



Dix ans d'ouverture des données publiques
Un bilan critique



Sommaire

Pourquoi un bilan critique (maintenant) ?	3
Partie 1 - Factualiser: qu'est-ce que l'ouverture des données publiques aujourd'hui en France ?	4
Acteurs, volumes, dispositifs : l'offre de données publiques ouvertes	4
Une mobilisation massive des "grands" acteurs publics qui contraste avec une adoption faible des "petits" acteurs	4
La métrique du territoire en lieu et place du nombre d'acteurs territoriaux ?	5
Le citoyen, un des gros producteurs d'Open Data ?	5
40 000 jeux de données ouvertes	6
Des données très variées avec peu d'angles morts	6
Qualité des données : le chemin est tracé mais la route est longue	7
Des dispositifs privilégiant spécialistes et horizontalité des données : un portail sinon rien ?	8
Le service public de la donnée, un dispositif original sur les jeux de référence	8
Démocratisation et nouveaux usages	9
La demande : un sujet encore mal appréhendé	9
Les succès d'usage : des profils proches en qualité, granularité et couverture	10
La longue traîne des données (presque) jamais utilisées	10
Les radars pédagogiques : des usages mais peu d'usagers	11
Partie 2 - Analyser: les impacts de l'ouverture des données publiques	12
5 grandes familles d'impacts	12
Les impacts économiques	12
Le marché de la donnée comme matière première: l'open data comme accélérateur de tendances déjà à l'oeuvre	13
Les gains économiques liés à la réutilisation des données	14
Une chaîne de valeur qui évolue avec de nouveaux acteurs	15
Des acteurs privés qui ouvrent leurs données	16
Les impacts sur la vie démocratique	17
Un outil de contrôle de la vie publique ... et de ses acteurs	17
Un outil d'empowerment des individus et des collectifs ?	18
L'open data n'est pas une machine à scandale ... ?	19
De quoi l'accroissement des demandes CADA est-il le signe ?	19
Les impacts sur les organisations (administrations, collectivités, etc.) et les relations entre acteurs	21
Les impacts sur les métiers, les fonctions et les méthodes de travail	22
La diffusion d'une culture de la donnée: pour un nouveau pouvoir d'agir	23
9 défis pour la prochaine décennie de l'ouverture des données publiques	25



Donner à voir, donner envie de se saisir de l'open data	25
Développer une offre de données vraiment adaptée à la demande	25
Réinterroger les frontières de l'open data	26

CRÉDITS

Auteurs : Charles Népote et Simon Chignard

Fondation Internet Nouvelle Génération

www.fing.org www.internetactu.net

8 passage Brulon 75012 Paris



Ce document est mis à disposition selon les termes de la licence Creative Commons Attribution 4.0 France : <https://creativecommons.org/licenses/by/4.0/fr/>

Vous êtes libre de partager, reproduire, distribuer et communiquer ce document, l'adapter et l'utiliser, à condition de l'attribuer de la manière suivante : "Open Data Impact : dix ans d'ouverture des données publiques" Fing 2020.

Ce document ne doit pas être attribué d'une manière qui suggérerait que la Fing vous approuve, vous ou votre utilisation de l'oeuvre.



Pourquoi un bilan critique (maintenant) ?

Au tout début des années 2010, quelques acteurs publics pionniers se sont engagés volontairement et résolument dans des démarches d'ouverture des données publiques. L'open data territorial est alors principalement l'affaire de quelques métropoles, territoires déjà bien mobilisés sur l'innovation numérique et les nouveaux usages. L'ouverture, en 2011, de la première version du portail national data.gouv.fr marque l'appropriation du sujet par l'Etat et les administrations centrales.

Comme dans toute aventure nouvelle et pionnière, ceux qui se sont lancés à cette époque ont dû défricher des champs nouveaux: quelles données doit-on ouvrir - et lesquelles faut-il plutôt garder pour soi ? comment animer une démarche d'open data ? quel est le cadre juridique et réglementaire applicable ? à quoi doit ressembler un portail open data ? Les réponses ont, pour la plupart, été co-construites dans l'échange et le partage d'expériences, que ce soit à travers des actions d'animation ou des événements (portés notamment par l'association Opendatafrance, La Fing, Etalab ou à l'époque Libertic) ou par des échanges directs et souvent interpersonnels entre les premiers animateurs de ces démarches.

Une décennie plus tard, le paysage français de l'open data a bien changé. La reconnaissance, en 2016, par la loi pour une République numérique du principe d'ouverture des données publiques par principe a marqué une étape importante. Dorénavant, l'open data n'est plus seulement un choix de quelques uns, c'est devenu une obligation pour l'ensemble des acteurs investis d'une mission de service public.

Le profil des acteurs qui se lancent a lui aussi changé: sans doute moins militant, il vient à l'ouverture des données publiques avec des attentes de résultats concrets. Les entreprises s'emparent du sujet, en tant qu'utilisatrices des données bien sûr, mais aussi parfois en ouvrant leurs propres données. Nous rentrons maintenant dans une phase d'acculturation plus large.

Au moment de tracer la feuille de route de la deuxième décennie de l'ouverture des données publiques, le programme Open Data Impact a proposé de faire un travail de bilan critique. Qu'est-ce que l'open data aujourd'hui ? Quels ont été les impacts des actions menées des huit dernières années ? Qu'est-ce qui a été réussi avec l'open data ? Les promesses ont-elles remplies ? Ce bilan a une particularité: il a été préparé par les acteurs du mouvement d'ouverture des données, administrations, entreprises, acteurs de la société civile. Il se compose de trois parties: la première vise à factueliser le diagnostic (qu'est-ce que l'ouverture des données publiques aujourd'hui en France ?), la seconde à analyser en détail les impacts (au sens large: impacts économiques, sociaux, organisationnels, sur les métiers, les fonctions, les cultures de travail). Enfin, en conclusion, nous proposons une série de défis auxquels sont confrontés aujourd'hui celles et ceux qui portent ce mouvement d'ouverture des données publiques.



Partie 1 - Actualiser: qu'est-ce que l'ouverture des données publiques aujourd'hui en France ?

Avant d'analyser les impacts (économiques, démocratiques, organisationnels, ...) de l'ouverture des données publiques, il nous semble important de actualiser ce que représente aujourd'hui l'open data en France. Alors, l'open data, "combien de divisions ?"

Acteurs, volumes, dispositifs : l'offre de données publiques ouvertes

Une mobilisation massive des "grands" acteurs publics qui contraste avec une adoption faible des "petits" acteurs

À partir de combien de jeux de données peut-on considérer qu'un territoire est engagé dans l'open data ? Au début des années 2010 la Grande-Bretagne, à travers le site "Openly Local", affichait une proportion record d'administrations estampillées "open data"... qui n'avaient en réalité, pour la plupart, publié qu'un seul jeu pour être sur la carte.

La France n'a pas connu cela en montrant très tôt des acteurs impliqués et appliqués à publier des données nombreuses et variées. En juin 2012, soit 2 ans à peine après la collectivité pionnière de Rennes, on comptait 5 des 15 plus grandes villes de France, 4 conseils départementaux et une région qui publiaient plus de 800 jeux de données (hors Etalab). En octobre 2018, au moment où l'ouverture des données devient la règle, deux ans après l'adoption de la loi pour une République numérique, le nombre de collectivités couvertes est le suivant :

- la quasi totalité des 20 plus grandes villes et métropoles
- 11 des 13 Régions (seules Normandie et Occitanie manquaient à l'appel)
- presque la moitié des départements
- plus de 50 EPCI (communautés de communes et d'agglomération)

Fin 2019, nous pouvons voir que la quasi-totalité des "grosses" collectivités ont ouvert leur données, ou sont en passe de le faire à un horizon de quelques mois.

Du côté de l'État, la totalité des ministères et l'essentiel de ses agences et services déconcentrés proposent de nombreux jeux de données ouverts, sous l'impulsion de la mission Etalab.

Chez les "**petits**" acteurs — territoires ruraux et territoires urbains de taille moyenne — la situation est tout autre. Le **nombre d'acteurs engagés est anecdotique** : quelques dizaines de territoires sur un potentiel de plusieurs milliers. La situation ne progresse que très lentement et on peut s'interroger la capacité réelle de ces territoires à se saisir de ce sujet. Les raisons en sont nombreuses : moyens humains, formation, structuration du service informatique, etc.

