monde des Etudes Stratégiques Transport, Paris

Juin 2013

Le phénomène de l'ouverture des données a commencé en Angleterre et aux Etats Unis dans les années 2010 avec la mise à disposition de données sur des sites gouvernementaux comme data.gov.uk en Grande Bretagne et data.gov aux Etats-Unis. Le portail américain donne accès à près de 400 000 jeux de données publiés par 180 administrations¹. Le président Barack Obama a publié le 09 mai 2013, un décret stipulant que « les données générées par le gouvernement sont rendues disponibles dans des formats ouverts et pouvant être interprétés par les machines, tout en protégeant lorsque nécessaire, la sécurité, la confidentialité et la vie privée».²

En France, l'idée de la mise à disposition des données publiques n'est pas nouvelle puisque par la loi du 17 juillet 1978, toute personne a un droit d'accès aux documents administratifs et un droit à la réutilisation des données publiques. Une directive européenne du 17 novembre 2003 concernant la réutilisation des données a été transposée en France à travers l'ordonnance du 06 juin 2005 et le décret du 30 décembre 2005 pour créer un droit de réutilisation de l'information à partir de licences.

Plutôt que de garder les données enregistrées quotidiennement par les services publics, il s'agit de les mettre à disposition de la communauté scientifique des développeurs ou du grand public qui pourrait rapidement y trouver un intérêt, voir contribuer à élaborer de nouveaux services créant ainsi de la valeur. L'objectif est d'associer les citoyens de manière transparente, collaborative et participative.

Facilitée par le web 2.0³, de nombreux pays, villes et même entreprises publiques ont contribué à cette mise à disposition de données. Différents portails open data ont fait leur apparition sur la toile. L'objet de cet article est de porter une attention particulière à la nature des données mises à disposition dans le domaine du transport urbain sur les sites des Autorités organisatrices de transport ou les sites d'opérateurs.

Dans un premier temps les données de mobilité mises à disposition en open data, à Paris, Londres, New York, Toronto, Madrid et Singapour seront passées en revue. L'objectif est de

¹ <http://www.lemagit.fr/technologie/gestion-des-donnees/2013/05/14/ladministration-obama-structure-officiellement-sa-politique-publique-liee-a-lopen-data/>

² <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-/>

³ Réseau Web interactif grâce à des interfaces utilisateurs facilitant le traitement des informations

caractériser les « standards » en terme de nature des données transport en open data. Cette analyse détaillée des données permettra dans un deuxième temps de mettre en évidence des stratégies d'acteurs publics ou d'opérateurs.

I - Des données de mobilité en libre-service mais lesquelles ?

Dans le secteur des transports publics de nombreuses données ont été mises à disposition sur les sites des opérateurs de transport, comme les sites des autorités organisatrices ou encore sur les sites des Villes. En matière de transport, il faut s'attacher à collecter les données relatives à la mobilité, c'est-à-dire au transport public : bus, métro, train mais également à la voirie, et au trafic routier et autoroutier, au vélo et enfin aux piétons. La quantité et la nature des données mises à disposition ont été recensées à partir des sites internet d'open data des gouvernements, des autorités organisatrices ou des opérateurs, en mai 2013.

Simon Chignard (2012) dans son livre « open data comprendre l'ouverture des données publiques » a classifié ces données en fonction de 2 axes. Il définit les données statiques et les données dynamiques. Il s'est attaché à déterminer si leur mise à disposition avait plutôt un objectif de transparence ou de recherche de contribution participative et a identifié une très grande diversité de données ouvertes. Dans l'étude présente, il s'agit de mettre en évidence et d'identifier :

- le type de données
- le degré d'accessibilité des données

Le classement des données réalisé dans l'étude présentée dans cet article a abouti à une segmentation en 4 catégories différentes. Il y a d'abord les données d'informations descriptives des réseaux : prix, produits, horaires théoriques. Il y a également les données de type géographique : emplacements géocodés, cartes des réseaux, places de parking Puis il est possible de recenser des données de type statistique telles que le nombre de voyages, le nombre d'entrants par station. Enfin, on recense des données en temps réel mises à disposition par flux avant tout traitement statistique.



Figure 1 : une segmentation des données en 4 catégories

Des premiers résultats montrent une surreprésentation des données d'information descriptives et de cartographie dans les données transport mises à disposition.

