



HAL
open science

“ Données recherche publics ” : les politiques d’open data à l’épreuve de la réutilisation

Samuel Goeta

► To cite this version:

Samuel Goeta. “ Données recherche publics ” : les politiques d’open data à l’épreuve de la réutilisation. Presses Universitaires du Septentrion. L’action publique saisie par ses “ publics ”, pp.137-155, 2018. hal-02025883

HAL Id: hal-02025883

<https://hal.science/hal-02025883v1>

Submitted on 19 Feb 2019

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

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

Chapitre 6. « Données recherchent publics » : les politiques d'*open data* à l'épreuve de la réutilisation.

Samuel Goëta

Depuis 2007, et la conférence dite de Sébastopol, les politiques d'*open data* sont constituées en élément essentiel des politiques d'innovation numérique. La réutilisation des données très variées publiées sur les nombreux portails *open data* tels que data.gouv.fr permettrait selon ses promoteurs de renforcer la transparence, de nourrir l'innovation et de transformer les pratiques de travail des administrations. En France, l'ouverture des données a été consacrée en 2016 par la loi pour une République Numérique qui oblige les administrations de plus de 50 agents et les collectivités locales de plus de 3 500 habitants à ouvrir par défaut « les bases de données, mises à jour de façon régulière, qu'elles [les administrations] produisent ou qu'elles reçoivent et qui ne font pas l'objet d'une diffusion publique par ailleurs » et « les données, mises à jour de façon régulière, dont la publication présente un intérêt économique, social, sanitaire ou environnemental. » Cette obligation d'ouverture des données, dont la charge de travail pour les agents reste largement sous-estimée, est justifiée par l'idée qu'il y aurait une demande de réutilisation de ces données. Mais, en pratique, cette demande de données est-elle aussi évidente ? Comment se révèlent ces publics ? Ces publics réutilisent-ils les données par eux-mêmes ?

À l'opposé d'une vision qui considérerait que les réutilisations apparaîtraient d'elles-mêmes, l'enquête¹ que j'ai menée dans le cadre de ma thèse de doctorat en sociologie montre que des instruments divers instaurent, chacun à leur manière, les publics de l'*open data*. Ruppert a montré que les politiques d'*open data* « imaginent » des *data publics* qui visualisent et créent les interfaces qui rendent les données intelligibles à un public plus large et permettent de promettre un renouveau de la transparence publique (Ruppert 2012). Or cette enquête a révélé que l'absence de publics de données peut constituer un problème important pour les responsables de projets

¹Cette enquête a permis de rassembler plusieurs types de matériau, afin d'appréhender les multiples facettes du travail d'ouverture des données. Des entretiens approfondis ont été menés dans des collectivités locales, des institutions et une entreprise, auprès de personnes en charge de la mise en œuvre du programme *open data*, de gestionnaires de données et de responsables informatiques. La démarche a été complétée par l'observation de comités de pilotage de deux projets *open data* ainsi que l'observation participante d'un programme en cours de préparation. Enfin, une série de documents internes et externes (plaquettes, articles de presse, billets de blogs et commentaires) a été assemblée et analysée.

d'*open data* qui ne sont pas seulement évalués en fonction du nombre ou de la qualité des données ouvertes, mais aussi des réutilisations qui en sont faites par le public. Considérer l'ouverture des données comme une « politique de l'offre de participation » (Gourgues 2012) nous permettra de voir comment sont instaurés les publics de données qui constituent le postulat essentiel des projets d'*open data*. Dans ce chapitre, nous nous intéresserons aux instruments par lesquels les publics de données qui constituent le fondement des politiques d'*open data* sont instaurés. Nous verrons tout d'abord comment l'existence des publics de données constitue un postulat des politiques d'*open data* qui sont évaluées en fonction de la réutilisation des données, laquelle ne peut se réaliser sans que le public n'exploite les données et crée des services à partir de ces dernières. Nous nous intéresserons ensuite aux instruments par lesquels les agents en charge de l'ouverture des données des administrations produisent les publics de données.

Le public dans les politiques d'*open data* : un présupposé

Les politiques d'*open data* conçoivent généralement l'existence d'un public de données comme une évidence. En effet, si l'on retrace quelques-uns des grands moments de définition de l'*open data*, l'existence d'un public pour les données n'est jamais questionnée. En 2007, Rufus Pollock, l'un des acteurs essentiels de l'ouverture des données, fondateur du réseau mondial *Open Knowledge* à l'origine notamment de l'*Open Definition* (définition d'un savoir ouvert) et de la solution open source CKAN qui équipe les portails *open data* des États-Unis et du Royaume-Uni, écrit un billet de blog « *Give us the data raw, give us the data now.* ». Il y considère que, dès lors que les données brutes sont publiées, il ne fait aucun doute qu'elles seront réutilisées : « *many interfaces can be written to that data (and not just a web one) and it is likely (if not certain) that a better interface will be written by someone else (albeit perhaps with some delay).* » En 2009, Tim Berners-Lee, l'inventeur du web, a repris son appel à l'ouverture des données brutes lors d'une conférence TED. Lors de cette conférence, il a demandé au public de crier « *we want raw data* » (nous voulons des données brutes) à l'attention des gouvernements. Depuis cette intervention, l'existence d'un public demandeur de données brutes s'est imposée comme un aspect essentiel de l'*open data*.

