

## **POLITIQUES DES DONNÉES URBAINES : DE L'OPEN DATA AU GOUVERNEMENT DES DONNÉES**

Antoine COURMONT, Sciences Po  
Centre d'études européennes (CEE)

# Politiques des données urbaines. De l'open data au gouvernement des données.

Si les administrations municipales ont toujours été productrices et utilisatrices de savoirs de gouvernement, d'indicateurs ou de statistiques, à partir du milieu des années 2000, de nouvelles entités ont fait leur apparition dans le gouvernement urbain : les données. Elles sont maintenant devenues un enjeu central pour de nombreuses métropoles, qui développent des stratégies, mettent en œuvre des instruments, transforment leur organisation et créent de nouveaux métiers pour gérer cette entité informationnelle. Comment les données sont-elles devenues un enjeu de gouvernement ? Qu'est-ce que l'émergence de ces nouvelles entités fait au gouvernement urbain ?

Cet article présente les résultats d'une thèse en science politique réalisée en partenariat avec la Métropole de Lyon dans le cadre d'une convention Cifre<sup>1</sup>. Durant trois ans, l'auteur a suivi quotidiennement la mise en œuvre de la politique d'open data du Grand Lyon. Ce matériau ethnographique a été complété par la réalisation de près de 70 entretiens, le dépouillement d'archives communautaires et une comparaison partielle avec des villes nord-américaines. La thèse suit la chaîne des données au travers de cinq étapes : produire, mettre en circulation, ouvrir, réutiliser et gouverner. A chacune de ces étapes, sont mises en évidence les politiques des données urbaines entendues comme les modes d'association de collectifs autour des données.

## Quand les données arrivent en ville : le libéralisme informationnel à l'échelle urbaine

L'émergence des données dans le gouvernement des villes est étroitement liée au développement des démarches d'*open data* et de *smart cities* au sein de nombreuses métropoles. La mise en circulation des données, le passage d'une logique de stock à une logique de flux, met en lumière l'application des principes du libéralisme informationnel à l'échelle urbaine.

L'*open data* est un mouvement international, initié dans les pays anglo-saxons, visant à mettre en circulation les données des administrations publiques afin de les rendre accessibles aux citoyens. La mise à disposition des données doit favoriser à la fois le développement économique et la transparence et la participation citoyenne. En France, la ville de Rennes est la première collectivité à s'engager dans une politique d'*open data* dès 2010. Aujourd'hui, plus d'une centaine de collectivités sont engagées dans des démarches d'ouverture de leurs données. Ce chiffre devrait s'étendre dans les prochains mois, puisque la loi NOTRE, votée en 2015, impose l'ouverture des données à l'ensemble des collectivités de plus de 3500 habitants.

---

<sup>1</sup> Antoine Courmont, *Politiques des données urbaines. Ce que l'open data fait au gouvernement urbain.*, Sciences Po, Paris, 2016, 423 p.

Parallèlement, sous l'impulsion d'entreprises des technologies de l'information telles que IBM<sup>2</sup>, la *smart city* a fait son apparition dans les métropoles. Si la ville intelligente reste une notion floue, on peut la définir de manière générique comme l'usage généralisé de données numériques pour rationaliser la planification et la gestion des villes<sup>3</sup>. Les discours sur la ville intelligente présentent la ville comme une totalité fonctionnelle et cohérente. La *smart city* est en effet pensée dans une vision systémique du fonctionnement urbain : la plasticité de la matière informationnelle doit permettre de gérer les interdépendances entre les multiples fonctions urbaines. La « couche informationnelle » relie – et dépasse – les différents secteurs d'action publique formant une « digital skin » enrobant les villes<sup>4</sup>. Pour cela, il est nécessaire de décloisonner les « silos » de données et de passer d'une logique de stock à une logique de flux, c'est-à-dire de mettre en circulation les données.

*Open data* et *Smart city* traduisent l'application de l'idéologie du « libéralisme informationnel » à l'échelle des villes. Théorisé par Benjamin Loveluck à partir d'une généalogie politique d'Internet, le libéralisme informationnel est une reformulation de la théorie politique libérale fondée sur les flux d'informations au sein des réseaux informatiques<sup>5</sup>. L'information doit être libre afin que ces flux d'informations résolvent de multiples problèmes dans une logique cybernétique. Le paradigme de la mobilité, mis en avant par John Urry<sup>6</sup>, s'applique maintenant aux flux d'informations : les données doivent être mobiles. La *smart city* caractérise ainsi la mise en circulation des données à l'échelle des villes.

## Les données entre attachements et détachements

Le libéralisme informationnel se base sur le postulat que la donnée existe, est autonome, détachée et circule aisément. Or, lorsque l'on confronte l'analyse des discours aux pratiques quotidiennes, le constat est loin d'être aussi évident. L'analyse des processus d'ouverture des données révèle que les données ne sont pas des entités déjà-là, autonomes qui peuvent aisément être mises en circulation. En effet, les données sont constituées d'attachements : elles sont attachées à des « métiers », à des systèmes d'information, à des applications, à des opérateurs, à des modèles économiques ou juridiques, etc. Par exemple, les données de circulation routière sont indissociables d'une longue chaîne qui associe des capteurs sous la chaussée, un réseau de fibre optique qui transmettent l'information, des algorithmes qui la traitent, des « traficiens » qui l'utilisent au sein du PC de régulation de la Métropole de Lyon pour piloter à distance les feux de circulation, d'une équipe qui la diffuse sur différents supports à destination des automobilistes, etc. Les données sont ainsi des dispositifs sociotechniques, constituées d'un réseau composé d'un ensemble d'actants reliés entre eux par des associations particulières. Toute donnée est indissociable du réseau sociotechnique qui la constitue.