Dans chaque ville étudiée, il a été possible d'identifier, pour chaque catégorie de donnée, quels étaient les modes les plus représentés. Cinq modes ont été retenus : le bus, les modes ferrés rail et train, le vélo, la route, et les bateaux, notamment pour la ville de Londres qui recense des données sur les déplacements en navette fluviale. On peut s'apercevoir très rapidement que les données relatives au bus et au métro sont surreprésentées notamment dans les catégories de données d'informations descriptives et cartographiques.

Dans certaines villes, certaines catégories sont rares ou inexistantes. A Paris, il y a beaucoup de données d'informations descriptives ou cartographiques mais très peu de statistiques et aucune donnée en temps réel. A Madrid, les données délivrées en temps réel sont rares et aucune donnée statistique n'est ouverte. New York propose de très nombreuses données dans toutes les catégories sauf dans la catégorie des données d'informations descriptives, sans doute car elles sont, par ailleurs, disponibles pour tous en dehors des sites d'open data.

La nature des données mises à disposition est très différente d'une ville à l'autre.

Pour identifier les villes les plus avancées dans la mise à disposition de données, la quantité de données transport a été comparée au nombre de données transport délivrées en temps réel (figure 2).

- ✓ La quantité de données disponible est portée sur l'axe des abscisses de la figure 2. Elle n'est pas un critère suffisant. Ce qui fait réellement la différence en matière de plus-value apportée par les données dans des applications « grand public » ou professionnelles, c'est la qualité de la donnée, et notamment de la donnée délivrée en temps réel. C'est, en effet, à partir de celles-ci que pourront être envisagées les applications apportant des notions de prédiction de trafic et une réelle plus-value par rapport à une donnée descriptive théorique.
- ✓ C'est la donnée temps réel qui contribue le plus à la valeur des applications développées. Dans la figure 2, l'axe des Y représente le nombre de données disponibles en temps réel alors que la taille des bulles donne une idée du degré d'accessibilité de la donnée. Cette dernière peut être disponible quasiment immédiatement à partir d'un site internet comme à Paris ou à travers un processus d'identification et de déclaration d'intention, voire de license, comme à Madrid ou Toronto.



Figure 2 : classement des villes en open data

- ✓ Une première lecture des figures 2 et 3, montre que Singapour, Londres et New York mettent à disposition des données en grand nombre. Par ailleurs toutes les catégories sont représentées notamment celle des données en temps réel. Paris, Madrid ou Toronto, restreignent encore ces dernières à certains modes ou à certaines catégories.