Ce postulat a depuis été repris dans les politiques d'*open data*. Les 17 et 18 juin 2013 à Loughe-Erne en Irlande du Nord, les membres du G8 (Allemagne, Canada, États-Unis d'Amérique, France, Royaume-Uni, Italie, Japon, Russie) ont adopté une charte sur l'*open data*. Ce texte a notamment fixé comme objectif commun aux huit gouvernements signataires de faire de l'ouverture des données la norme et non l'exception dans les administrations. En France, cette charte a joué un grand rôle dans l'adoption du principe d'*open data* par défaut dans la loi pour une République numérique de 2016. Le préambule de cette charte détaille les bénéfices attendus de l'*open data* : création de services, transparence de l'action publique, meilleure gouvernance, amélioration du débat public, lutte contre la corruption, soutien à l'innovation des entreprises et de la société civile, prospérité renouvelée. Le

préambule fonde les cinq principes de la charte sur la demande du public, des citoyens qui réclament de pouvoir accéder aux données :

« Les citoyens exigent aujourd’hui de pouvoir accéder à l’information et aux services par voie électronique au moment et selon les modalités qui répondent à leurs attentes. C’est de plus en plus souvent le cas pour les données publiques elles-mêmes. Nous arrivons à un tournant qui marque le début d’une nouvelle ère au cours de laquelle il sera possible de se servir des données ouvertes pour acquérir des connaissances, générer des idées et produire des services afin de créer un monde meilleur pour tous². »

Cet extrait de la charte du G8 sur l’*open data* le montre bien, les politiques d’ouverture de données sont fondées sur le postulat de l’existence d’une demande de réutilisation des données. Cette demande de données est censée concerner l’ensemble des citoyens. Or, en pratique, on s’aperçoit très vite que ces politiques d’*open data* sont en fait associées à une « communauté », à un public spécifique, qu’il s’agit de faire vivre. Ce que soulignent notamment les éléments de langage d’Etalab, la mission en charge de l’ouverture des données de l’État. Lors de la sortie de la nouvelle version du portail data.gouv.fr, le site est présenté comme une plateforme qui permet de « faire vivre » les données et de rencontrer des innovateurs permettant de faire naître de nouveaux services. Dans sa communication, Etalab insistait tout autant sur la valeur des données publiées sur data.gouv.fr que sur celle de la « communauté » qui s’y active. Cette « communauté » et les réutilisations des données qu’elle est censée produire ne peut pas être réduite à un objet de communication. Son existence et son activité constituent un des trois indicateurs à partir desquels la mission Etalab est évaluée pour son action. En effet, dans le projet de loi de finances 2017, en plus d’ouvrir des données, ses agents doivent accroître le nombre d’utilisateurs actifs de la plateforme et de réutilisations recensées sur data.gouv.fr.

INDICATEUR 7.2

Ouverture et diffusion des données publiques

(du point de vue du citoyen)

	Unité	2014 Réalisation	2015 Réalisation	2016 Prévision PAP 2016	2016 Prévision actualisée	2017 Prévision	2017 Cible
Nombre de ressources en open data (site "data.gouv.fr")	Nombre	37 338	93 000	40 000	96 993	100 000	40 000
Nombre de contributeurs actifs (site "data.gouv.fr")	Nb de comptes actifs	NR	13 500	12 000	14 271	17 000	13 000
Nombre de réutilisations (site "data.gouv.fr")	Nb	1 008	1 400	2 500	1 445	1 700	3 500

Figure 1 : Indicateurs d’évaluation des politiques d’*Open data*

Ainsi, sans usagers des données, les objectifs qui sont assignés à la politique d’ouverture des données ne peuvent se réaliser. Ce problème ne se pose pas uniquement à l’échelle des gouvernements, mais aussi au sein des collectivités locales qui mettent en place des politiques d’*open data*. Par exemple, à Montpellier et à

²Source : <http://www.modernisation.gouv.fr/sites/default/files/fichiers-attaches/charte-g8-ouverture-donnees-publiques-fr.pdf>.

Rennes, les projets d'*open data* étaient assignés à des objectifs en termes de création de services.

« L'élu a très vite compris qu'il y avait un enjeu important et intéressant, c'est ce que ça pouvait donner lieu à la création de nouveaux services pour les habitants et c'était une bonne chose. Et c'était ça l'objectif qu'ils nous ont donné au départ. Ce n'était pas transparence politique ou développement de nouvelle filière économique. C'était plutôt voilà des services pour les habitants. » (Chef de projet *open data*, Rennes Métropole)

« Quand, je suis arrivé, Jacques [son chef] a dû me dire "ça serait bien que d'ici la fin de l'année, on soit à soixante jeux de données et quatre applications créées.". On ne me demande pas de faire du chiffre, mais ça nous donne un objectif, une ligne de conduite et moi, j'essaie de suivre. » (Chef de projet *open data*, Montpellier)

Si les données ne trouvent pas de public et ne sont pas réutilisées, c'est tout l'édifice de l'ouverture qui semble progressivement s'effondrer. Les responsables de projets d'*open data* ne doivent pas seulement instaurer ces données, mais doivent aussi souvent instaurer leurs publics.

Comment sont produits les publics des données ouvertes ?