Ce contraste entre gros et petits a fait écrire et dire que l'Open Data concerne en fait peu d'acteurs et que le gros des troupes n'est pas encore sorti du bois. De fait la Loi pour une



République Numérique, considère (en simplifiant) que toute collectivité de plus de 3500 habitants est tenue d'ouvrir ses données de manière spontanée. Il n'est cependant pas interdit de s'interroger non pas sur l'esprit de cette loi mais bien sur les conditions de sa mise en oeuvre ? Le législateur a voulu promouvoir une politique très volontariste d'ouverture des données, développer un esprit où la transparence devient la règle et non l'exception. Pour autant, les difficultés rencontrées par les plus petits territoires interrogent. Doit-on souhaiter une démarche systématique et volontaire, de la part de petits acteurs dont l'éventuelle DSI est déjà surchargée ? Une alternative aurait été, pour ses plus petits acteurs, de renforcer le droit d'accès, permettant à tout citoyen de réclamer et d'obtenir plus facilement des données, plutôt que d'imposer un travail amont de publication de données ?

Sommes-nous condamnés à se passer de milliers d'acteurs de l'Open Data ? Mais après tout, est-ce les acteurs qui sont importants ou bien la couverture de l'infrastructure de données ? Faut-il donc regarder aussi précisément le nombre d'acteurs territoriaux ou les territoires eux-mêmes ? Dit autrement, les territoires français sont-ils bien pourvus en données ouvertes ?

La métrique du territoire en lieu et place du nombre d'acteurs territoriaux ?

En s'interrogeant sur les territoires, la situation change en effet radicalement. Si nous revenons aux petits acteurs, leurs territoires sont en réalité irrigués par d'importantes masses de données ouvertes. S'ils n'ont pas eux-même ouverts de données, les échelons administratifs supérieurs en publient beaucoup pour eux. Dans les départements, dans les Régions, dans les Métropoles et dans beaucoup d'importantes intercommunalités, au niveau de l'État enfin, de très nombreuses données sont publiées à une maille communale voire infra-communale. Certaines de ces données sont d'ailleurs collectées par les petits acteurs : les données électorales, les données de sécurité, du cadastre, etc. sont par exemple collectées de longue date par les communes. Les communes, *petites et grandes*, représentent ainsi l'un des principaux "pourvoyeur" d'Open Data, à défaut d'en être les éditeurs.

Le citoyen, un des gros producteurs d'Open Data ?

La plupart des données Open Data et leurs producteurs existent de longue date (INSEE, données des ministères, etc.). Mais depuis 15 ans, l'internet en général et le web en particulier, a permis la naissance d'immenses réseaux de co-production de données par les citoyens. On parle de "crowdsourcing", littéralement co-production par les foules, pour désigner ces pratiques de production de contenus ou de données en commun. Avant l'internet, des communautés d'amateurs réunissaient leurs données sur quelques sujets spécifiques comme l'astronomie, la botanique, etc. L'internet a sans cesse repoussé plus loin les limites de ces pratiques. Wikipedia, qu'on ne présente plus, en est l'ancêtre commun. OpenStreetMap (2004), base de données géographiques libre, en est l'enfant le plus fécond : sa couverture est exceptionnelle en zone dense (meilleure que ses équivalents commerciaux comme Google Maps) ; la progression du projet montre en outre qu'une couverture planétaire des infrastructures routières est réaliste.

D'autres projets suivent ces exemples. Open Food Facts, né en France, est un véritable wikipedia de l'alimentation. Il se targue de plus de 20 000 contributeurs ayant sourcé plus d'1 million 60 000 produits. C'est le point de départ de 99% des apps sur l'alimentation aujourd'hui, dont Yuka (10 millions d'installation) ou ScanEat.

Wikidata, quant à lui, est un projet annexe de Wikipedia, qui vise à mettre en données toute chose



existante.

Ces projets produisent des données de haute qualité et bâtissent des écosystèmes riches. Ils sont en outre pleinement interconnectés : les 4 projets majeurs (Wikipedia, OpenStreetMap, Wikidata et Open Food Facts) échangent des données et/ou référencent les données des uns des autres, augmentant ainsi la facilité et la valeur d'usage de ces données.

Auparavant vu avec méfiance ou amusement, **ces données participent aujourd'hui pleinement à l'écosystème professionnel des données**. Même les acteurs publics y trouvent des usages et contribuent souvent en retour.

Si l'on intègre les nombreux jeux de données fournis par l'État à la maille communale ou infra-communale, ainsi que les données communautaires (OpenStreetMap, etc.), nous pouvons affirmer qu'**aucune partie du territoire métropolitain français est dépourvue de données ouvertes**.

Certes, la densité des données connaît des écarts selon que vous soyez au coeur d'une Métropole ou d'un territoire rural excentré. Cependant, comme nous le verrons plus tard, l'Open Data se caractérise en France par un socle de données homogène sur tout le territoire.

L'adoption des grand pourvoyeurs de données à l'Open Data en France est donc résolument massive et permet une couverture très complète. Qu'en est-il de l'offre en terme de données ?

40 000 jeux de données ouvertes

Le comptage précis des données (par type, par thématique) se révèle extrêmement difficile, bien que la question de la mesure et la rationalisation de l'offre se soient très tôt posées aux acteurs — nous renvoyons notamment aux efforts non aboutis mais continus de catalogage des données (*répertoire des informations publiques* (promu par les textes dès 2005), recommandations de catalogage de l'APIE (2011), catalogue INSPIRE (2013), standard international DCAT (2013) et "Catalogue simplifié" (2018) proposé par Open Data France).

La France compterait à ce jour plus de 40 000 jeux de données publiés (plus de 35 000 recensés sur data.gouv.fr). En nombre, plus des trois quarts proviennent des collectivités, qui publient essentiellement des jeux de petites et moyennes tailles. L'État publie comparativement moins de jeux, en nombre, mais, par nature, de grandes quantités en volume, notamment à travers de grands jeux de référence dont la couverture est nationale.

Des données très variées avec peu d'angles morts

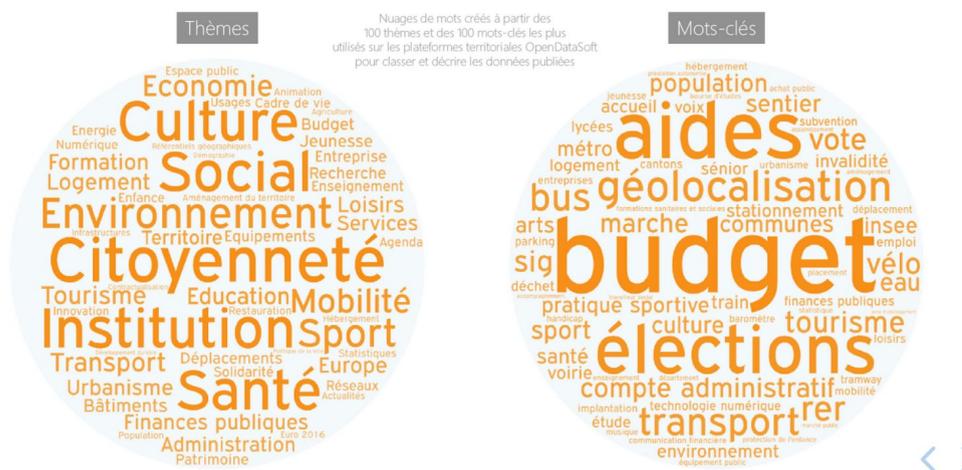
Une vision thématique des données montre un ensemble très varié. Certains domaines sont particulièrement bien couverts : géographie, économie, transports, etc. De fait, la France est une machine administrative bien huilée dont la production de données est une pratique pluri-centenaire, voire plus ancienne encore pour quelques sujets comme la fiscalité ou la cadastration. On ne trouve ainsi que peu d'angles morts ou de domaines pauvres en données. Tout au plus pouvons-nous évoquer quelques thèmes un peu en retrait : la thématique sociale ; la justice ; dans une moindre mesure, l'éducation ; la culture enfin où



l'on observe de forte disparités entre des acteurs en pointe (BNF...) et des acteurs plus réfractaires.

Certains domaines, à l'inverse, sont fort avancés, en termes de jeux, de normes, d'usages et d'écosystème, comme les transports, l'environnement, les données électorales et l'information géographique au sens large.

L'indexation des données territoriales ouvertes



Qualité des données : le chemin est tracé mais la route est longue

Habituellement la qualité ne dépend que de l'usage. C'est aussi le cas dans le domaine de la donnée : les producteurs, souvent les métiers, se sont contentés d'une qualité de données utile à leurs usages. L'ouverture des données, cependant, a créé des usages nouveaux, imprévus, pour lesquels la qualité initiale est jugée insuffisante. Peut-on reprocher aux acteurs publics de ne pas avoir dépensé sans compter pour des besoins de qualité qu'ils n'avaient pas ?