---

<sup>2</sup> Ola Söderström, Till Paasche et Francisco Klauser, « Smart cities as corporate storytelling », *City*, 2014, vol. 18, n° 3, p. 307-320.

<sup>3</sup> Anthony Townsend, *Smart Cities - Big Data, Civic Hackers, and the Quest for a New Utopia*, New York, W. W. Norton & Company, 2013, 400 p.

<sup>4</sup> Chirag Rabari et Michael Storper, « The digital skin of cities: urban theory and research in the age of the sensed and metered city, ubiquitous computing and big data », *Cambridge Journal of Regions, Economy and Society*, 2015, vol. 8, n° 1, p. 27-42.

<sup>5</sup> Benjamin Loveluck, *Réseaux, libertés et contrôle. Une généalogie politique d'internet*, Paris, Armand Colin, 2015.

<sup>6</sup> John Urry, *Mobilities*, Cambridge, Polity Press, 2007, 336 p.

Dès lors, pour la mettre en circulation, il est nécessaire de recomposer ce réseau en déliant certaines associations – le détachement - et en constituant de nouvelles – le réattachement. Loin d’être joué à l’avance, le processus d’ouverture des données est le résultat d’une série d’épreuves, au résultat toujours incertain, au cours desquelles les caractéristiques des données, des producteurs, des utilisateurs, sont jugées et redéfinies par les acteurs.

Il faut en premier lieu parvenir à identifier la donnée. Les données ne sont en effet pas déjà-là, prêtes à être diffusées : il n’existe pas de catalogue répertoriant de manière exhaustive le patrimoine de données de l’administration. En outre, la notion de données ne va pas de soi pour les agents, nombre d’entre eux n’ont pas conscience de détenir et d’utiliser quotidiennement des données. Pour mettre en circulation une donnée, il faut ainsi d’abord savoir ce qu’est une donnée, caractériser une entité informationnelle comme une donnée et savoir qui la détient. Un travail d’enquête, au résultat incertain, est alors mené par les porteurs du projet d’*open data* pour lister les données existantes au sein de l’administration et identifier leur producteur.

Ensuite, contrairement aux discours des militants de l’*open data* qui incitent les producteurs à diffuser leurs données sans se préoccuper des usages qui en découleront, en pratique, le processus d’ouverture des données se caractérise par la construction d’utilisateurs imaginés. Les producteurs préfigurent des usages potentiels à partir desquels ils jugent de l’opportunité de mettre à disposition leurs données. Ces scénarios d’usage varient en fonction des producteurs et des données. Selon une direction, la mise à disposition des données peut mener à une modification des rapports de pouvoir, à une remise en cause de l’action de la collectivité, ou encore à des usages malveillants pour les citoyens. D’autres producteurs de données peuvent être davantage attentifs à des aspects économiques (crainte d’espionnage industriel), juridiques (responsabilité du producteur engagée en cas de mauvais usage), ou sécuritaire (risque pour la sécurité publique et l’intérêt général). Enfin, pour certaines données spécifiques aux processus métiers internes à la collectivité, les producteurs n’imaginent aucun usage potentiel par des acteurs externes et ne voient donc pas l’intérêt de les mettre à disposition. Afin d’intéresser les producteurs à l’*open data*, l’enjeu est alors de pondérer ces risques perçus par les avantages potentiels à l’ouverture de leurs données. La donnée est considérée comme publiable quand le producteur estime que les avantages à associer la donnée à de nouveaux utilisateurs sont plus grands que les risques inhérents.

Enfin, il est nécessaire d’extraire la donnée de son environnement initial et de la mettre à disposition sur une infrastructure de diffusion. Les données ne sont en effet jamais indépendantes d’une infrastructure technique au sein de laquelle elles sont produites et utilisées. Ces systèmes d’information n’ont pas toujours été conçus pour permettre l’extraction et la mise à disposition des données qu’ils contiennent. Une transformation de l’infrastructure est souvent réalisée pour permettre la mise à jour automatique des données sur le portail, un changement de format de données ou la diffusion de données volumineuses et en temps réel, tout en assurant la sécurité du système d’information de l’institution. Ces opérations constituent un véritable investissement technique, financier et humain qui peuvent les rendre particulièrement coûteuses<sup>7</sup>. Certaines données ne sont alors pas mises à disposition du fait de contraintes techniques ou organisationnelles. En outre, les données sont transformées avant leur publication pour rendre possible le détachement de leur

---

<sup>7</sup> Le coût total de la politique d’*open data* de la Métropole de Lyon n’a pas été calculé. A titre d’exemple, la mise à disposition des données en temps réel de la RATP est le résultat de six mois de travail pour un coût d’un million d’euros.

Source : [http://lemonde.fr/economie/article/2017/01/05/la-ratp-ouvre-enfin-ses-donnees-temps-reel\\_5057926\\_3234.html](http://lemonde.fr/economie/article/2017/01/05/la-ratp-ouvre-enfin-ses-donnees-temps-reel_5057926_3234.html) (Consulté le 19 mars 2017)

environnement de production et leur attachement à de nouveaux utilisateurs. La suppression des attributs « métiers » est une de ces opérations de « brutification<sup>8</sup> » qui visent à faciliter la réutilisation de la donnée. Cette sélection s'effectue en préfigurant un usage de la donnée ouverte, ci-après par un citoyen souhaitant connaître les jours des marchés.