Tim Berners-Lee a appelé le public de la conférence TED à crier « *we want raw data* », mais, passée l'exhortation, la demande de données brutes peut se révéler beaucoup moins criante lorsqu'on s'intéresse à la mise en œuvre des politiques d'*open data*. On l'a vu, les politiques d'*open data* « imaginent » les publics de données sans pour autant interroger leur existence pratique. Cela est particulièrement problématique alors que l'*open data* prend part à un projet majeur de renouvellement de la transparence et de la réédition des comptes dans lequel les publics de données pourraient compléter voire remplacer le contrôle de l'action publique réalisé par les experts et les fonctionnaires dont c'est la mission (Ruppert 2012). Pour les responsables de projet d'*open data* et les producteurs de données ouvertes, l'absence de publics constitue un problème très concret qui fait l'objet d'évaluations et qui, s'il n'est pas résolu, peut remettre en question tout l'édifice de l'ouverture des données.

Après avoir instauré des données, les agents administratifs et les responsables de projets d'*open data* doivent aussi souvent instaurer des publics pour ces données (Goëta 2016). Nous allons nous intéresser ici à trois instruments (Lascoumes, Le Galès 2005) qui contribuent à instaurer les publics de données : les métadonnées, la visualisation des données et les concours de réutilisation. Les deux premiers instruments s'articulent autour des portails de données ouvertes et se situent en amont de l'ouverture des données tandis que les concours se placent en aval pour inciter un public averti à créer des outils et des services qui permettront à un public sans compétence technique d'interagir avec les données.

Les métadonnées : anticiper les demandes des ré-utilisateurs des

données

Les métadonnées sont des informations standardisées et structurées qui qualifient les données et sont censées apporter toutes les informations nécessaires à leur compréhension et leur appropriation. Dans les projets d'*open data*, la production de métadonnées fait partie des préconisations officielles qui fondent les politiques d'ouverture de données. Au niveau national, les recommandations formulées par Etalab dans son *vademecum* de l'ouverture des données envisagent les métadonnées comme un outil essentiel pour que les usagers découvrent la « bonne » donnée parmi la masse de fichiers compris sur les portails d'*open data* et parviennent à l'utiliser.

« La qualification des métadonnées et l'indexation sont une étape essentielle pour faciliter la réutilisation des données publiques. Les données sont très difficiles à retrouver si elles ne sont pas indexées et elles sont difficilement réutilisables si elles ne sont pas décrites avec précision. Ces informations complémentaires décrivant les données sont appelées "métadonnées". Etalab propose ainsi des champs de descriptions normalisées à tous les producteurs de données publiques afin de leur permettre de spécifier le contexte et le contenu des données. Il leur est notamment demandé de caractériser leurs données (titre, description, mots-clés...) en répondant aux questions suivantes : qui a produit les données ? Quand les données ont-elles été produites ? Quelle est la période temporelle concernée ? Quelles sont les zones géographiques couvertes ? Quelles sont les thématiques des données ? Par ailleurs, pour faciliter la réutilisation la plus large possible des données publiques, Etalab recommande que tout jeu de données soit accompagné d'une description du contenu du jeu de données. Ce document annexe peut se révéler très important pour les ré-utilisateurs³. »

Au milieu de l'extrait précédent, on trouve une liste de questions qui correspondent aux champs proposés par le portail pour décrire les métadonnées. Ces champs sont liés à un standard de métadonnées, le *Dublin Core Metadata Initiative* (DCMI), qui réclame le remplissage d'un certain nombre de champs standardisés tels que le nom du jeu de données, sa description textuelle, sa couverture géographique, la période couverte par les données, le contact de la personne responsable, des mots-clés ou encore la date de mise à jour.

En sciences, et dans certains secteurs de l'administration comme l'information géographique où le partage des données est ancré dans les pratiques depuis plusieurs décennies, les métadonnées sont souvent considérées comme le « graal » (Edwards *et al.* 2011) par lequel la réutilisation des données ouvertes se déroulera sans difficultés. Dans les projets d'*open data*, l'ouverture concerne souvent des données qui n'étaient précédemment jamais sorties des administrations qui les produisent, ne serait-ce que pour circuler dans d'autres services. La production de métadonnées est donc un défi : comment qualifier ces données, expliquer les champs qui les composent et décrire les méthodes de production ? Formaliser ce savoir tacite des services peut demander un investissement considérable. Néanmoins, les

³Etalab, *Vademecum sur l'ouverture et le partage des données publiques*, septembre 2013.

recommandations officielles en la matière sont généralement peu contraignantes et les gestionnaires de données avec lesquels je me suis entretenu ne percevaient pas le remplissage des métadonnées comme un problème. En effet, les agents sont seulement tenus de remplir les champs normalisés par le DCMI, l'investissement dans la production de métadonnées décrivant chacun des champs composant les données relevant de la décision du producteur des données. Et pourtant, la production de métadonnées plus ou moins complètes constitue une des étapes clés lors de laquelle les agents peuvent concrètement décider d'inciter à la réutilisation des données en facilitant leur appropriation par le public. Sans elles, les ré-utilisateurs doivent comprendre par eux-mêmes les conditions de production des données et les catégories employées par les agents.