“Ouvrez, dieu reconnaîtra les siens” pourrait résumer les premières années de l'Open Data : il s'agissait d'expérimenter les effets de l'ouverture des données et donc d'ouvrir rapidement tout ce qui pouvait l'être. La nouveauté et les idéaux très positifs véhiculés par les données ouvertes (ouverture, innovation, nouveaux usages, etc.), ont également participé à une sorte de course à l'échalote, dans laquelle la qualité devenait un sujet annexe. Les acteurs s'accordent aujourd'hui pour identifier la qualité comme un goulet d'étranglement. Elle est devenue une préoccupation qui monte : standardisation des données par Open Data France (Socle Commun des Données Locales) et Etalab, outils de validation, ETL, outils de contrôle en tous genre, etc. Le sujet commence enfin à prendre de l'ampleur. Mais le mouvement restera lent et long parce que la qualité nécessite l'agrégation de compétences très hétérogènes et que, loin d'un sujet technique, c'est en réalité un sujet qui contient beaucoup d'humain.



Des dispositifs privilégiant spécialistes et horizontalité des données : un portail sinon rien ?

“Faire de l’open data”, comme on l’entend souvent, est trop souvent devenu synonyme de faire un portail.

Le portail Open Data est le dispositif principal de médiation de l’offre et de la demande. Il concentre toutes les fonctions et contenus utiles à la diffusion des données : catalogue, page dédiée pour chaque jeu, moteur de recherche ; outils sociaux (commentaires, publication des réutilisations), prévisualisation, outils géographiques et autres, pour les plus riches. De fait, le portail rationalise la publication, en standardise certains éléments (fiches, etc.) et facilite la communication. Le portail est aussi un bon vecteur de sérendipité voire favorise les rencontres inattendues (se perdre dans la donnée). Les acteurs ont déployé de gros effort pour en faire un dispositif qui abaisse le seuil d’usage des données. L’objectif est pour partie réussi : l’éditorialisation des données est l’un des piliers de la réussite de l’Open Data. OpenDataSoft, la plateforme qui motorise un gros tiers des sites Open Data publics, connaît d’ailleurs une importante réussite en entreprise : non que ces dernières publient en Open Data mais parce qu’elles ont compris l’intérêt d’éditorialiser en interne leur propres données.

Mais le portail trouve aussi ces limites. Qui connaît vraiment les portails Open Data ? L’audience de ces sites est loin derrière les sites institutionnels ou verticaux des différents producteurs de données. Le portail n’est souvent connu que des spécialistes. Malgré les efforts déployés, les portails demeurent des objets très techniques. Peu d’acteurs ont pensé et mis en oeuvre une politique d’éditorialisation des données sur les sites plus fréquentés. Les sites institutionnels et serviciels des acteurs publics, les plus consultés, ne référencent que très rarement des données.

Dans ce contexte de succès mitigé des portails généralistes, sont nés depuis quelques mois plusieurs portails “verticaux” : adresse, transports, cadastre, entreprise, etc. Il est encore un peu tôt pour juger, mais semblent ces services semblent appréciés du public, étant plus simples d’accès (monothématiques) et plus ciblés.

Le service public de la donnée, un dispositif original sur les jeux de référence

La littérature et le débat sur les données abondent de références à des jeux dont l’importance fait souvent consensus. Données essentielles, données de référence ou encore données pivot, leur vocabulaire et leur doctrine n’est pas tout à fait fixée mais la France a fait un grand pas en avant à travers le service public de la donnée.

Créé par l’article 14 de la loi pour une République numérique, il “ *vise à mettre à disposition, en vue de faciliter leur réutilisation, les jeux de données de référence qui présentent le plus fort impact économique et social. Il s’adresse principalement aux entreprises et aux administrations pour qui la disponibilité d’une donnée de qualité est critique.*” Des engagements de service sont par ailleurs précisés par un décret¹ qui mentionne notamment :

- côté producteur : la fréquence de mise à disposition des données (quotidiennement par exemple pour Sirene) ; le délai de réponse à une question (1 mois) et enfin celui

¹ <https://www.legifrance.gouv.fr/eli/arrete/2017/6/14/PRMJ1713859A/jo/texte>



d'information des usagers de toute modification substantielle des caractéristiques des données, de leurs modalités de mise à disposition, et de la structure de la base (3 mois).

- côté diffuseur, le taux de disponibilité du service et niveau de performance, soit une disponibilité des données en téléchargement 99 % du temps mensuel et 99,5 % du temps mensuel pour l'API.

Neuf jeux de données représentent aujourd'hui ce service : la Base adresse nationale, la base Sirene des entreprises, le Code Officiel Géographique, le Plan Cadastral Informatisé, le Registres parcellaire graphique, le Référentiel de l'organisation administrative de l'État, le Référentiel à grande échelle, le Répertoire National des Associations et le Répertoire Opérationnel des Métiers et des Emplois.

L'intention et les premiers jeux présentés montrent une démarche ambitieuse et utile. De fait, ces services connaissent un important succès d'usage. Pour autant, ce service "se construit progressivement" nous dit encore sa page de présentation. L'État devra ainsi transformer l'essai pour un service pleinement réussi :

- certaines voix s'élèvent sur des évolutions qui n'ont pas été discutées avec les usagers voire une baisse de la qualité de service (s'agissant par exemple de Sirene)
- la délivrance de ces données ne s'accompagne pas encore d'une politique de standardisation homogène (encodage, normalisation des CSV, etc.)
- à ce jour, il manque un suivi régulier publié qui permettrait aux usagers de valider les promesses de qualité de service.

Il serait également utile de mieux connaître le planning de déploiement du service (nouveaux jeux, etc.) ainsi que le process qui conduit au choix des données. La page d'accueil du SPD annonce un processus de construction "itératif avec les producteurs et les utilisateurs". La feuille de route du SPD dans une version future reste à déterminer.

Démocratisation et nouveaux usages

La demande : un sujet encore mal appréhendé

La connaissance de la demande est nécessaire à l'amélioration de l'offre et de connaissance des impacts. Mais elle traduit également une inquiétude récurrente des acteurs : "que vont-ils faire de nos données ?" entend-on souvent, notamment chez les métiers.

La connaissance de la demande est cependant difficile et l'open data souffre du manque d'analyses et de métriques. Le sujet est couramment traité comme un sujet d'analyse d'audience des portails ou bien de monographies — par ailleurs très utiles. Il est cependant exceptionnel de voir des acteurs analyser précisément la demande à travers des outils et méthodes adaptées : enquêtes, entretiens, etc. À ce jour presque aucun portail ne propose à ses usagers de signaler le pourquoi du téléchargement de telle ou telle donnée.

Les succès d'usage : des profils proches en qualité, granularité et couverture

Certains portails proposent judicieusement aux usagers, pour chaque jeu de données, de publier les résultats de leurs usages. Par exemple, tel jeu de données a permis la création



de telle application. Cette pratique présente un intérêt évident pour valoriser les usages et promouvoir toujours plus d'usages à venir. Le nombre de réutilisations de chaque jeu donne un indicateur, certes partiel, mais révélateur de la demande de certains jeux. La plateforme data.gouv.fr donne justement la liste des jeux ayant le plus fait l'objet d'usages référencés².

On y retrouve, pour un gros tiers, des **données de l'État publiées de longue date** et traditionnellement très utilisées comme les données INSEE (6/38), la base des codes postaux, ou certaines données spécifiques au travail et fonctionnement de grands ministères : Intérieur (résultats des élections, accidentologie), Santé (Fitness, démographie des médecins, médicaments), Économie et finance (Fantoir, Valeurs foncières)... L'open data a plus encore démocratisé leur usage et les réutilisations publiées montrent des usages nouveaux de ces données.

Mais on y trouve aussi de nombreuses **données d'un usage précédemment limité à certains publics**, que l'Open Data a fait sortir de l'ombre. La base des prix des carburants est un exemple emblématique : précédemment vendue pour plusieurs dizaines de milliers d'euros annuels, sa mise en Open Data en a développé les usages.

Enfin, pour un quart, cette liste fait ressortir le succès rapide de jeux de données **directement nés de la dynamique Open Data**. On compte tout d'abord la Base Adresse Nationale (BAN), projet impulsé par Etalab, mettant en partenariat l'IGN, La Poste et OpenStreetMap. L'important succès de cette base est aujourd'hui bien établi, la plateforme comptant chaque jour des millions de requêtes par des milliers d'acteurs.

Qu'ont en commun ces jeux de données fortement demandés ? Ils correspondent tout d'abord à des usages : usages économiques, recherche, études, etc. Mais ils ont surtout en commun de posséder une qualité de diffusion et de données qui facilite voire permet la multiplication de leurs usages.

De fait, toujours pour data.gouv.fr, ce succès se retrouve également dans les statistiques d'audience qui montrent que 90% du trafic porte sur 100 jeux de données³.

La longue traîne des données (presque) jamais utilisées

Que nous racontent les dizaines de milliers de jeux de données qui ne sont utilisés que rarement ou pas du tout ?