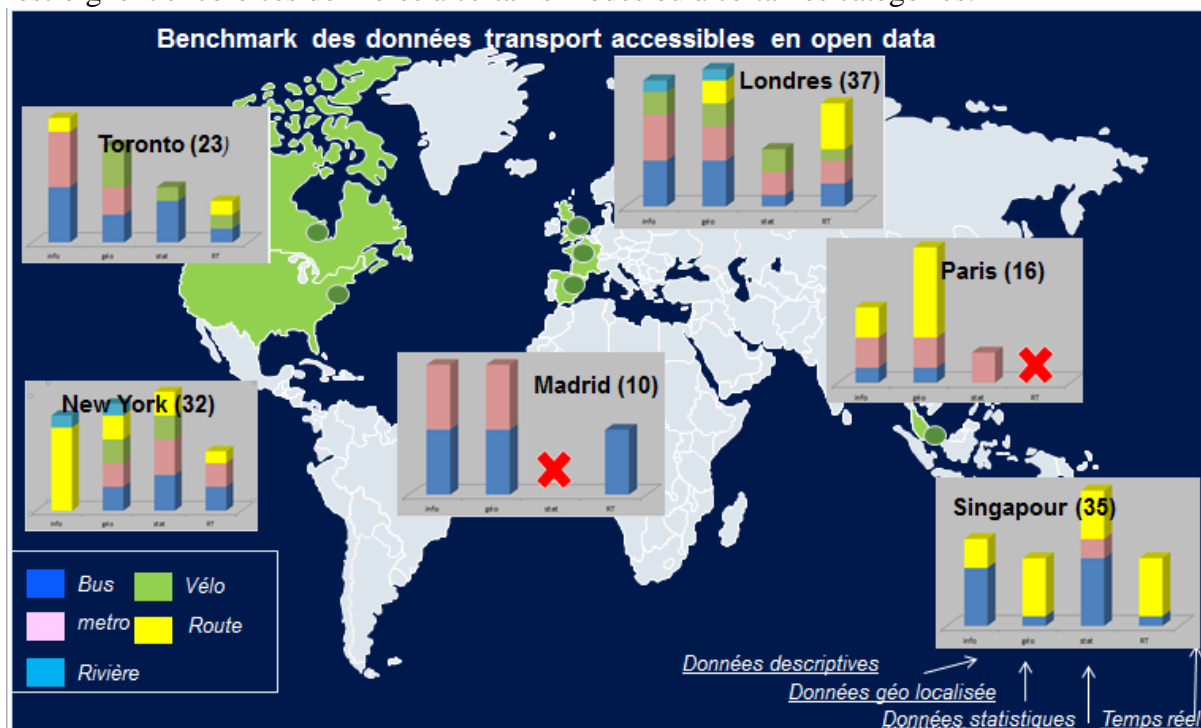


Figure 3 : classement des villes en open data

Une lecture approfondie de la figure 3 permet de mieux comprendre la stratégie de chacun :

- ✓ Sur les 16 données transports recensées pour Paris, 9 sont des données cartographiques : les données telles que la localisation des places de livraison, la localisation du mobilier urbain, les cartes de réseaux, les rues sont parfaitement disponibles ainsi que les données d'information descriptives générales sur les horaires théoriques. C'est cependant la seule ville de l'étude où aucune donnée en temps réel n'a pu être collectée. En matière statistiques, les données sont très parcellaires: statistiques annuelles de fréquentation des stations de métro et qualité de l'air par trimestre...
- ✓ Madrid met à disposition peu de données, une dizaine, dont 2 en temps réel exclusivement dédiées au bus. Madrid a fait le choix de ne pas mettre à disposition de statistiques.
- ✓ Avec près de 23 données disponibles et plusieurs modes représentés, Toronto dénote par le fait qu'il n'y ait que 3 données en temps réel : l'activité vélo, les horaires de bus réel et les images de trafic. Par contre toutes les catégories sont représentées avec beaucoup de données d'informations descriptives que ce soit sur les parkings, le réseau vélo, les bus, des données cartographiques de réseaux, des données statistiques de fréquentation mensuelle et annuelle en heure de pointe et en heure creuse.
- ✓ Dans le peloton de tête des villes mettant le plus de données à disposition, Singapour surprend. En effet alors que NYC et Londres mettent quantité de données à disposition à la fois, pour tous les modes et dans toutes catégories, Singapour a clairement ouvert les données liées au trafic routier et au réseau de bus. De très nombreuses statistiques de trafic sont disponibles, très certainement du fait que l'autorité organisatrice est également en

charge du péage urbain. Par ailleurs LTA a fait le choix de développer elle-même un certain nombre d'application en temps réel pour ses réseaux de transport urbain. Ainsi, toutes les données ne sont pas mises à disposition de la communauté des développeurs. La plupart des données en temps réel, concernent le trafic routier : vitesse instantanée, image caméra en temps réel du trafic sur les autoroutes.

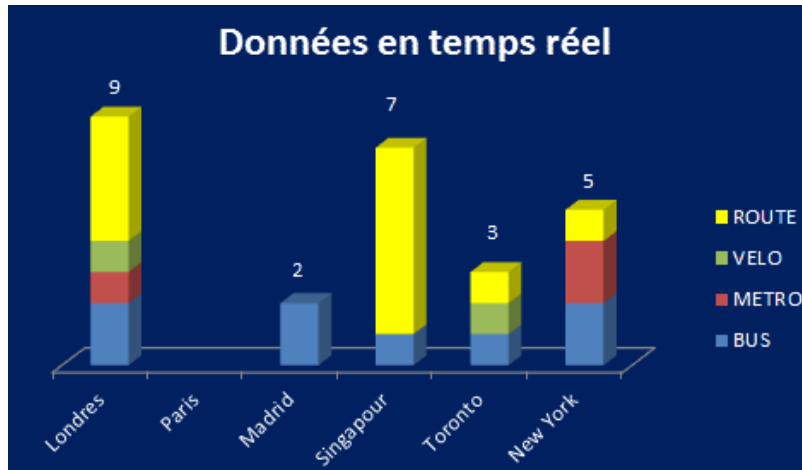


Figure 4: des données temps réel parcellaires

Si on regarde les données disponibles en temps réel uniquement (figure 4), on note que les données sont rares, une dizaine au maximum. Le nombre de données par mode, est faible et les données relatives au réseau de bus semblent être un incontournable notamment l'horaire d'arrivée estimé. Les données de trafic viennent en deuxième position avec les images de trafic routier. Pour compléter cette étude, il faut se demander quelles sont les utilisations possibles de ces données et chercher à comprendre si la mise à disposition de ces dernières permet de créer de la valeur et pour qui ?