Pour les services dédiés à la production de données tels que les Services d'Information Géographique (SIG), le remplissage des métadonnées peut constituer un travail à part entière, pour lesquels certains agents disposent d'une véritable expertise acquise après plusieurs années de partage de données, et être soumis à la réglementation qui impose une normalisation des métadonnées. En particulier, la directive européenne INSPIRE sur les données environnementales de 2007 demande aux agents de remplir les métadonnées selon une norme commune et de décrire le contenu des données, les objets qui y figurent, selon des modèles standardisés. Lors d'une réunion interne de présentation de la nouvelle version du portail de la ville de Paris, une gestionnaire de données géographiques a ainsi souligné le décalage entre les pratiques de production de métadonnées en vigueur dans son service et celles mises en place pour le projet *open data*. Son service maintient déjà un catalogue de données dans lequel l'accès et la réutilisation sont conditionnés à la lecture de métadonnées extensives qui contiennent des restrictions d'usages. Selon elle, la plupart des données publiées sur le portail *open data* ne comportent pas une description suffisante des données et ne préviennent pas assez les ré-utilisateurs des limites de leur réutilisation.

« Les métadonnées, ça me parle beaucoup, c'est mon sujet, mon domaine. [...] Au niveau de Paris, j'ai du mal à me retrouver. Nous quand on nous a posé la question de la publication de nos données, tout le monde était extrêmement frileux. La donnée, c'est une donnée métier. Une fois qu'elle sera sortie, comment cette donnée va être interprétée si elle n'est pas suffisamment documentée ? » (Extrait d'une réunion interne de présentation du portail *open data*, ville de Paris)

Dans les services où le partage des données constitue une pratique nouvelle, la production de métadonnées fait rarement l'objet d'une telle expertise. Leur production constitue donc un travail inédit pour les agents qui doivent qualifier les données dont ils assument la gestion quotidienne. Le choix d'investir ou non dans la production de métadonnées complètes capables d'anticiper jusqu'à un certain point les besoins et les questions des ré-utilisateurs constitue ainsi une phase essentielle de l'instauration des publics de données.

La visualisation : transformer les données pour les rendre intelligibles à

un plus large public

Pour éviter que les données ne s'adressent uniquement à des publics disposant de connaissances techniques avancées, les portails d'*open data* proposent donc souvent des fonctionnalités de visualisation qui permettent aux usagers d'afficher les données sous la forme de tableaux, de graphiques ou de cartes sans même avoir à ouvrir le fichier dans un autre logiciel que le navigateur. Ces fonctionnalités sont au cœur de l'offre de certains acteurs économiques comme *OpenDataSoft* dont le service propose aux usagers de visualiser directement les données. Pour certaines institutions et collectivités locales, ces fonctionnalités sont un moyen, en apparence simple, d'élargir le public des données et de faire en sorte que la politique d'*open data* ne cible pas uniquement les personnes ayant acquis les compétences pour exploiter des données. C'est le cas de la région Île-de-France où j'ai pu assister à une réunion de l'ensemble des correspondants du réseau *open data* lors de laquelle les animateurs ont insisté à plusieurs reprises sur l'intérêt de ces fonctionnalités pour permettre à un public plus large de se saisir des données.

« Nous pouvons visualiser ces données quand elles sont géocodées sur une carte. [...] François complète "on est plus sur une prévisualisation de la donnée pour pouvoir la faire parler, la visualiser autrement que sous une forme tabulaire et avoir une première vision de ce qui a dans nos données. Alors il faut faire un minimum de traitement sur le jeu de données pour avoir des visualisations". » (Réunion *Open data BootCamp*, région Ile-de-France, 27 novembre 2013)

La dernière phrase l'indique, les fonctionnalités de visualisation n'apparaissent pas de manière automatique, elles réclament une première phase de traitement lors de laquelle les agents doivent transformer les données pour les adapter aux spécifications des services de visualisation des données. Pour saisir l'épaisseur de ce travail, tournons-nous vers un document produit par la région Île-de-France intitulé « bonnes pratiques sur Excel »⁴ qui indique comment produire des visualisations de données en utilisant le portail régional produit par la société *OpenDataSoft*. Cette dernière demande un certain nombre de transformations des fichiers produits dans un tableur comme Excel pour qu'ils soient lisibles par les machines et donc visualisables. Dans le document « bonnes pratiques sur Excel », l'équipe en charge de l'ouverture des données de la région demande ainsi de ne pas faire figurer de valeurs en dehors du tableau de données. Ces valeurs sont qualifiées d'« orphelines » : pour exister selon ces spécifications du format CSV, elles doivent figurer dans un autre tableau, dans une autre série d'enregistrements (Figure 2).

⁴Christophe Libert Idf. (2013). *OpenData: quelques bonnes pratiques sur Excel*. Accédé à l'adresse <http://fr.slideshare.net/christophelibertidf/bonnes-pratiquesexcel-cc27juin2013> le 24 octobre 2013.

- **Attention aux données « orphelines » !**

	A	B	C	D	E
1	Directions	1er semestre 2012	2e trimestre 2012		
2	Service A	25 368 €	16 357 €		26,03%
3	Service B	35 987 €	19 963 €		33,70%
4	Service C	14 555 €	8 350 €		
5					
6	NB. Suppression le 3/03/12 du dispositif...				

Figure 2 : Recommandations sur la structure du fichier.

Source : Extrait du document « bonnes pratiques sur Excel » de la région Île-de-France.