Une première famille concerne les **données (quasi) inutilisables**. D'un côté les éditeurs n'ont pas toujours pu ou su filtrer des jeux de données toujours plus nombreux. De l'autre, l'injonction à ouvrir des données ainsi que le nombre de jeux de données publiées — métrique simple et parlante — ont poussé à publier des données mal préparées dont l'usage est très difficile voire quasi impossible. Presque tous les portails en possèdent : documentation lacunaire, erronée voire inexistante, format propriétaire ou complexe/ésotérique, données à la qualité médiocre et non documentée. Les problèmes ne manquent pas ! Un seul d'entre eux suffit parfois à rendre un jeu inutilisable. Que pourra le plus beau jeu de données du monde sans sa documentation ? Quel potentiel d'usage pour des données parfaitement documentées mais dans un standard cryptique ?

² <https://www.data.gouv.fr/fr/datasets/?sort=-reuses&reuses=many>

³ Source : <https://stats.data.gouv.fr/>



Faut-il s'étonner que ces jeux ne soit jamais utilisés ? Ces jeux sont d'autant plus problématiques qu'ils représentent des quantités non négligeables et freinent les usages en rendant moins visibles les jeux de qualité.

La seconde famille est représentée par les **données hyper-locales ou hyper-spécialisées**. Documentées, réellement utilisables, elles attendent sagement leurs usagers et leurs usages. Faute de standards partagés et d'une couverture spatio-temporelle trop faible, leurs usages relèvent d'un hasard peu souvent observé.

Les radars pédagogiques : des usages mais peu d'usagers

Certains acteurs publient les données issues des radars pédagogiques (ils informent du dépassement d'une limite sans verbaliser). Ces données présentent une donnée passionnante qui pourraient intéresser des publics variés : agents de la voirie, de la sécurité, sociologues, géographes, comités de quartiers, aménageurs, conseil municipal, associations d'usagers du vélo ou de piéton, commerçants, etc. Il est aisé de saisir le potentiel de ces données dans la compréhension du territoire et du débat sur ses choix d'aménagement. Dans les faits, ces données attendent ce mouton rare qui saura les utiliser : ce geek ou curieux qui a poussé la porte du portail Open Data, qui réside sur le territoire, qui a une idée bien précise ou qui vient là pour s'amuser ; ou bien encore ce thésard qui travaille sur l'influence des radars sur la vitesse et croit encore pouvoir comparer des centaines de jeux de données (dont le format diffère à chaque commune) ; etc. Sur de petits territoires, cet entonnoir de conditions donne rarement des usages, plus encore dans un temps long.

Cette longue traîne des données inutilisées est installée, diffuse, invariable. Doit-on s'en plaindre ? Elle regorge aussi de trésors inestimables.



Partie 2 - Analyser: les impacts de l'ouverture des données publiques

5 grandes familles d'impacts

Nous avons identifié cinq grandes familles d'impacts de l'ouverture des données publiques, avec autant de questions ouvertes:

- les **impacts économiques**: l'open data a-t-il fait émerger un marché de la réutilisation des données ? peut-on évaluer quantitativement ce marché ? quels sont les acteurs économiques privés qui se sont emparés du sujet ?
- les **impacts démocratiques**: l'open data a-t-il renforcé la transparence de la vie publique ? est-ce un outil de surveillance des acteurs de la vie publique (élus, agents des administrations, ...) et/ou un outil de capacitation des citoyens ?
- les **impacts sur les organisations** qui ont ouvert des données ou les utilisent, et leurs relations avec des tiers: l'open data a-t-il permis de nouer des relations avec des écosystèmes nouveaux ? En quoi les données ouvertes modifient-elles les relations existantes, par exemple avec les délégataires de service public ?
- les **impacts sur les métiers, les fonctions et les méthodes de travail**: quels sont les métiers principalement impactés par l'ouverture des données ? Voit-on émerger de nouvelles fonctions, voire de nouveaux métiers ? L'open data a-t-il modifié les méthodes de travail au sein des organisations ?
- enfin, les impacts sur la diffusion d'**une culture de la donnée** au sein des organisations: l'open data comme brique première de la littérature des données ? La donnée ouverte a-t-elle ouvert la voie à des sujets comme la ville intelligente (smart city) ?

En préambule, il faut souligner que beaucoup de changements ont eu lieu sur la période considérée (la dernière décennie). De nouveaux thèmes liés au numérique sont apparus ou ont pris davantage d'importance (*smart city*, intelligence artificielle, *big data*). Le paysage administratif a lui aussi évolué (réforme territoriale introduisant de nouvelles compétences, ...) tout comme la société elle-même et ses attentes (transparence, ...). Nous avons donc veillé à distinguer dans les impacts que nous avons étudiés ceux qui sont directement imputables à l'ouverture des données publiques et ceux qui sont plutôt indirects, où l'*open data* a pu jouer un rôle aux côtés des autres évolutions survenues sur la période.

Les impacts économiques

Sur le plan économique, l'ouverture des données publiques a eu des impacts différenciés selon le marché considéré. Nous en avons listé trois: le marché de la donnée comme matière première (qui préexistait à l'ouverture des données publiques), le marché de la réutilisation (directe ou indirecte) des données publiques et enfin celui des entreprises qui ont fait de l'open data un élément de leur stratégie sans qu'elles y soient nécessairement contraintes par la loi.



Le marché de la donnée comme matière première: l'open data comme accélérateur de tendances déjà à l'oeuvre

Il existait un marché de la donnée comme matière première bien avant le mouvement d'ouverture des données publiques.

Les producteurs de données publiques, et en particulier les principaux d'entre eux (IGN, INSEE, DGFIP, Météo-France, ...) ont été parmi les premiers impactés par l'ouverture des données publiques, notamment en raison de l'instauration du principe de gratuité. Il est incontestable que leurs modèles économiques ont été impactés par la diminution progressive des recettes liées aux redevances perçues sur la vente de données publiques. Il serait cependant imprudent de ne pas tenir compte des évolutions intervenues sur le marché de la donnée et d'imputer l'ensemble des transformations au seul essor de l'open data.

En ce sens, on peut considérer de ce point de vue que l'open data est venu *accélérer* des tendances déjà à l'oeuvre et qui ont contribué à diminuer progressivement la valeur marchande des données publiques. Ces tendances sont: la fin de l'exclusivité (le fait que des sources de données alternatives existent, comme par exemple OpenStreetMap en matière de données cartographiques), l'émergence d'un marché mondialisé de la donnée (par exemple les données météorologiques produites par les services américains qui couvrent aussi le territoire européen) et la diminution de la propension à payer des acheteurs. Certaines grandes entreprises du numérique, qui avaient fait l'acquisition de bases de données publiques au début des années 2010 (Google, Microsoft, ...) ont ensuite développé leur capacité à mettre à jour les données de manière autonome.

D'un point de vue budgétaire, la situation est contrastée selon les producteurs concernés: certains, à l'instar de l'INSEE, ont obtenu des compensations budgétaires pour la perte de revenus sur la base d'immatriculation des entreprises (base Sirene), par exemple. D'autres, comme l'IGN font encore l'objet d'une injonction paradoxale, pour reprendre les termes de la Cour des Comptes⁴: ouvrir les données, tout en poursuivant la recherche de ressources propres.

Pour les clients historiques de ce marché de la donnée matière première, ceux qui s'acquittaient autrefois de redevances pour réutiliser les données, l'effet est multiple. Ils ont tout d'abord réalisé une économie nette, la ressource devenant gratuite. Mais certains d'entre eux ont pu par ailleurs regretter que la relation avec les producteurs se soit banalisée et qu'en perdant la relation commerciale ils perdaient aussi la capacité à agir sur la qualité des données fournies, par exemple. Enfin, les redevances agissaient aussi comme une barrière à l'entrée: la donnée devenant gratuite, de nouveaux réutilisateurs sont apparus, parfois en concurrence directe avec les clients historiques du marché.

⁴ Référé de la Cour des Comptes au Premier ministre sur la valorisation des données de l'IGN, de Météo France et du CEREMA: <https://www.ccomptes.fr/system/files/2019-03/20190311-refere-S2018-3287-valorisation-donnees-IGN-Meteo-France-Cerema.pdf>



Les gains économiques liés à la réutilisation des données

L'évaluation du potentiel macro et micro-économique de l'ouverture des données varie beaucoup d'une source à l'autre et donne lieu à de nombreuses spéculations. Il est cependant faux d'affirmer qu'aucune évaluation macro-économique sérieuse n'ait pu être menée. Bien au contraire, la synthèse réalisée en 2017 pour l'European Open Data Portal⁵ en a évalué plus d'une dizaine au niveau international. Ces évaluations ont été classées en trois groupes: conservatrices, ambitieuses et optimistes.

La synthèse des études macro-économiques

L'analyse menée pour l'European Open Data Portal classe les études menées en trois groupes:

- les évaluations dites "conservatrices" (Deloitte, Open Data Institute, Pollock) qui évoquent une valeur entre 15 et 20 milliards d'euros au niveau européen,
- les évaluations "ambitieuses" (Vickery, European Data Portal) évaluent la valeur entre 35 et 50 milliards d'€ au niveau européen,
- les évaluations "optimistes" (Mc Kinsey, Omydiar) évoquent une valeur de 200, voire 600 milliards d'€. Cette dernière évaluation, due à Mc Kinsey est unanimement jugée peu fiable.