*Enquêteur : Et sur la mise à disposition de la donnée, tu disais sélectionner certains attributs, ça on prend, ça on prend pas pour diffuser. Il y a certains attributs que tu invisibilises en fait ?*

*Pauline : Oui, oui. Par exemple, pour les marchés alimentaires, t'avais vraiment énormément de champs à remplir parce que tu disais par exemple, ce marché, imaginons, il a lieu le lundi matin uniquement, tu avais lundi, mardi, mercredi, tous les jours de la semaine avec oui, non, non, non. Après, est-ce qu'il est désinfecté le lundi matin ? Donc, lundi, oui, non, non, non, non. Tu vois à de multiples reprises. Est-ce qu'il est lavé ? Etc. Donc, c'était énorme, tu n'allais pas mettre tous ces champs, c'était ridicule. Même savoir s'il était désinfecté, lavé, c'était vraiment de la donnée métier, donc ça je l'ai pas mis. J'ai mis, ouverture, début du marché, heure de début, heure de fin de marché, et, c'est tout. **En gros, j'ai essayé de me mettre dans la place d'un usager.** Je me suis dit, voilà, il s'en fout de la surface, il s'en fout de la mise en circulation après, parce qu'il y a un temps entre la fin du marché et le début de la mise en circulation où tu peux nettoyer. Donc, voilà, j'ai essayé de la simplifier, de ne mettre que quelques données qui pouvaient être intéressantes pour un usager<sup>9</sup>.*

Les usages projetés influencent les caractéristiques de la donnée publiée. C'est également le cas lorsque ce sont des attributs considérés comme « sensibles » qui sont retirés. Par exemple, sur demande du service des arbres, l'attribut indiquant leur état phytosanitaire a été retiré de la donnée « arbre d'alignement ». Il considère que cette information « métier » nécessite « *d'être un professionnel de la botanique, de l'élagage pour analyser ces choses là* »<sup>10</sup>. Ces opérations de « nettoyage » de la donnée révèlent que la donnée « brute » est un mythe. Il n'existe pas de données autonomes de tout environnement social. La donnée est toujours travaillée pour répondre à un usage déterminé et/ou empêcher un usage préfiguré.

\*

\*\*

Loin d'être joué à l'avance, le processus d'ouverture des données est ainsi l'aboutissement d'une série d'épreuves, au résultat toujours incertain, au cours desquelles les caractéristiques des données, des producteurs, des utilisateurs, sont jugées et redéfinies par les acteurs. Ces différentes *épreuves de diffusibilité* conduisent ainsi à recomposer le réseau des données afin qu'elles soient considérées comme « diffusibles ». Ce processus est réversible : des données non catégorisées comme telles peuvent le devenir, et, inversement, des données ouvertes peuvent être redéfinies comme non-diffusibles.

Le cas de la donnée « Silos à verre » est particulièrement révélateur de ces épreuves de diffusibilité. Produite par la direction de la propreté de la communauté urbaine, cette donnée géographique a été successivement publiée, dépubliée et republiée sur le site *open data* en l'espace de quelques mois. Si elle avait été identifiée par l'administrateur de données SIG, la direction de la propreté a jugé a posteriori que la qualité de la donnée n'était pas suffisamment

---

<sup>8</sup> Jérôme Denis et Samuel Goëta, « La fabrique des données brutes. Le travail en coulisses de l'open data. » dans Clément Mabi, Jean-Christophe Plantin et Laurence Monnoyer-Smith (eds.), *Penser l'écosystème des données. Les enjeux scientifiques et politiques des données numériques.*, Paris, Editions de la Maison des Sciences de l'Homme, 2015, .

<sup>9</sup> Source : Entretien avec Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

<sup>10</sup> Source : Entretien Matthias, administrateur de données, Grand Lyon (17 janvier 2014)

fiable pour être mise à disposition du public. Une campagne de vérification a été mise en œuvre pour s'assurer de la fiabilité de l'information avant de publier à nouveau la donnée.

Cette modification du réseau sociotechnique transforme la donnée. A mesure que la donnée circule d'une épreuve à l'autre, elle diffère par le réseau qu'elle déploie. La donnée comme entité stable et immuable est une fiction. La donnée n'est ainsi pas préexistante à son ouverture. Au début du processus de diffusibilité, la donnée n'existe pas. Dès lors, il est impossible d'affirmer que la donnée a une essence, c'est-à-dire certaines propriétés desquelles, il serait possible de déterminer, dès le début de ce processus, sa diffusibilité. Loin d'être joué à l'avance, le processus d'ouverture des données est le résultat d'une série d'épreuves, au résultat toujours incertain, au cours desquelles les caractéristiques des données, des producteurs, des utilisateurs, sont jugées et redéfinies par les acteurs. Ces différentes *épreuves de diffusibilité* conduisent ainsi à recomposer le réseau des données afin qu'elles soient considérées comme « diffusibles ».