Pour que le fichier soit correctement interprété, les informations qui figurent sur la première ligne doivent être effacées ou déplacées si elles ne font pas partie du tableau de données comprenant les enregistrements (Figure 3).

- **En-têtes sur la 1ère ligne (= titres de colonnes)**

	A	B	C	D	E
1	Unité communication		Service innovation numérique		
2					
3		ID	Année	Budget	
4		CP-13	Année 2013	295 562 €	
5		CP-12	Année 2012	183 687 €	
6		CP-11	Année 2011	255 665 €	
7		CP-10	Année 2010	199 355 €	
8		CP-9	Année 2009	222 887 €	
9		CP-8	Année 2008	231 300 €	
10					
11					

Figure 1 : Recommandations sur la structure du fichier.

Source : Extrait du document « bonnes pratiques sur Excel » de la région Île-de-France.

Les gestionnaires de données doivent aussi ne pas faire figurer de cellule vide dans la première ligne pour que chaque colonne comporte un titre comme le réclament les fonctionnalités de visualisation du portail.

- **Pas de cellule vide dans les titres de colonnes**

	A	B	C	D
1	ID	Année		Budget
2	CP-13	Année 2013	25 368 €	295 562 €
3	CP-12	Année 2012	35 987 €	183 687 €
4	CP-11	Année 2011	14 555 €	255 665 €
5	CP-10	Année 2010	16 357 €	199 355 €
6	CP-9	Année 2009	19 963 €	222 887 €
7	CP-8	Année 2008	8 350 €	231 300 €



Figure 2 : Recommandations sur la structure du fichier.

Source : Extrait du document « bonnes pratiques sur Excel » de la région Île-de-France.

Les opérations évoquées précédemment permettent uniquement l'affichage des données sous la forme d'un tableau comportant des fonctionnalités de tri ou sous

forme de graphiques. Pour que les données soient visualisées sous la forme de cartes, les agents doivent effectuer de nouvelles transformations afin que les adresses contenues dans les fichiers puissent correspondre à des coordonnées géographiques qui puissent être affichées sous la forme de points sur une carte. Cette opération dite de « géocodage » impose une nouvelle couche dans le travail de transformation des données. Bien qu'ils s'appuient sur les interfaces de programmation (API) des services de cartographie en ligne comme *Google Maps* ou *Open Street Map* pour automatiser la mise en correspondance des adresses avec les coordonnées géographiques, les agents doivent intervenir à plusieurs reprises pour uniformiser le format des adresses, corriger les erreurs et contrôler les résultats. Ce processus crée généralement des erreurs et des incohérences qui sont prises en compte par les API de géocodage et par les responsables de projet *open data* qui peuvent mettre en place des procédures de nettoyage et de validation des données.

La visualisation crée donc de nouvelles frictions dans le processus de l'ouverture des données. Du point de vue des agents, la prise en charge de ce travail dans le processus de l'ouverture constitue un nouvel investissement (Thevenot 1986) dans l'intelligibilité des données par les publics. En effet, ces fonctionnalités de visualisation anticipent que les usagers ne disposent pas nécessairement des compétences techniques ni de la disponibilité pour exploiter et visualiser ces données à travers des outils ad hoc. C'est un certain type de publics de données qui est inscrit à travers l'intégration de ces outils à même les portails. À travers le cas du *Transparency Agenda* (TA) du gouvernement britannique, Ruppert (2012) a montré comment les politiques d'*open data* produisent des publics de données en érigeant la visualisation de données comme une technologie neutre pouvant renouveler les conditions de l'*accountability* de l'État.

« La production des publics de données est médiée par des développeurs informatiques, des journalistes, des *think tanks*, des lobbyistes, des lanceurs d'alerte, des visualiseurs de données et des blogueurs qui sont les experts prédominants à la place des auditeurs, des analystes et des statisticiens. [...] L'Initiative sur la Transparence positionne les nombres, les feuilles de calcul et les outils de visualisation comme des technologies neutres qui permettent la production de récits « objectifs » des « faits » sans l'intervention des experts⁵. »

En créant un agencement entre les données de l'État et des *data publics*, les technologies de médiation fondées sur la visualisation et l'interaction avec les données créent un nouvel agencement entre les données de l'État et leurs publics. Ces outils, souvent présentés comme aussi « neutres » que les données, configurent le citoyen comme un acteur de l'évaluation et de l'audit des politiques publiques, disposant des outils et des données brutes, inaltérées permettant d'inspecter le

⁵Traduction personnelle de l'auteur : « The production of data publics is mediated by software developers, journalists, think tanks, lobbyists, watchdog organisations, data visualisers, and bloggers who are the predominant experts rather than auditors, policy analysts, academics and statisticians. [...] TA similarly positions numbers, spreadsheets and visualisation software as neutral technologies that enable the production of "objective" accounts of "facts" without the intervention of experts » (Ruppert 2012).

fonctionnement de l'État. Mais les cas précédents montrent que la production des publics ne consiste pas seulement à la mise en place d'un nouvel agencement. Elle s'effectue à travers un travail très concret effectué par les agents qui adaptent les données aux réseaux sociotechniques qui permettent de produire ces publics. Ce travail se poursuit en aval de l'ouverture des données par des dispositifs qui incitent le public à produire des services.