Ainsi, en suivant les conclusions des évaluations jugées conservatrices, on peut considérer que les bénéfices économiques de l'ouverture des données en France sont évalués à *minima* autour de 3 à 5 milliards d'euros⁶.

Des études microéconomiques permettent d'affiner la compréhension des gains économiques, secteur par secteur ou industrie par industrie.

Dans le domaine des **transports**, TfL (Transport for London, l'autorité organisatrice des transports de la capitale britannique) a mandaté le cabinet Deloitte pour étudier les gains liés à l'ouverture des données de transport du territoire. Il en ressort que la grande majorité des bénéfices sont liés à des réductions de coûts (liés à des pertes de temps ou d'opportunités) davantage que des gains de monétisation directe. Dit autrement, les données ouvertes ont permis aux londoniens de gagner du temps, de mieux planifier leur trajet, de se voir proposer une offre d'information voyageurs très riche.

En France, l'ouverture des données de la base Sirene a permis de multiplier par 20 le nombre de téléchargements de la mise à jour quotidienne des données selon les calculs de

⁵

https://www.europeandataportal.eu/sites/default/files/analytical_report_n9_economic_benefits_of_open_data.pdf

⁶ A titre de comparaison: le marché du jeu vidéo, premier marché de l'industrie culturelle en France, est évalué à **5 milliards d'€** en 2018.



l'Insee⁷. Il reste encore à mener des travaux détaillés pour détailler et quantifier les impacts économiques de l'ouverture des grandes bases de données de référence.

Enfin, pour compléter ce tableau macro- et micro-économique, on peut mentionner d'autres études qui se sont attachées à décrire les mécanismes de création de valeur par l'open data. Ainsi, les principaux mécanismes identifiés⁸:

- l'**efficience**, notamment par la réduction des coûts de transaction: nombre d'acteurs, tant publics que privés, consacraient des efforts importants à identifier des données, leurs producteurs et à négocier les conditions de réutilisation avec ceux-ci. L'open data a permis de réduire drastiquement ces coûts de transaction,
- l'**innovation**: les données ouvertes constituent des matières premières qui permettent de créer de nouveaux services ou de modifier les services existants. On note notamment une dynamisation de la filière avale: de "nouveaux" réutilisateurs apparaissent grâce à l'open data, ils n'avaient pas auparavant la capacité ou le souhait d'acquérir des données,
- la **réduction de l'asymétrie d'information**, particulièrement bien illustrée récemment par l'ouverture de la base des valeurs foncières. Ces données n'étaient autrefois accessibles qu'à certains acteurs du marché (notaires, agents immobiliers). Leur plus grande disponibilité permet à chacun de connaître la réalité des transactions,
- la **boucle de rétroaction**: la publication d'une donnée permet aussi aux acteurs de modifier leurs comportements, par exemple dans le cadre d'une politique de name and shame. Mais cette boucle permet aussi, de manière plus opérationnelle, d'améliorer la qualité des données mises à disposition,
- la **collaboration**: les données peuvent aussi être l'occasion de collaborer entre différents acteurs. C'est particulièrement le cas pour les bases de données collaboratives, mais cela se révèle aussi vrai pour la création de standard de données.

Une chaîne de valeur qui évolue avec de nouveaux acteurs

L'open data a fait évoluer la chaîne de valeur traditionnelle de la donnée publique. On l'a vu notamment avec la question des producteurs et de leurs modèles économiques. Mais ce sujet ne concerne pas uniquement l'amont de la chaîne de valeur, mais bien son ensemble.

On constate ainsi que de nouveaux acteurs sont apparus avec l'émergence de l'ouverture des données publiques, en particulier ceux qui font fonction d'agrégation, de préparation ou de médiation autour de la donnée.

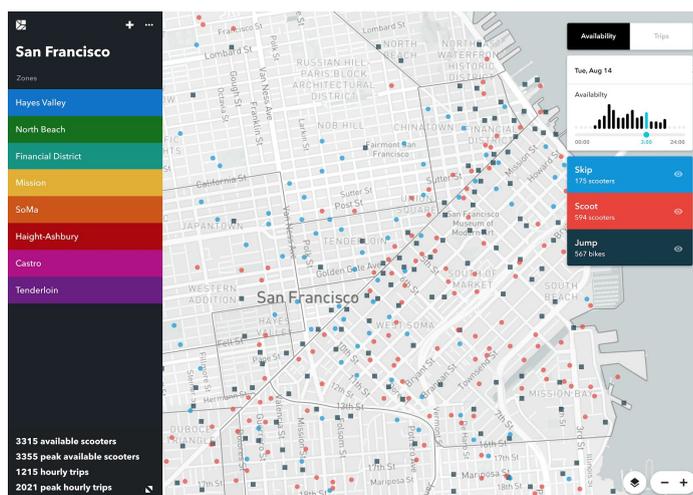
Les sociétés **Opendatasoft** ou **Koumoul** illustrent bien l'émergence de ces nouveaux acteurs. Elles éditent des solutions logicielles (portails, plateforme de visualisation de

⁷ "Accès aux données sur les entreprises: quelles conséquences du passage à la gratuité ?", Guillaume Mordant, revue Variances, février 2018

⁸ « The Generative Mechanisms of Open Government Data », European Conference on Information Systems, Jetzek, Avital, Michel, 2013



données, ...) et leur essor est en partie imputable à l'open data, même si elles adressent aussi d'autres marchés. D'autres acteurs, spécialisés dans un domaine particulier, commencent à émerger. Prenons l'exemple de la mobilité, où les sociétés **Remix** et Populus AI ne produisent pas des données, mais fournissent des services aux collectivités pour les traiter, les analyser et leur donner du sens.



La société Remix propose des outils pour analyser l'usage (ici des scooters et trottinettes électriques), réguler l'activité sur leur territoire et planifier l'offre de transports publics.

Il n'est pas d'ailleurs surprenant que la mobilité soit le premier domaine où ce type de nouveaux acteurs apparaissent, dans la mesure où l'ouverture des données de transport (qui n'est pas en France un long fleuve tranquille ...) est initiée depuis plusieurs années maintenant.

Il faut enfin noter que des acteurs déjà implantés se saisissent aussi parfois des données ouvertes pour proposer de nouveaux services, par exemple le média La Gazette des Communes (avec son observatoire), la plateforme Meilleurs Agents (qui intègre les données des transactions foncières à ses analyses et recommandations) ou encore Le Bon Coin.

Des acteurs privés qui ouvrent leurs données

Les acteurs publics ne sont pas les seuls à ouvrir leurs données, certains acteurs privés ont aussi des politiques d'open data plus ou moins volontariste - cf. le cahier n°5 du programme Infolab, 2017.

On peut considérer que cette ouverture s'est fait en trois temps. Premières concernées, les entreprises qui interviennent déjà en relation étroite avec les acteurs publics, notamment dans le cadre de délégation de services publics: énergie (ex. Engie), gestion de l'eau et des déchets (ex. Suez), transport public (ex. Keolis), etc. Certaines sont aussi contraintes de le faire par des réglementations sectorielles, d'autres ont répondu à des demandes pressantes de leurs commanditaires. Ensuite, les entreprises qui intègrent l'ouverture des données dans une approche plus générale: soit d'innovation ouverte (ex. BPCE), soit de responsabilité sociale (ex. MAIF). Enfin, il faut signaler les initiatives d'acteurs qui peuvent difficilement être



qualifiées d'open data mais qui s'en revendiquent cependant (ex. UBER Movement, DataVille AIRBNB). Dans ces cas-là, on voit que la donnée est le support, l'occasion d'engager des relations avec de nouveaux acteurs, notamment les territoires.

Les impacts sur la vie démocratique

Parmi les promesses initiales de l'open data, la revitalisation de la vie démocratique n'était pas la moindre ! La transparence, la participation devaient poser les bases d'une meilleure relation entre les citoyens et la chose publique au sens large.

Mais avant de tenter d'en évaluer les impacts, il faudrait se mettre d'accord sur les objectifs attendus. Or c'est bien là qu'est la première difficulté. Comme pressenti depuis le début des années 2000, l'open data n'est pas un mouvement [au sens de "groupe de personnes qui aspirent à un changement] homogène, en particulier dans sa relation à l'action publique:

- certains voient dans l'open data un outil au service d'une **société plus libérale**⁹: permettre à des acteurs privés ou associatifs de participer à la fourniture du service public, voire même à le mettre en concurrence,
- d'autres ne cachent pas leur **défi**ance vis-à-vis des administrations et des services publics jugés au mieux inefficaces et au pire corrompus. L'open data, dans sa dimension de transparence, leur est apparu comme un moyen de documenter leur position, un outil au service de leur cause.
- enfin, les derniers, tenants d'une vision plus positive des services publics, ont imaginé que l'open data pourrait introduire **plus de participation** et d'inclusion des citoyens dans la chose publique.