En conclusion on peut dire qu'il y a essentiellement des données d'informations descriptives et cartographiques disponibles en open data. Les données statistiques sont peu abondantes et les données temps réel sont rares. Ces dernières concernent essentiellement les bus et le trafic routier. Les horaires estimés des bus et les images des caméras filmant la circulation sont les données disponibles les plus courantes dans cette catégorie.

II -3 profils d'autorités organisatrices dans la stratégie de mise à disposition des données

Le retour d'expérience Etalab (2011)⁴ met en évidence le foisonnement d'acteurs pouvant utiliser les données : chercheurs, journalistes, développeurs, entreprises privées développant des applications pour le grand public, mais également sociétés privées désireuses de faire du géomarketing, ou collectivités, pour de l'aide à l'exploitation. S.Chignard (2012) a défini 4 classes de réutilisation : la « consultation des données », la « réalisation d'applications », la « médiation » qui n'est autre que la représentation de la donnée brute, et la « réutilisation spécialisée » en business to business.

⁴ Etalab est le service du Premier Ministre, chargé de l'ouverture des données publiques et du développement de la plateforme française Opendata

Il s'agit de mettre en évidence les stratégies qui sous-tendent la mise à disposition des données. Au vu de l'étude des 6 villes présentées précédemment, on peut dégager 3 tendances par l'analyse de la donnée mise à disposition :

- Les AOT qui ouvrent les données brutes sans condition pour faciliter les apports externes.

Ces dernières font confiance au marché pour produire de la valeur en délivrant une donnée brute non traitée : cette donnée est multiple, toutes catégories, tous modes : c'est le cas de Londres avec TFL (Transport for London) et de NYC avec MTA (Metropolitan Transportation Authority). La donnée est disponible, facilement accessible, non traitée, libre d'accès et de droit : la « médiation » n'est pas aisée car elle nécessite une manipulation par l'utilisateur. Ce dernier est néanmoins libre de pouvoir réutiliser cette donnée et de créer de la valeur dont il récupère les profits. Dans ce type de ville on peut observer un foisonnement d'applications grand public dont le standard est le calculateur d'itinéraire avec horaires réels. L'AOT bénéficie indirectement des retombées positives en termes de modernité et de facilité d'usage du réseau

- Les AOT qui ouvrent partiellement leurs données et produisent de la valeur en interne
- Ces dernières traitent une partie de leurs données en produisant elles même un certain nombre d'applications de visualisation et de médiation. Elles ne mettent à disposition du public qu'une partie de la donnée brute : C'est le cas de Singapour avec LTA (Land Transport Authority) qui a produit en interne des applications ou intégré à son site un service de conciergerie ⁵(figure5). Une partie de la valeur créée est récupérée par celui qui produit la donnée, c'est-à-dire l'autorité organisatrice. L'autorité choisit néanmoins de mettre à disposition un certain nombre de données en temps réel, notamment celles relatives au trafic routier. L'Autorité organisatrice n'a pas forcément toutes les compétences en interne pour explorer tous les champs d'applications possibles et se heurte à des limites en termes de recrutement de profils spécialisés et de développement de compétences dans tous les domaines de la mobilité.



Figure 5 : intégration d'applications au site open data de l'Autorité Organisatrice (LTA, Singapour)

⁵ Voir le site mytransport.sg : <http://www.mytransport.sg/content/mytransport/home/dataMall.html>

- Les AOT qui cherchent un positionnement en vendant une donnée déjà transformée. Celles-ci hésitent à mettre à disposition la donnée par crainte de voir une partie de la valeur potentielle de cette donnée leur échapper. Pour autant, elles hésitent également à mettre en œuvre les investissements qui pourraient potentiellement amener à récupérer une partie de cette valeur. Faut-il retraiter la donnée ? Investir dans du développement ? Dans quel type de projet investir et est-ce encore du domaine de compétence d'une Autorité organisatrice ?