Les publics des concours : des assemblages temporaires

En 2014, l'association Regards Citoyens épinglait dans un billet de blog la Direction de l'Information Légale et Administrative (DILA) pour avoir publié des données ne respectant pas les grands principes de l'*open data* tout en communiquant sur l'ouverture de ses données en organisant plusieurs événements incitant le public à réutiliser les données. Sans entrer dans le détail de cette affaire, le billet de blog affirmait que si les administrations publiaient leurs données en respectant les principes de l'*open data*, les données devraient trouver leurs publics sans même qu'il n'y ait besoin d'avoir recours à de tels dispositifs :

« Si les données ont bien été libérées sous conditions *Open data*, les réutilisations arriveront sans doute d'elles-mêmes. Ne perdez pas donc votre temps avant même l'ouverture à préparer des communications, hackathons, sites officiels de réutilisation... [...] Les jeux de données que vous avez rendus publics vont certainement intéresser des ré-utilisateurs⁶. »

Dans bien des situations, le principe évoqué précédemment selon lequel les réutilisations arrivent « d'elles-mêmes » ne se réalise pas comme promis. En effet, utiliser des données ouvertes en vue de la création de services nécessite des compétences avancées en termes de développement informatique. Sans compétence en développement informatique, elle se résume à de la visualisation de données sous forme de graphiques ou de cartes basiques malgré la disponibilité d'outils grand public. Dans bien des situations, les responsables de projet d'*open data* se rendent compte que les développeurs ne créent pas d'eux-mêmes des services qui exploitent les données mises à disposition.

« On crée un petit peu un écosystème d'innovation et ça, c'est vraiment important pour qu'un *open data* puisse naître parce que, s'il n'y a pas un terreau derrière, il ne se passe rien. » (Responsable de l'innovation, Montpellier)

Lorsque la taille de la ville est réduite et que l'industrie informatique n'est pas très implantée, identifier et mobiliser les possibles réutilisateurs des données devient une tâche à part entière du travail du chef de projet *open data*. Les concours constituent une des solutions apportées au problème qui consiste à trouver un public pour les données. Ces concours incitent, de manière financière ou symbolique, les

⁶Regards Citoyens. (2014). Apprenons des échecs de la DILA, épisode 1 : « Comment faire de l'*Open data* ? ». Accédé en juin à 2014 à l'adresse <http://www.regardscitoyens.org/apprenons-des-echecs-de-la-dila-episode-1-comment-faire-de-lopen-data/#guide>.

développeurs et les entrepreneurs à réutiliser les données sous la forme de services et d'applications. À travers ces événements compétitifs, souvent qualifiés de marathon de développement ou « hackathons » quand ils sont limités à quelques jours, la participation par le développement de services est incitée, voire stimulée par l'attribution de prix dont la valeur est soit financière soit symbolique. Ces prix sont généralement attribués par un jury composé d'usagers et/ou d'experts du numérique. Historiquement, les concours répliquent un des rites des hackers, les « marathons » de développement qui se déroulent pendant les conférences des communautés issues du logiciel libre. Lors de ces événements, les participants se regroupent pour développer les logiciels pendant d'intenses sessions d'écriture de code : « les participants s'adonnent frénétiquement au hacking et rendent visibles les liens qui existent entre eux, leur donnant une plus grande intensité » (Coleman 2013).

Les concours imposent des règles et encadrent les productions du public. Tout d'abord, cela peut paraître évident, mais les porteurs de projet doivent obligatoirement exploiter au moins un jeu de données ouvertes par l'organisation à l'origine du concours. Cette condition est essentielle pour que de nouveaux assemblages se créent entre les données et leurs publics et donc répondent à la problématisation. Par ailleurs, les concours imposent souvent des conditions quant aux types de résultats qui peuvent être primés. À Rennes, l'enveloppe de 50 000 € était distribuée en six prix thématiques correspondant aux critères établis par les partenaires du concours⁷. Par exemple, le prix de l'accessibilité attribué par la région Bretagne récompensait « l'application qui favorisera l'accès aux services ou aux transports pour les personnes à mobilité réduite, déficients visuels ou auditifs » ; celui de l'éco-mobilité soutenait « les applications favorisant les modes de déplacements doux ». Les définitions des prix orientent ainsi les participants vers certaines données, souvent dans le domaine des transports, et réclament parfois un certain type de réutilisation, sous la forme d'applications mobiles ou de visualisations de données. La définition des prix du concours constitue une nouvelle épreuve lors de laquelle les responsables de projet d'*open data* doivent se positionner quant à l'utilité des projets d'*open data* et à leurs objectifs. En encadrant la participation dans le dispositif du concours vers la création de service, c'est un certain type de publics de données qui émerge, non pas pour favoriser la transparence et l'*accountability*, mais pour développer des services qui complètent voire pour certains pourraient se substituer au service public (Bates 2012 ; Birchall 2014). Le dispositif du concours encadre donc la participation : les réutilisations qui échappent à ses modalités sont de fait défavorisées et la probabilité de leur création sera plus faible en l'absence d'un dispositif de soutien.

Au-delà des développeurs et des porteurs de projet, le concours permet aussi d'instaurer un autre public : celui des usagers de service. Dans la communication qui accompagne les projets d'*open data*, les données ne sont pas uniquement présentées comme des ressources mises à disposition de leurs usagers directs, qui sont souvent

⁷Rennes métropole en accès libre, « Les prix et le jury », <http://www.data.rennes-metropole.fr/le-concours/les-prix-et-le-jury/>, consulté le 19 novembre 2013.

qualifiés de « réutilisateurs », mais aussi comme une offre de service qui peut toucher un public large, bien au-delà des personnes coutumières de l'exploitation de données.