### **L'open data par défaut facilite-t-il la diffusibilité des données ?**

Plusieurs militants de l'open data plaident pour la mise en place d'un principe d'open data par défaut, qui vient d'être en partie acté dans la loi pour une République numérique. L'ouverture deviendrait alors la règle, la fermeture l'exception. En mars 2012, le maire de New-York, Michael Bloomberg, a signé une loi en ce sens : toutes les données de l'administration municipale doivent être diffusées sur le portail open data de la ville. Si cette loi a indéniablement favorisé l'ouverture des données, sa mise en œuvre est plus complexe que son postulat originel. En effet, les caractéristiques de diffusibilité des données ne se décrètent pas mais doivent être acquises au cours d'épreuves de diffusibilité. Les données new-yorkaises subissent les mêmes épreuves d'identification, de publicisation et d'extraction qui restreignent leur ouverture. L'impensé du principe d'open data par défaut est que les données ne peuvent être ouvertes par défaut, puisqu'elles acquièrent précisément ces caractéristiques de diffusibilité au cours du processus d'ouverture.

## **Les données à l'épreuve de leur réutilisation**

Poursuivons la chaîne des données ouvertes pour s'intéresser à leurs réutilisations. Loin des promesses initiales de développement économique et de renouvellement démocratique, à l'heure d'un premier bilan des démarches open data, les acteurs du secteur portaient un regard désappointé face à la faiblesse (relative) des réutilisations des données ouvertes. Une observatrice attentive de l'open data en France témoignait ainsi en janvier 2013 d'un « baby blues de l'open data » : « *les organisations traversent une phase de déprime et de doute, le data blues* » ; « *la multitude de difficultés techniques, juridiques, culturelles, et organisationnelles a laissé un goût amer aux réutilisateurs de données*<sup>11</sup>. » Ces réactions soulignent le fait qu'une offre de données ne rencontre pas automatiquement une demande de données : ce n'est pas parce qu'une donnée est mise à disposition qu'elle sera (ou pourra) être réutilisée. Quels sont les défis auxquels sont confrontés les réutilisateurs de données ? Comment font-ils pour les surmonter ? Ces questions soulèvent plus largement des enjeux de communication et de coordination entre des mondes sociaux hétérogènes. Comment les données produites et utilisées pour répondre aux besoins de l'administration peuvent-elles être utilisées au sein d'autres univers sociaux ?

Pour répondre à ces interrogations, plutôt que de se focaliser uniquement sur les réutilisateurs, il est nécessaire de faire entrer d'autres entités dans l'analyse, dont en premier lieu les

<sup>11</sup> Source : <https://libertic.wordpress.com/2013/09/24/vers-la-fin-du-baby-blues-de-lopen-data> (Consulté le 14 mai 2016)

données elles-mêmes. Loin d'être neutres, celles-ci ne se laissent pas aisément réutiliser. De multiples opérations de nettoyage, de croisement, de standardisation, d'articulation, etc. sont nécessaires pour permettre leur attachement à un nouvel environnement. Certaines entités sont associées, d'autres dissociées, transformant d'autant les données. Tout l'enjeu est alors de réussir à aligner ces médiations pour « faire tenir » le réseau de ces données ouvertes et assurer leur solidité.

Au travers de l'étude de plusieurs réutilisations de données mises à disposition par le Grand Lyon, différentes manières de composer avec les données ouvertes sont révélées. Ces politiques plurielles de réutilisation sont présentées dans trois catégories selon les opérations qu'elles font subir aux données : la consolidation, l'homogénéisation et l'articulation.

### Consolidation

Marquée de l'empreinte institutionnelle de son producteur, la donnée ouverte ne se laisse pas aisément utiliser au sein d'autres univers organisationnels ou professionnels. La dissociation des liens entre production et usage met en péril la solidité de la donnée. Puisqu'on ne peut s'assurer que les modalités de production de la donnée répondent aux besoins des réutilisateurs, les données sont menacées de déliquescence<sup>12</sup>. Plusieurs situations soulignent les difficultés de réutilisation des données : l'absence dans le jeu de données initial des informations nécessaires pour une utilisation alternative de la donnée, l'échelle de la donnée qui est limitée au périmètre de compétence de l'institution ou encore la structuration hétérogène des données qui restreint le développement de services. Ces facteurs de déliquescence des données ouvertes mettent en lumière leur héritage institutionnel. Produites pour répondre à des besoins spécifiques, au sein d'un environnement particulier, les données ouvertes sont difficilement réutilisables dans un cadre d'action alternatif.

Il existe toutefois de multiples exemples de réutilisations de données ouvertes. Pour parvenir à passer d'un monde social à un autre tout en préservant leur solidité, les données ouvertes doivent devenir des objets-frontières, dotés d'une flexibilité interprétative et d'une infrastructure partagée entre ces mondes sociaux.