Figure 6

Application en interne de données temps réel
Centre de contrôle du trafic à Los Angeles



Figure 7

Application grand public avec données TR
Application Tripease et London Jamcams

Paris n'a pas mis à disposition de données temps réel. Pour autant le Syndicat Mixte des Transport d'île de France, STIF revend du service avec un modèle de licence et de tarification à l'usage pour le téléchargement par des opérateurs privés de calculs d'itinéraires. C'est ainsi que Mappy, acteur privé de cartographie et de services géo localisés, achète des calculs d'itinéraire au STIF. Les potentiels sont limités car aujourd'hui, le calcul d'itinéraire est devenu un standard et la donnée temps réel prend toute sa valeur quand elle est intégrée. Or, Paris n'ayant pas mis de données en temps réel à disposition et n'ayant pas développé les applications grand public, l'utilisateur parisien n'a pas encore accès à ce type de service comme il pourrait les connaître à Londres (voir figure 7) ou à Singapour. Certes, certains opérateurs comme Transdev ont mis sur le marché des applications utilisant des informations de leur réseau en temps réel, puisqu'elle disposait en interne de la donnée. Malheureusement ces applications ne peuvent couvrir tous les modes avec la même qualité d'information et de service par manque de disponibilité de la donnée sur les autres réseaux. C'est le cas de l'application bus IDF. Les problèmes rencontrés relèvent alors de la gouvernance du partage des données.

En conclusion

Le foisonnement possible des données, mais surtout la possibilité de les traiter en combinaison avec d'autres informations collectées à travers différents canaux laisse entrevoir différents leviers d'action. Les données de mobilité, comme les informations d'ordre démographiques, météorologiques, billettique, une fois combinées et retraitées, permettent d'envisager des applications diverses. Pour mettre en œuvre ces projets des compétences variées sont nécessaires : recherche opérationnelle, géomarketing, design et ergonomie, et bien sûr traitement informatique analytique. Ces projets ne peuvent voir le jour sans une forte implication des acteurs avec une gouvernance forte. Comment produire du service et participer à la création de valeur grâce à l'open data ?

Le Grand Lyon, en faisant partie prenante du projet Optimod participe directement aux côtés de 6 autres intervenants de la sphère universitaire et privée, à la création d'une application

grand public et d'un navigateur pour le fret urbain. Ce projet réalisé à partir de données transverses incluant bus, métro, train, vélo, route, auto partage, co voiturage, parking devrait permettre de prévoir le trafic routier à 1H. Il s'agira, à terme, d'un navigateur multimodal temps réel, proposant des changements d'itinéraire en temps réel sur plusieurs modes... Par ailleurs le projet intègre la mise à jour de l'information nécessaire au gestionnaire des feux en voirie. Dans cet exemple, une collectivité a fait le choix d'investir aux côtés de partenaires multiples, pour utiliser ses propres données et celles d'autres autorités organisatrices, et ainsi produire des services à valeur ajoutée pour le compte du public et pour des entreprises privées. Mais ce type de projets reste encore assez rare. L'open data, in fine, pose le problème de la gouvernance des projets de mobilité dans un contexte de morcellement de la compétence transport.

RÉFÉRENCES

The Economist, 2010, "A Special report on managing information: Data, data everywhere", 25 février 2010

Etalab, Clement JP, Faure GE (2011) retour d'expérience atelier EtaLab, Mairie de Paris

Chignard,S. (2012), L'open data, comprendre l'ouverture des données publiques, Fyp Editions

<http://opendata.paris.fr/opendata/jsp/site/Portal.jsp>

<http://xn--ppnadata-m4a.se/>

<http://datos.gob.es/>

<http://digitaliser.dk/resource/432461>

<http://data.gov.sg/>

<http://amsterdamsmartcity.com/projects/theme/label/open-data>

<http://daten.berlin.de/>

<http://govinfo.nlc.gov.cn/>

<http://www.toronto.ca/open>

<http://data.seoul.go.kr/>

<http://esd.ny.gov/NYSDataCenter.html>

<http://www.edb.gov.SG>

<http://okfn.org/opendata/>

<http://fr.slideshare.net/AgentNumerique/atelier-etalab-retour-dexprieence-open-data-paris>