« On fait des petits fascicules, ça nous permet de communiquer. En fait, les gens, quand tu leur parles d'*open data* (là je parle de citoyens normaux), ils ne savent pas ce que c'est généralement et puis, quand je leur explique, ils ne comprennent pas forcément. Mais quand tu leur montres que ça sert à faire une application pour calculer l'itinéraire pour les aveugles, que c'est parce que les données sont ouvertes, que les gens peuvent faire ça, là, ça devient de suite beaucoup plus concret. » (H.B., Chef de projet *open data*, Montpellier)

Comme dans le cas évoqué précédemment des outils de visualisation intégrés dans les portails, les services et les applications qui sont créés lors des concours contribuent à produire de nouveaux publics de données. Comme l'a bien montré Ruppert (2012), ces technologies de médiation fondées sur l'interactivité et la visualisation des données produisent un certain type de public qui n'interagit pas avec les données, mais avec des « boîtes noires » qui orientent l'utilisateur vers certaines conclusions, vers certains « faits ». Elles produisent des publics multiples composés d'une part des « ré-utilisateurs », le public qui produit les interfaces par lesquelles les données sont rendues lisibles et d'autre part des usagers qui interagissent avec des services.

Néanmoins, l'instauration des publics par les concours n'a rien d'acquis. Penser les concours avec les concepts d'intéressement et d' enrôlement souligne le caractère temporaire de ces assemblages. À Rennes, à la suite du concours, de nouvelles données ont bien été publiées sur le portail à la suite du concours. Mais ces nouvelles données ne semblent pas trouver de publics contrairement à celles qui avaient été publiées lors du concours. Alors que le concours avait fait émerger des services et des publics, la faible réutilisation des nouvelles données publiées amène des questionnements en interne quant à la persistance des publics de données. Un membre du jury a décrit la phase qui suivait le concours comme un « trou noir » dans lequel le travail de production des publics qui animait une partie de l'administration municipale lors du concours a perdu rapidement en priorité.

« Après le concours, c'est le trou noir. Pourquoi c'est le trou noir ? Le concours est sorti donc, après, à cette époque-là, il y a plein d'interrogations et on va dire qu'autant les bonnes étoiles étaient alignées à une époque autant là d'un seul coup elles commencent à se désaligner chez tous les acteurs. [...] L'impression de réactivité qu'on avait donnée jusqu'ici, d'un coup on rentre dans un schéma où c'est beaucoup plus lent tu vois. Tout ce qui avait été pensé dans l'après-concours et qui était pour moi vraiment indispensable, par exemple une vraie communication autour de certaines applis, n'a pas été fait. » (D.V., Responsable associatif, Rennes)

En l'absence de nouveau soutien technique, financier ou communication, certains développeurs n'ont plus mis à jour leurs services ce qui a remis en cause le public des usagers. En effet, les évolutions régulières des systèmes d'exploitation mobiles

peuvent rendre une application rapidement obsolète pour les usagers. Par ailleurs, dans le domaine des transports, si les applications n'ont pas été mises à jour avec la dernière version des données, les services perdent leur utilité pour des usagers en situation de mobilité à la recherche d'horaires fiables. En l'absence d'intéressement à la réutilisation des données, certains porteurs de projet ont progressivement abandonné le développement de leurs services que ce soit à Rennes ou à Montpellier.

« Les applications manquent de pérennité. [...] Comme il n'y a pas de financement et bien ce n'est pas réparé. Et comme ce n'est pas réparé et bien ça tombe en panne, ça n'évolue pas. Il n'y a pas de com en plus : les applications c'est bien si les gens savent qu'elles existent ; si elles ne savent pas, elles sont mortes. Pour faire connaître une application, il faut faire beaucoup de buzz et beaucoup de communication. » (K. B., Responsable innovation numérique, Montpellier)

Par rapport aux fonctionnalités de visualisation ou aux métadonnées proposées dans les portails, l'instauration des publics n'est pas uniquement permise par l'instrument du concours, elle est stimulée et incitée par un dispositif d'intéressement qui crée un assemblage temporaire entre les données et ses publics. Dès lors que l'intéressement s'arrête, le public aussi.

Conclusion

Pour les responsables politiques et administratifs de projets d'*open data*, la réutilisation des données constitue une épreuve qui interroge les agencements sociotechniques qui permettent l'ouverture des données. En effet, la mise à disposition de données librement réutilisables constitue le résultat d'un travail souvent invisible grâce auquel les agents transforment les fichiers de gestion de l'administration en un objet de savoir pouvant constituer le matériau de services, de visualisations ou d'articles journalistiques susceptibles de nourrir l'innovation, de renouveler la transparence ou de moderniser l'action publique. Ce travail est généralement justifié par le fait que les données trouveront sans difficultés un public de développeurs, entrepreneurs, data journalistes ou encore militants associatifs qui produiront des réutilisations, c'est-à-dire des interfaces qui présentent les données sous une forme réputée intelligible par tous. La première partie nous a montré que les acteurs qui ont formulé les premiers grands principes de l'*open data* ont fait de l'existence de ce public de « ré-utilisateurs » un présupposé, un postulat qui n'est généralement pas interrogé. Ce postulat se retrouve dans de nombreux textes officiels qui justifient les politiques d'*open data*, il devient même parfois un indicateur de succès. Le nombre de réutilisations produites par « la communauté » est ainsi un des trois indicateurs qui figure dans les projets de loi de finances successifs pour évaluer l'action de la mission Etalab en charge de l'ouverture des données de l'État. Ses agents, comme celles et ceux qui dans les collectivités locales chapeautent un projet d'*open data*, doivent donc susciter l'apparition de ce public de ré-utilisateurs pour atteindre les objectifs qui leur sont fixés.