Selon que vous vous positionniez dans l'une ou l'autre de ces catégories - qui de plus ne prétendent pas à l'exhaustivité ! - votre jugement sur l'impact de l'open data sur la vie démocratique pourra être fort différente. Tentons quand même de donner de premiers éléments de bilan.

Un outil de contrôle de la vie publique ... et de ses acteurs

*"La lumière du soleil est le meilleur des désinfectants"*¹⁰. Cette expression, que l'on doit à Louis Brandeis (1856-1941), juge à la Cour suprême des Etats-Unis, a vite été adoptée comme slogan par ceux qui font la promotion de la transparence, comme la Sunlight Foundation. Ceux-ci ont vu dans l'open data un outil pour accroître la transparence de la vie publique et contribuer à la lutte contre la corruption.

Les ONG Transparency International et Web Foundation ont publié en 2018 une étude comparative sur 5 pays membres du G20 qui ont signé des engagements en ce sens (Afrique du Sud, Allemagne, Brésil, France et Indonésie). Elles ont évalué la disponibilité et

⁹ Cf. par exemple l'inspiration du projet de Big Society défendue à l'époque par David Cameron, Premier ministre britannique.

¹⁰ L'expression intégrale est: "Publicity is justly commended as a remedy for social and industrial diseases. Sunlight is said to be the best of disinfectants; electric light the most efficient policeman"



la qualité de dix jeux de données jugés clés pour la lutte anti-corruption : informations sur le lobbying, les dépenses publiques, les bénéficiaires effectifs des sociétés, les marchés publics ou encore le financement des partis politiques. Si aucun des 5 Etats étudiés n'a entièrement rempli ses engagements, la France se distingue cependant comme le pays qui a le plus progressé ces dernières années, en particulier avec l'adoption des lois Valter et Lemaire en faveur d'une plus grande ouverture des données publiques. Mais, comme le souligne Transparency International, l'open data en France n'est pas au service de la lutte contre la corruption. Les avancées de ces dernières années, comme la création de la Haute autorité pour la transparence de la vie publique (HATVP) doivent davantage au climat de suspicion créée par des affaires médiatiques (comme celle mettant en cause l'ancien ministre Cahuzac) qu'au mouvement d'open data.

La transparence relative aux élus et à leurs actions reste un sujet délicat. Si ces dernières années ont vu la publication d'un certain nombre de jeux de données (dont le Répertoire national des élus, longtemps réclamé), la transparence de la vie parlementaire fait toujours l'objet du combat mené par des associations comme Regards Citoyens.

Il est cependant un domaine particulier où l'action publique a engagé une véritable dynamique de transformation, c'est celui de la commande publique. Le Code des marchés publics a évolué: les acheteurs publics sont maintenant tenus de "rendre accessible sous un format ouvert et librement réutilisable les données essentielles du marché public". Les initiatives se multiplient sur les territoires (Régions Bretagne, Occitanie et Bourgogne France Comté) et à l'international à travers le PGO et le "Contracting Five", qui regroupent cinq pays (France, Royaume Uni, Mexique, Ukraine, Colombie) afin de travailler conjointement à l'ouverture et la mise à disposition des données relatives à la commande publique dans un standard international¹¹.

Un outil d'*empowerment* des individus et des collectifs ?

L'open data peut aussi être vu comme un outil, au service des individus et des collectifs qui pourraient se saisir des données ouvertes. Force est cependant de constater que les associations ayant intégré l'open data dans leur plaidoyer (comme Transparency International ou Bloom) le font davantage pour souligner le manque d'ouverture des données concernant leurs thématiques respectives (la lutte contre la corruption et la surpêche).

Certains collectifs produisent ou collectent eux-mêmes de la donnée pour documenter leur cause, parfois en l'absence de données publiques officielles sur le sujet. On peut par exemple citer le projet "A guichets fermés" de la Cimade, qui dénonce la difficulté à obtenir un rendez-vous en préfecture pour les personnes étrangères. L'association a développé un outil qui interroge très régulièrement les modules de prise de rendez-vous dans les préfectures et re-publie ces données en open data par la Cimade. Dans le domaine de la

¹¹ Cf. "La transparence de la commande publique", Romain Talès, blog Etalab, <https://www.etalab.gouv.fr/la-transparence-de-la-commande-publique-se-transforme-au-service-de-lopen-data>



démocratie sanitaire, l'association de patients Renaloo a pu mettre en avant des inégalités territoriales et sociales dans l'accès à la greffe du rein.

L'open data n'est pas une machine à scandale ... ?

Les bases de données constituent-elles une matière première pour révéler des scandales et jouent-elles à ce titre, un rôle dans le débat public ? Nous reprenons ici le concept de "machines à scandales" tel que décrit par Sylvain Parasio en 2012¹².

Il est souvent dit que l'open data ne constitue pas une machine à scandale, dans la mesure où les administrations seraient très précautionneuses dans le choix des données à ouvrir en priorité et que les données ouvertes seraient la plupart du temps inoffensives, contrairement aux données divulguées (*leaks*) ou collectées par des acteurs tiers (on pense notamment aux *Migrants Files*).

L'autre limite, majeure, à la mobilisation des bases de données dans le débat public est bien la question centrale des acteurs et de leur capacité (technique, organisationnelle, financière, ...) à se mobiliser sur ces questions. Traiter des données, les "faire parler", leur donner du sens demande du temps et des ressources qui ne sont pas à la portée de chacun dans la société civile. De manière frappante, hormis quelques acteurs historiques (comme Regards Citoyens ou Transparency International) ou des mobilisations particulières (la Cimade avec le projet à Guichets fermés déjà cité), force est de constater que les acteurs de la société civile (associations non gouvernementales, syndicats, mouvements politiques) se sont peu ou pas mobilisés sur cette question des données ouvertes.

Deux contre-exemples viennent cependant montrer tout le bénéfice qu'il y aurait à mieux mobiliser les données dans le débat public:

- L'analyse des inégalités de moyens dans les collèges publics, révélée par une étude de France Stratégie à l'aide des données ouvertes officielles: élèves, professeurs et personnels des collèges publics sont-ils équitablement répartis ?
- L'analyse des décisions d'obligations à quitter le territoire français (OQTF) a permis de questionner l'impartialité de certaines décisions judiciaires.

De quoi l'accroissement des demandes CADA est-il le signe ?

La CADA constate, dans son dernier rapport annuel, un accroissement des demandes liées à l'ouverture des données publiques. Plusieurs organisations de la société civile, ainsi que des journalistes, se sont aussi émus de l'accroissement du délai de traitement des demandes. De quoi cet accroissement est-il le signe ? Sans prétendre trancher cette question complexe, et dont les réponses sont sans aucun doute multifactorielles, on peut toutefois avancer trois hypothèses, pour certaines contradictoires. La première, généralement tenue pour vraie par l'ensemble des acteurs, considère que l'accroissement des demandes est en partie liée à la nouveauté des dispositions issues de la loi pour une République numérique. Les administrations, confrontées à un cadre nouveau et encore mal

¹² "Des machines à scandale" Éléments pour une sociologie morale des bases de données, Sylvain Parasio, revue Réseaux 2013/2-3 (n° 178-179)



connu (soulignons toutefois que la loi a été votée il y a plus de 3 ans aujourd'hui...), auraient tendance à s'en remettre de manière croissante à la CADA, en ne répondant pas aux demandes de documents ou de données.

La seconde hypothèse serait de penser que cet accroissement des demandes est le signe d'une défiance accrue de la société civile vis-à-vis d'administrations perçues comme opaques et peu soucieuses d'établir une relation apaisée avec les administrés, et au-delà, la société dans son ensemble. La troisième hypothèse serait de voir dans cet accroissement des demandes auprès de la CADA un signe de vitalité démocratique, la mise en oeuvre opérationnelle d'un droit de savoir sous une nouvelle forme, celle des données et des bases de données.

Il reste difficile de trancher cette questions de manière tout à fait claire. Précisons cependant deux points qui semblent essentiels. Le premier est que nous manquons de données pour mesurer exactement le phénomène. En effet, les demandes CADA ne représentent sans aucun doute qu'une partie de l'ensemble des demandes adressées aux administrations. Certaines demandes obtiennent une réponse favorable directement, sans que la CADA n'ait à en connaître, certains demandeurs renoncent aussi sans doute à saisir la CADA même en cas de refus ou de non-réponse de l'administration. Situation paradoxale: il manque donc des données pour quantifier, décrire et *in fine* expliquer ce phénomène que nous constatons.

Le deuxième élément susceptible de nous éclairer est de regarder le temps long. La loi CADA n'est pas nouvelle et, même avec le nouveau contexte introduit ces dernières années, les relations entre les administrations et le public ne sont pas nées il y a 3 ans !