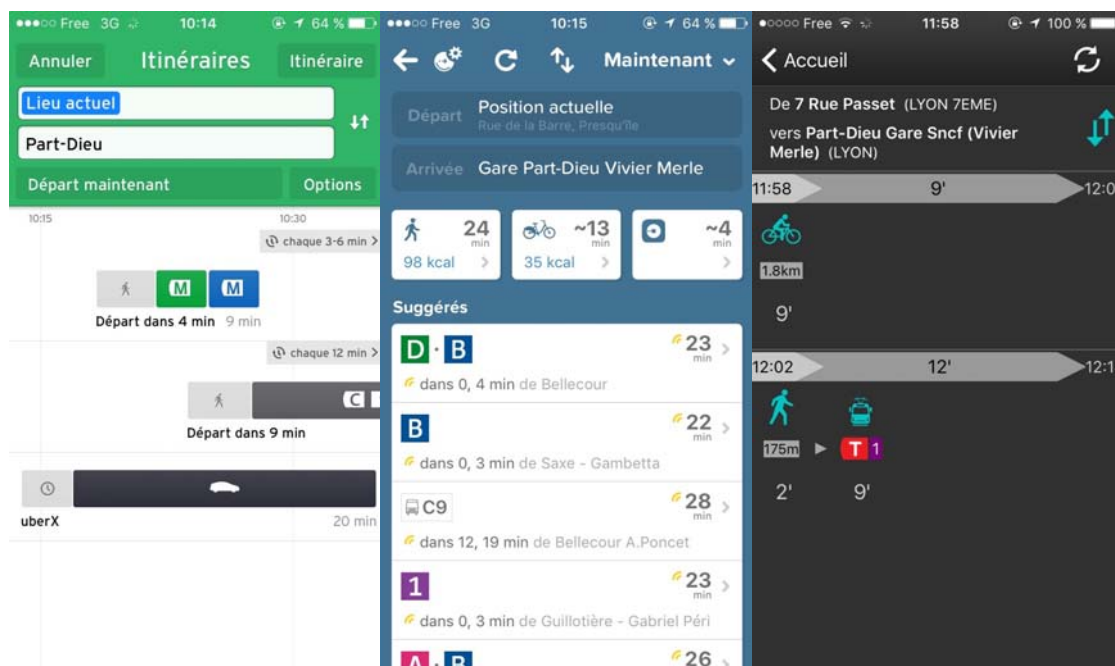
Préalablement à tout usage secondaire des données ouvertes, les réutilisateurs réalisent deux opérations de «sourcing» et de «nettoyage». La première consiste à identifier, comprendre et estimer la fiabilité de la donnée pour l'usage que l'on souhaite en faire. La seconde se compose de l'ensemble des opérations de préparation de la donnée afin de l'intégrer au sein d'un nouvel environnement informationnel. Ces deux opérations renforcent la consolidation des données ouvertes en testant leur flexibilité interprétative, c'est-à-dire en s'assurant qu'elles peuvent être utilisées dans un cadre d'action alternatif sans perdre leur signification initiale. Ces opérations ne sont toutefois pas suffisantes pour que la donnée devienne un objet-frontière utilisé au sein d'autres mondes sociaux. Le partage d'une infrastructure, composée notamment de standards et de conventions d'usage, entre les producteurs et les réutilisateurs, est nécessaire pour que la donnée devienne objet-frontière. La donnée peut facilement dépasser les frontières organisationnelles si elle répond aux conventions propres à un milieu professionnel. Les cas des données géographiques et des données de mobilité soulignent que leur usage est facilité par l'utilisation de formats standardisés. Ces derniers sont autant de conventions entre des communautés professionnelles qui permettent à la donnée d'être utilisée au-delà des frontières organisationnelles. Un géomaticien n'aura pas de difficultés à

---

<sup>12</sup> Emmanuel Didier définit la déliquescence comme la perte de prévisibilité des liens qui unissent les éléments qui composent un agrégat.

Emmanuel Didier, « La consistance du futur. La prévision statistique aux États-Unis et la Grande Dépression », *Raisons politiques*, 2012, vol. 4, no 48, p. 65 83.

insérer une couche de données géographiques dans son SIG ; dès lors qu'elles sont fournies dans un format standardisé, un développeur parviendra à utiliser les données des horaires des transports en commun pour mettre en place un service d'information aux usagers du réseau.



*La consolidation des données ouvertes de transport en commun permet le développement par des start-up d'applications de services de mobilité qui viennent compléter et prolonger l'offre fournie par les pouvoirs publics (ici, les applications Transit, Citymapper et Optymod).*

Devenue objet-frontière, la donnée ouverte permet à différents mondes sociaux de maintenir leurs différences et leur autonomie, tout en se coordonnant dans le temps et dans l'espace. Les acteurs peuvent alors s'accorder autour d'une représentation commune de l'espace urbain tout en préservant leurs points de vue respectifs. Cela facilite par exemple la coordination des différents acteurs d'un projet urbain (collectivités, agence d'urbanisme, cabinets d'architecture ou d'ingénierie, etc.) qui accèdent aux mêmes données sur le portail *open data* du Grand Lyon et partagent ainsi un langage commun. La mise en place de cet espace d'équivalence renforce la capacité du Grand Lyon à imposer sa représentation de l'espace urbain à l'ensemble des acteurs intervenant sur son territoire.

## Homogénéisation

L'homogénéisation est la production d'un nouvel agrégat à partir de données hétérogènes. Contrairement à la consolidation, il ne s'agit pas ici d'utiliser la donnée ouverte pour un nouvel usage tout en préservant sa représentation de l'espace urbain, mais de la considérer comme une source d'information parmi d'autres pour construire une nouvelle donnée uniformisant des représentations hétérogènes. L'homogénéisation est ainsi la construction d'un agrégat basé sur des conventions d'équivalence alternatives à celles établies par les producteurs des données ouvertes. Ces opérations d'homogénéisation font émerger une nouvelle représentation de l'espace urbain, au sein de laquelle l'héritage institutionnel des données n'est plus visible. La perte de cette maîtrise de représentation de l'espace par les acteurs publics reconfigure le pouvoir urbain au profit de nouveaux acteurs positionnés en situation d'intermédiaire entre producteurs et utilisateurs de données.