Comment la technique conditionne et transforme le droit d'accès, depuis près de 40 ans maintenant.

La lecture des tous premiers rapports d'activité de la CADA montre bien que, dès le début des années 1980, la question de la technique conditionne l'accès à l'information. La technique en vogue, à cette époque, c'est la photocopie. La CADA qui va d'ailleurs jusqu'à affirmer que « l'absence de cet appareil [photocopieur] vide parfois de sens la liberté d'accès, aussi leur nombre devrait être multiplié, leur prix étudié en fonction du nombre d'exemplaires et leur délivrance obligatoire ».

La possibilité de reproduire des documents administratifs n'est pas seulement un moyen comme un autre d'exercer ce droit d'accès, cela ouvre d'autres usages possibles: photocopier un document (plutôt que le consulter sur place) permet de se constituer un dossier « au cas où ». A contrario, le format des planches cadastrales ne permet pas de les photocopier facilement, diminuant donc de facto la possibilité d'y accéder autrement qu'en allant les consulter sur place !

Les techniques ne sont pas neutres, au sens où elles façonnent l'exercice pratique du droit et les usages qui sont alors permis.



Les impacts sur les organisations (administrations, collectivités, etc.) et les relations entre acteurs

Les données ouvertes impactent les organisations et parfois aussi les relations entre les acteurs. Commençons tout d'abord par celles qui ouvrent des données, les administrations et les collectivités. Il a souvent été fait le constat que les portails des données ouvertes ont d'abord, en premier lieu, permis de faciliter la circulation de certaines données au sein des organisations. C'est l'idée que l'usage interne constituerait le premier usage, non au sens de "le principal" mais bien au sens de "celui qui vient en premier, en amont des autres".

Certains auront pu s'émerveiller que des dispositifs, pensés vers des publics extérieurs à l'administration, bénéficient ainsi d'abord à ses agents. Il faut sans doute dépasser cette première réaction. Tout d'abord, il convient de penser tous les publics dans leur ensemble: l'open data ce n'est pas faire "que pour les autres", pour les startups ou certaines associations du territoire. La donnée peut être, et est déjà, un outil au service de la performance interne. Autre élément, qui nous semble tout aussi important: l'usage interne et l'usage externe ne sont pas opposables, ils sont bien au contraire complémentaires. Une bonne donnée ouverte est aussi une bonne donnée utilisée en interne, parce qu'elle aura sans doute fait l'objet d'une meilleure documentation, qu'elle aura déjà été pensée pour que des tiers la réutilisent, même si les tiers en question sont des voisins de bureau. A l'inverse, nous n'avons vu que de très rares exceptions où une donnée peu ou pas utilisée en interne constituait une excellente donnée ouverte, bien qualifiée et bien documentée.

L'open data transforme aussi les relations entre acteurs, entre organisations. C'est particulièrement vrai pour les organismes de type observatoires publics. On assiste à un transfert de leur coeur de métier, moins centré sur la collecte des données ou l'établissement de partenariats ad hoc (avec les producteurs initiaux comme les CAF ou l'INSEE) et davantage d'efforts consacrés à leur valorisation non seulement sous la forme de notes ou de rapports, mais aussi dans la réalisation de tableaux de bord interactifs.

La mobilisation des écosystèmes, corollaire de l'ouverture des données sur un territoire, constitue aussi l'occasion, pour les collectivités et les administrations, de rencontrer un nouveau public, au-delà des espaces de discussion ou d'échanges que certains peuvent entretenir avec des usagers / bénéficiaires. On pense notamment à des formats tels que les *open data camps*, ou les *hackathons* qui, bien qu'ils soient tout à fait recommandés de discuter les productions réelles et durables, constituent toutefois des opportunités pour s'ouvrir à d'autres publics. En ouvrant des données on peut aussi ouvrir de nouvelles relations.

Enfin, la donnée ouverte peut aussi permettre de modifier des rapports de force existants. Citons l'exemple d'une grande région française qui accède maintenant à des données extrêmement détaillées (ligne par ligne, train par train) sur la ponctualité des trains régionaux dont l'exploitation est confiée à la SNCF. Le donneur d'ordre est mieux équipé, dans son dialogue avec le délégataire, grâce aux données. Il ne suffit toutefois pas



d'accéder à de nouvelles sources de données, encore faut-il disposer des compétences internes pour savoir les mobiliser.

Les impacts sur les métiers, les fonctions et les méthodes de travail

L'ouverture des données publiques a aussi impacté les métiers et les fonctions, avec une approche par cercles concentriques.

Les fonctions les plus directement impactées sont celles dont la gestion des données figuraient déjà dans leur métier, voire en constituaient la mission première. On peut notamment citer les fonctions liées au Système d'information géographique (SIG) ou les fonctions statistiques. On a pu observer des différences dans la manière dont ces fonctions se sont emparés du phénomène d'open data. Entre sentiment de dépossession et nouvelle opportunité de faire avancer leurs missions, les postures ont été parfois assez différentes. Ainsi, dans le domaine de l'information géographique, le mouvement open data a pu s'appuyer sur un socle juridique préexistant, celui de la directive Inspire. Les SIG ont d'ailleurs été longtemps les principaux contributeurs, en volume de données, aux portails open data des collectivités. Certains ont pu cependant s'émouvoir que l'open data n'accorde pas assez d'importance, au début, aux enjeux de qualité de la donnée ou de sa documentation. Mais aujourd'hui on constate que, sur la plupart des territoires, les initiatives open data s'intègrent en bonne intelligence avec les données géographiques des SIG. Car c'est une évidence: la plupart des données utiles comportent une dimension géographique, ne serait-ce que sous la forme d'une adresse ou de coordonnées géographiques.

Les relations ont parfois été plus complexes avec les services chargés de la production statistique et ceux qui ont pour mission d'observer les territoires ou les phénomènes (observatoires locaux ou nationaux, observatoires thématiques). Ils ont pu, dans un premier temps, être réticent à laisser publier des données brutes car elles ne répondent pas au standard de qualité et d'exigence qui est attendu des sciences statistiques. En un sens, une certaine culture statistique ne voyait pas d'un bon oeil que des données granulaires puissent être mise à disposition de tous, craignant que de fausses interprétations puissent en être tirées. On a pu sentir chez certains, un sentiment de crainte de dépossession face à l'émergence de l'ouverture des données publiques.

Nouvel acteur dans le paysage des agents administratifs, les chargés de mission open data constituent bien souvent une fonction inédite sur les territoires et dans les administrations centrales. Parfois rattachés au service informatique, à la communication ou encore à l'innovation numérique, ils ont dû développer un nouveau métier au sein des organisations, dans une fonction qui nécessite à la fois une transversalité interne (travailler avec l'ensemble des services, mais aussi manier l'informatique, la communication, l'animation de communautés) mais aussi une forte ouverture à des tiers (les communautés d'utilisateurs de données, les écosystèmes locaux, etc.).

Les données ont aussi constitué un facteur de changement dans les méthodes de collaboration et de travail. L'open data accompagne, plus qu'il ne précède, la mise en place



de nouvelles cultures de travail, centrées sur l'ouverture et le travail collaboratif. Le "faire ensemble / faire collectivement" est d'ailleurs l'un des traits constitutifs de la culture des données (développée au chapitre suivant). Certaines administrations, mais aussi des entreprises, se sont pleinement emparés des données comme un sujet de transformation des pratiques et des organisations. La donnée a le potentiel de modifier, de manière plus ou moins importante, les processus de prise de décision, en impactant notamment le "qui décide". Il ne s'agit bien sûr pas de retirer à l'autorité politique la capacité de prendre et d'assumer les décisions structurantes, mais plutôt de "démocratiser la décision" en permettant, à chaque échelon hiérarchique, de mettre à disposition les données qui permettent de prendre des micro-décisions, ou des décisions sur un périmètre ou une période donnée (par exemple moduler certains horaires d'ouverture des équipements publics en fonction d'événements exceptionnels attendus).

La diffusion d'une culture de la donnée: pour un nouveau pouvoir d'agir

L'open data a enfin produit un impact imprévu, en tous cas presque absent des discours de promotion initiaux : l'open data est l'un des premiers vecteurs, en qualité comme en quantité, d'une culture de la donnée — également qualifiée de "littératie de données", ou "d'acculturation aux données". Si les data sont aujourd'hui un puissant vecteur de développement économique, de capacitation citoyenne ou encore, de production de connaissances, elles suscitent également des inquiétudes légitimes, ainsi que des luttes de pouvoir. L'idée montante d'une « littératie » des données consiste à permettre à tout décideur, collaborateur, innovateur, entrepreneur, chercheur ou citoyen d'en comprendre les enjeux, d'en discuter les sources et les usages, et d'en tirer parti dans sa propre activité. À l'heure où le numérique parachève son entrée complète dans la société, comme phénomène social total, la culture des données serait ainsi une sorte de nouveau pouvoir d'agir, voire LE nouveau pouvoir d'agir.