Ce besoin d'homogénéisation émerge lorsqu'il y a une volonté d'utiliser les données à une échelle plus large que celles pour lesquelles elles ont été produites. Cela peut-être le souhait d'utiliser des données communales à l'échelle intercommunale, des données métropolitaines à



l'échelle nationale ou internationale ou plus largement sur un périmètre autre que les territoires institutionnels.

Par exemple, l'entreprise de cartographie numérique Here Maps produit des bases de données géographiques destinées à la navigation routière. Leur production repose sur de nombreuses sources d'informations. L'usage de l'open data est réservé exclusivement aux mises à jour de la base de données de Here. Ses cartographies, qui couvrent près de 190 pays, sont vendues à de nombreux constructeurs automobiles, mais aussi à des entreprises numériques (Amazon, Bing, Yahoo, etc.). Pour être associées aux bases de données géographiques de l'entreprise Here, les données en open data doivent respecter des normes précises édictées par la société. Cela doit assurer un degré de qualité similaire sur l'ensemble des territoires couverts par la société. A partir de sources de données hétérogènes, ce travail de qualification et d'intégration fournit alors une représentation homogène de l'ensemble des espaces couverts par l'entreprise.

Ces opérations d'intégration des données ouvertes au sein de ces bases de données nécessitent un travail important réalisé par des personnes dédiées au sein de chacune de ces organisations. Le travail de ces « petites mains<sup>13</sup> » est indispensable pour fluidifier les nombreuses frictions<sup>14</sup> qui apparaissent lors de la mise en relation d'infrastructures informationnelles.

L'intégration de sources de données hétérogènes au sein d'une base de données unifiée impose l'établissement de conventions d'équivalence<sup>15</sup>, instaurant une représentation uniforme de l'élément représenté.

À la différence des systèmes statistiques nationaux, l'établissement de ces conventions d'équivalence n'est plus l'apanage de l'État ou des acteurs publics. Il est pris en charge par des acteurs privés qui imposent leur représentation de l'espace urbain au risque, pour les acteurs publics, d'une perte de maîtrise de leurs politiques publiques.

Par exemple, la base de données de l'entreprise Here repose sur une typologie des voies alternatives à la hiérarchisation du réseau routier établie par le Grand Lyon. Cette différence de représentation n'est pas neutre : elle est une force de prescription importante sur le comportement des automobilistes par l'intermédiaire de l'itinéraire mis en avant par leur GPS. Particulièrement visibles lors de situations de congestion, ces différences de classification entraînent un report du trafic automobile par les GPS sur des voies que le Grand Lyon considère comme non adaptées. La hiérarchisation des voies établie par l'acteur public ne fait plus convention, ce qui réduit sa capacité de régulation du réseau<sup>16</sup>. Face à ces

---

<sup>13</sup> Jérôme Denis et David Pontille, « Travailleurs de l'écrit, matières de l'information », *Revue d'anthropologie des connaissances*, 2012, vol. 6, n° 1, p. 1-20.

<sup>14</sup> Paul Edwards, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.*, Cambridge Mass., MIT Press, 2010.

<sup>15</sup> Alain Desrosières, « Entre réalisme métrologique et conventions d'équivalence : les ambiguïtés de la sociologie quantitative. », *Genèses*, 2001, vol. 2, n° 43, p. 112-127.

<sup>16</sup> Face à cette situation, le Grand Lyon se trouve réduit à agir sur l'infrastructure, en réclamant à l'Etat le déclassement de la voie, afin de pouvoir réduire drastiquement la vitesse de circulation et installer des feux de circulation, ce qui contribuerait à modifier substantiellement les calculs d'itinéraires.

Source : <http://www.rue89lyon.fr/2016/04/01/autoroute-a6-a7-quand-gerard-collomb-voulait-convaincre-les-gps-dignorer-fourviere> (Consulté le 16 mai 2016)

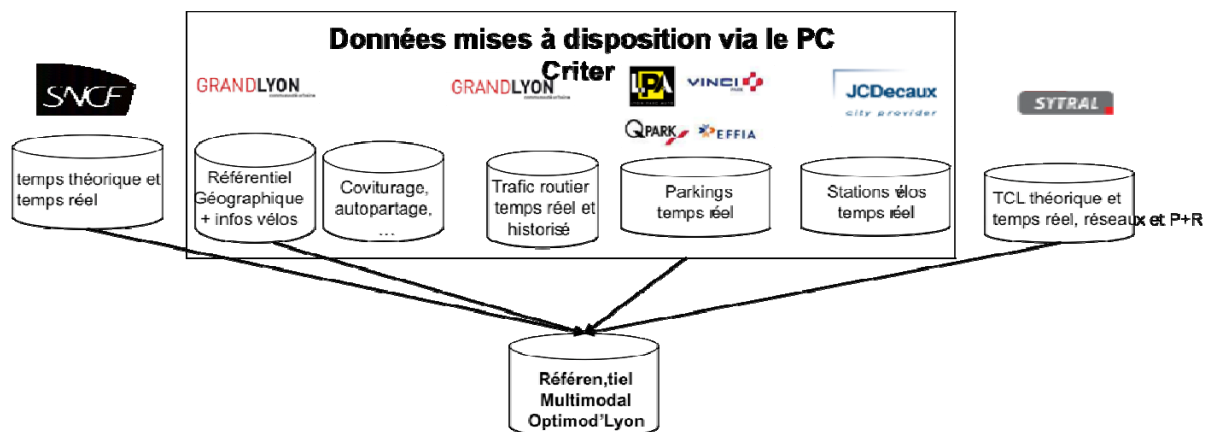
nouvelles conventions d'équivalence, le Grand Lyon perd la maîtrise de la politique de mobilité de l'agglomération.

### Articulation

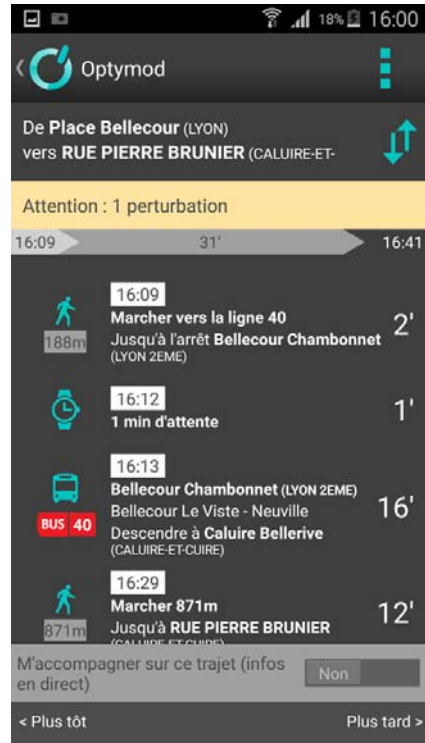
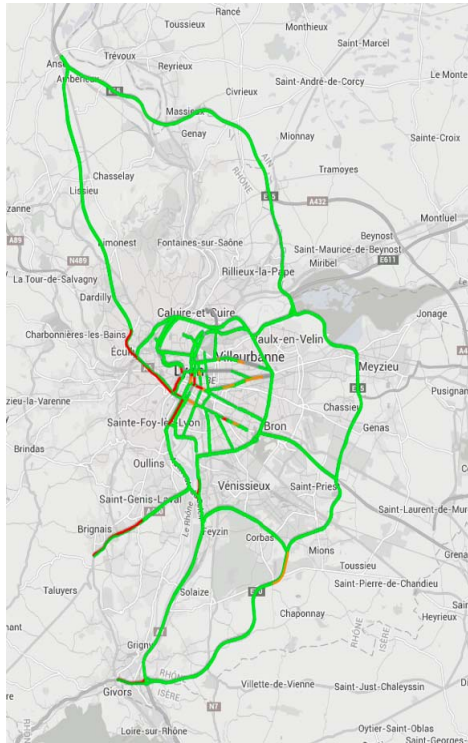
L'articulation est la troisième modalité de réutilisation des données ouvertes. Elle se caractérise par la mise en relation de données hétérogènes par l'intermédiaire d'un attribut commun. L'articulation produit une nouvelle représentation de l'espace urbain par l'association de données qui conservent toute leur hétérogénéité.

Par exemple, le projet Optimod, un système de transport intelligent (ITS) mis en place par la communauté urbaine de Lyon, qui rassemble, articule et analyse les données de l'ensemble des modes de transport de l'agglomération pour fournir des services d'information multimodaux aux usagers. Le projet vise ainsi à articuler les données de l'ensemble des modes de transport de l'agglomération lyonnaise pour produire une représentation métropolitaine de la mobilité.

Tout l'enjeu est alors d'assigner aux données un usage secondaire pour lequel elles n'ont pas été conçues. Les différences de structuration rendent impossible l'homogénéisation des données dans un modèle commun. L'enjeu est de préserver l'héritage de la donnée en interconnectant les jeux de données sans modifier les façons dont ils sont produits. Un travail d'articulation des jeux de données autour d'un dénominateur commun est alors nécessaire. Un référentiel géographique décrivant l'ensemble du réseau de voirie a ainsi été constitué pour permettre la mise en relation des différentes bases de données de transport initialement incompatibles.



A l'issue de ces opérations d'articulation, une nouvelle représentation de l'espace urbain émerge. En associant les données de l'ensemble des opérateurs de transport de l'agglomération lyonnaise, la « Centrale de mobilité » offre une nouvelle représentation de l'espace de la mobilité urbaine. L'objectif de ce projet est d'offrir de nouveaux services d'information aux voyageurs afin de favoriser le report modal de la voiture vers les transports en commun ou les modes doux. Cette nouvelle représentation de l'espace urbain transforme l'objet de la politique publique. En plaçant l'utilisateur au centre d'un réseau de données, on glisse de la régulation des flux des réseaux de transport vers la régulation des individus.



Ce changement de représentation fait évoluer la cible des politiques publiques. Alors qu'auparavant les données étaient des instruments de régulation des réseaux de transport, elles doivent dorénavant permettre à chaque usager, envisagé comme un acteur rationnel, de comparer les modes de transport à sa disposition. La régulation des transports urbains se déplace ainsi des réseaux aux individus. Par le *big data*, ce n'est plus le « nous » qui est gouverné, mais chaque individu qui devient gouvernable.

\*  
\* \*

Ces trois politiques de données ouvertes ne constituent pas seulement différentes modalités de réutilisation des données. En passant d'un monde social à un autre, la donnée ouverte se dote d'autres qualités. Cette évolution des données modifie les représentations de l'espace urbain qu'elles véhiculent, et, influe, in fine, sur les politiques publiques.