Dans ce contexte, l'open data a rendu disponible de grandes quantités de données sur des thèmes extrêmement variés : elles sont capables de toucher tous publics de tous territoires. Ces données sont souvent récentes, en augmentant ainsi l'attrait. Ces données sont enfin éditorialisées : tout n'est pas parfait mais l'effort pour les rendre compréhensibles et "trouvables" est réel. En d'autres termes, l'open data a rendu les données non seulement visibles mais lisibles. Et ça change tout.

En rendant lisible d'énormes quantités de données, l'open data crée de très nombreuses possibilités de rencontre des publics avec ces données, mais aussi et surtout avec le monde de la donnée en général. De fait, l'observation de 10 ans de projets open data regorge d'exemples de publics a priori éloignés de la data, qui en sont devenus des usagers éclairés sinon régulier. Telle cette association de promotion du vélo à Marseille, qui progressivement, s'est saisie des données publiques urbaines et des données coproduites de la communauté OpenStreetMap pour en faire un terrain d'action et de dialogue avec la collectivité. En quelques années, communauté du vélo et communauté OpenStreetMap se sont rapprochées jusqu'à collaborer fréquemment voire se fondre pour partie. Les données open



data de la collectivité ne sont pas en reste, maintenant scrutée et analysée par ces passionnés de la petite reine.

Cet exemple montre enfin un autre aspect qui a été évoqué en creux dans la partie précédente : un producteur de données ouvertes s'acculture bien plus rapidement aux enjeux contemporains de la donnée. Ainsi de cette association sur le vélo, mais ainsi, également de toutes les collectivités et métiers producteurs de données ouvertes.

Produire des données ouvertes est un chemin qui nécessite d'avoir une vue très complète des problématiques de la donnée : problématiques techniques, standardisation, problématiques locales (langues, formats), problématiques juridiques, organisationnelles, de gouvernance, etc. Qu'elle soit amateur ou professionnelle, une démarche d'ouverture de données est une manière très intéressante de faire un bond en avant dans la culture de la donnée. De nombreux responsables open data des collectivités témoignent de ses changements culturels mais aussi du changement de leur propre culture.

Il y a encore quelques années, l'open data et la smart city étaient deux sujets presque orthogonaux, portés la plupart du temps par des équipes différentes. La vision centralisatrice de la smart city, comme un poste de pilotage de la ville, a encore cours dans de nombreux projets mais beaucoup d'acteurs ont désormais compris que l'intelligence d'une ville c'est d'abord l'intelligence de ses habitants. L'open data, longtemps ignoré voire parfois même méprisé des tenants d'une vision centralisée de la smart city, y fait aujourd'hui son retour en force. Nous irions même plus loin en nous demandant si, à travers ses immenses capacités d'acculturation, l'open data n'est pas aujourd'hui le premier pilier de la smart city.

En ouvrant leurs données, en s'acculturant donc aux données, les acteurs sont parvenus à comprendre l'enjeu stratégique des données. Chaque producteur de données dispose d'un actif stratégique qu'il doit employer à bon escient. Aujourd'hui un mouvement semble se dessiner d'acteurs publics stratèges de leurs données. Puisque les données sont le nouveau pouvoir d'agir, comme le montre chaque jour l'ubérisation toujours plus poussée de pans entiers de notre société — largement alimentée par les data —, les acteurs publics comprennent qu'ils doivent reprendre la main. L'open data est le premier outil pour les y aider.



9 défis pour la prochaine décennie de l'ouverture des données publiques

En parallèle de la réalisation de ce bilan critique, les participants du programme Open Data Impact ont aussi travaillé à l'identification d'un ensemble de défis.

Ces défis, qui correspondent à autant de questions pour la prochaine décennie de l'ouverture des données publiques, requièrent une mobilisation collective déterminée et constante. Nous les avons classé en trois thématiques principales: donner à voir / donner envie de se saisir de l'open data, développer une offre vraiment adaptée à la demande et réinterroger les frontières de l'open data.

Donner à voir, donner envie de se saisir de l'open data

L'ouverture des données n'a encore rien d'évident pour beaucoup de monde. Et pourtant de nombreux exemples, aujourd'hui pas suffisamment lisibles, nous montrent tout son intérêt. Comment le faire savoir et étendre toujours plus ses usages ?

Défi 1 : se donner des métriques pertinentes.

Comment "mesurer" ce que produit l'ouverture des données publiques ? Les deux types de métrique aujourd'hui utilisées (le nombre de réutilisations et la mesure de l'audience des portails open data) sont insuffisantes pour capter tout ce que produit l'open data. Il est temps d'imaginer et de tester d'autres métriques.

Défi 2 : les données ouvertes, pilier des villes et territoires intelligents, enfin expliquées et expérimentées par le plus grand nombre.

Comment raconter les usages ? Comment les faciliter, les rendre encore plus accessibles ? Beaucoup d'acteurs s'y essaient, partiellement, en ordre dispersé. Les portails intègrent progressivement plus de services. Ce défi propose de mieux donner à voir, de donner envie et de faciliter le passage à l'action.

Développer une offre de données vraiment adaptée à la demande

Défi 3 : rendre les données vraiment réutilisables.

Faut-il redire que la préparation représente 80% du temps consacré à l'usage de données ouvertes ? Et qu'un nombre considérable de jeux de données n'est quasiment pas réutilisable. Faut-il redire que sans standardisation, sans qualité de service, un usage professionnel régulier des données ouvertes n'est pas envisageable ? La facilité d'usage des données, un des plus grands défis de l'open data aujourd'hui ?



Défi 4 : pour un open data tiré par la demande !

L'énergie consacrée à ouvrir un maximum de données a probablement atteint ses limites. Il suffit d'observer ces nombreux jeux jamais utilisés. Peut-on imaginer à l'inverse un open data tiré par la demande ? par ce que l'on sait déjà avoir du succès ? Comment identifier, écouter cette demande ? Comment y répondre dans de bonnes conditions ?

Défi 5 : comment inciter les producteurs à s'engager durablement, autrement que par la contrainte ?

Un écueil possible de l'introduction du principe d'open data dans la loi serait de perdre le sens de cette obligation pour les acteurs publics. Ce qui est en jeu ici, c'est l'incitation durable des producteurs dans une démarche d'ouverture, parce qu'il nous semble que c'est l'une des clés du sujet pour la prochaine décennie. Les réutilisateurs attendent plus que le simple respect des obligations légales, ils attendent aussi un véritable engagement (sur l'amélioration des données, la pérennité de l'offre, l'effort de montée en qualité, etc.).

Réinterroger les frontières de l'open data

Après la première décennie d'ouverture des données, nous proposons de réinterroger les dogmes et les frontières de l'open data. Mieux connaître les réutilisations - sans pour autant fliquer les usagers -, imaginer un open data qui se matérialise autrement que par des portails open data, proposer des données qui suscitent le débat et qui ne sont plus inoffensives: autant de nouvelles pistes à creuser et à tester ensemble à travers 4 défis.

Défi 6 : Tracer l'utilisation sans l'entraver ?

Conditionner les usages à une identification est une des lignes jaunes initiales de l'open data. Est-il pourtant possible de tracer plus finement et plus systématiquement les usages sans les entraver ? Faut-il suivre les utilisateurs ... ou les utilisations ?

Défi 7 : Des données à l'information.

La notion de "donnée brute" s'est imposée très tôt comme la condition première de données "open data". Cette notion est pourtant fragile et nous savons depuis bien longtemps qu'une donnée a nécessairement subi des transformations avant d'être publiée. En allant plus loin on s'interroge sur la possibilité de publier des informations, plus faciles à comprendre et interpréter, en complément des données soit-disant brutes. Sans doute pouvons-nous aller plus loin encore en publiant des informations ciblées, raffinées, adaptées à des usages pratiques et immédiats, etc. Sans pour autant remettre en cause l'intérêt des données brutes.

Défi 8 : L'open data sans plateforme ni application.



Peut-on se suffire des portails open data ? Les données ouvertes ont-elles toujours vocation à produire une application ? Nous pensons que d'autres formes d'open data peuvent apparaître, en dehors des plateformes et des usages traditionnels.

Défi 9 : Des données qui grattent : pour un open data qui ne soit pas inoffensif !

Les pionniers de l'open data avaient intuitivement fait oeuvre de prudence et les publications courageuses comme les usages "poils à gratter" sont plutôt rares. Comment les susciter ? Quelles données posent problème ? Egalité femmes-hommes, lutte contre les discriminations, lutte contre le changement climatique: comment les données contribuent-elles utilement aux grands débats de société ?

PARTENAIRES



GRAND POITIERS
Communauté urbaine
grandpoitiers.fr



La Fing a le soutien de ses grands partenaires