- La donnée consolidée est un objet-frontière permettant la coordination d'acteurs pluriels autour d'une représentation commune de la ville. La consolidation offre au producteur de nouvelles capacités de régulation en rassemblant ces acteurs autour de sa donnée.
- La donnée homogénéisée propose une représentation alternative de l'espace urbain par la mise en place d'un nouvel espace d'équivalence qui échappe au producteur initial. Ce dernier perd alors la maîtrise de la représentation de la ville au risque de conflits entre ces points de vue hétérogènes.
- La donnée articulée rassemble une multiplicité de points de vue sur un même objet. Par ce biais, elle rend visible la singularité de chacune de ces entités et offre la possibilité de gouverner spécifiquement chacune d'entre elles.

## Gouverner les données pour gouverner la ville à l'ère du numérique

La mise en circulation des données fait émerger un enjeu de gouvernement des données. Alors que de plus en plus d'actions et de comportements sont régulés par des flux d'informations, si l'acteur public ne maîtrise pas ces flux, il perd la maîtrise de ses politiques publiques. L'exemple des GPS et des itinéraires de substitution met en lumière tout l'enjeu que constitue la régulation des flux de données pour conserver la maîtrise des politiques publiques. Au travers du gouvernement des données, c'est la capacité de l'acteur public à gouverner la ville à l'ère du numérique qui est en jeu.

Gouverner la donnée, c'est réguler la circulation de l'information. Cela suppose de maîtriser les flux de données et de les orienter pour, au choix, faciliter ou restreindre leurs usages. Trois postures de régulation des flux de données peuvent être distinguées :

- Le premier choix est celui du laisser-faire : il faut laisser circuler les données sans entrave. Le rôle de l'acteur public est de permettre cette mise en circulation sans préjuger des usages qui seront faits, puisque le « marché » devrait s'auto-organiser. Benjamin Loveluck distingue trois modèles d'organisation des flux d'informations :
  - o La captation : un acteur collecte, s'approprie et centralise les données pour proposer de nouveaux services (ex. Google)
  - o La dissémination : l'architecture technique distribuée empêche toute centralisation au profit d'un seul acteur. Inspirée de l'architecture originelle d'Internet, on la retrouve dans les réseaux de pair à pair ou dans la technologie *blockchain*.
  - o L'auto-institution : basée sur le modèle des communs, l'architecture technique est encadrée par des procédures et des règles qui permettent aux acteurs de se saisir de la circulation de l'information et de la formation du collectif (ex. Wikipédia, OpenStreetMap).
- Le deuxième choix est celui du protectionnisme : on ferme les frontières de l'accès aux données pour en conserver la maîtrise exclusive. C'est le choix effectué pour les données jugées sensibles : les données restent attachées exclusivement à leur environnement initial.
- Le troisième choix est celui du tiers de confiance : une posture de tiers neutre entre producteurs et utilisateurs, qui crée une relation de confiance là où règnent incertitudes et méfiance. C'est le positionnement mis en avant par la Métropole de Lyon au travers du projet Optimod de centrale de mobilité : placée entre différents producteurs publics et privés et des utilisateurs, elle garantit aux premiers qu'il ne sera pas fait un mauvais usage de leurs données, et aux seconds qu'ils auront accès aux données de manière pérenne. Garantir la confiance est loin d'être facile, cela passe par plusieurs dispositifs (par exemple les licences, les labels, l'infrastructure technique, etc.).

Ces trois modèles de régulation des flux de données sont des idéaux-types qui s'appliquent différemment selon les secteurs d'action publique, les acteurs en présence, les données concernées. Gouverner la donnée, c'est donc composer entre ces différentes approches. Il n'y a pas un modèle unique de régulation des flux de données.

Le gouvernement des données métropolitaines en est à ses prémices. Toutefois, une tension latente entre autonomisation et sectorialisation se manifeste dès à présent : la donnée doit-elle

être l'objet d'une politique spécifique ou n'est-elle qu'un enjeu à prendre en compte de manière différenciée dans chaque politique sectorielle ?

Au sein de la Métropole de Lyon, le repositionnement stratégique des services informatiques et la création en son sein d'une nouvelle fonction d'administrateur général des données<sup>17</sup> qui a pour rôle de piloter la politique publique de la donnée métropolitaine, semble faire pencher la balance vers la première option. Toutefois, parallèlement, les directions du Grand Lyon insèrent de leur propre chef ce volet « donnée » dans leurs politiques publiques, sans que la direction des services informatiques, censée piloter la gouvernance de la donnée n'intervienne directement. La direction de l'eau par exemple a intégré cette question lors du renouvellement du contrat de délégation de service public. Indépendamment des services informatiques du Grand Lyon, elle a mis en avant ce nouvel aspect, exigé l'accès aux données du délégataire, défini des formats et des protocoles d'échange de données, etc.

De nombreuses questions restent ouvertes et constituent autant de pistes de recherche à explorer. Elles mériteraient des travaux comparant des villes, des institutions, des secteurs d'action publique et des natures de données pour mettre en évidence les capacités de régulation des acteurs publics. L'histoire du gouvernement de la donnée métropolitaine reste largement à écrire.

---

<sup>17</sup> Recruté en avril 2015 au sein du Grand Lyon, ce poste s'inspire directement de celui du *Chief Data Officer*, nommé dans certaines entreprises et administrations municipales nord-américaines.