Face à l'épreuve de la réutilisation des données, les responsables de projet d'*open data* mobilisent des instruments qui contribuent à s'assurer de l'émergence de ré-utilisateurs. Les deux premiers cas s'inscrivent dans l'espace du portail et ne les placent directement en situation d'échange avec le public. Nous avons ainsi pu voir que les métadonnées sont souvent présentées comme une « baguette magique » censée permettre une réutilisation sans frictions. Cet instrument projette un public très générique de ré-utilisateurs contrairement aux fonctionnalités de visualisation qui permettent aux projets d'*open data* de cibler un public plus large. En effet, lorsque les portails offrent des fonctionnalités de visualisation des données, le public peut interagir directement avec ces dernières sans avoir recours à des outils tiers. Mais ces fonctionnalités ne peuvent s'activer automatiquement, elles réclament préalablement un travail de préparation des données et de configuration des outils qui en assurent la visualisation.

Enfin, le cas des concours présente une autre forme d'instauration des publics. Là où les métadonnées et les outils de visualisation agissent sur l'intelligibilité des données pour les usagers, les concours stimulent la création d'applications et de services. Par des incitations financières et symboliques, ils tentent de créer des assemblages entre les données et les machines qui se révèlent souvent temporaires et incertains. Du point de vue des responsables de projet d'*open data*, les concours peuvent déjà permettre de fournir la démonstration de l'existence des publics de données. Mais ces démonstrations restent fragiles et l'instauration du public qu'elles permettent est finalement provisoire. Par ailleurs, les concours cadrent la participation : ils favorisent l'émergence d'un public essentiellement d'entrepreneurs qui témoigne de l'orientation de certaines politiques d'*open data* vers la création de valeur économique plus que vers les enjeux démocratiques liés à la transparence de l'action publique et à la participation politique de citoyens informés par les données. En encourageant les développeurs à exploiter les données publiques pour créer des services, les concours organisés par les institutions publiques peuvent prêter à la critique d'une libéralisation des services publics masquée sous une couche technologique supposément neutre (Bates 2012).

On peut ainsi s'interroger sur la logique et la portée des nombreux *hackathons* qui ont été conduits par les principaux ministères français au cours de l'année 2016 sur des sujets très variés (impôt, garde nationale, pesticides, douanes...) Un événement de ce type a même été organisé en décembre 2016 pendant le sommet de l'*Open Government Partnership* au sein même de la salle des fêtes du Palais de l'Élysée pendant toute une journée. Il faudrait retracer les moments et les acteurs qui ont contribué à la prolifération de ce type d'appels à la participation des citoyens. Ce chapitre pourrait être prolongé sur le terrain en conduisant des ethnographies de ces événements pour comprendre les motivations des participants acceptant de travailler gratuitement dans leur temps de loisir ou de travail, décrire les interactions entre les différents acteurs impliqués (participants, prestataires, élus, agents publics...).

Plus fondamentalement, on pourrait interroger la séparation entre un public de ré-utilisateurs jugé compétent pour manier les données publiques ouvertes et le public plus large réduit à un rôle de consommateur de services. La notion de *data literacy*,

qui a émergé une première fois dans un rapport de l'Organisation des Nations Unies de 2014 intitulé « *A world that counts* », va à l'encontre de cette division du public. En effet, la *data literacy*, promue en France par la Fondation Internet Génération (FING) qui a organisé à Aix-en-Provence en septembre 2017 la deuxième conférence internationale sur le sujet, considère que la capacité de trouver des données, de les comprendre, les interpréter, les visualiser afin de soutenir un argument doit devenir une compétence fondamentale au même titre que lire ou écrire. Les formes d'action publique qui découleront de cet idéal d'inclusion et de citoyenneté technologique (Barry 2006) mériteront elles aussi toute notre attention.

Bibliographie

- Barry, Andrew, *Political Machines. Governing a Technological Society*, London, The Athlone Press, 2001.
- Bates, Jo, « "This is what modern deregulation looks like": co-optation and contestation in the shaping of the UK's Open Government Data Initiative » *Journal of community informatics*, 8 (2), 2012.
- Birchall, Clare, « Radical Transparency? », *Cultural Studies*  *Critical Methodologies*, 14 (1), 2014, p. 77-88.
- Coleman, Gabriella E., *Coding Freedom. The Ethics and Aesthetics of Hacking*, Princeton University Press, 2013.
- Edwards Paul, Mayernik, Matthew S., Batcheller, Archer L., Bowker, Geoffrey C., Borgman Christine L., « Science friction: Data, metadata, and collaboration » *Social Studies of Science*, 41 (5), 2011, p. 667-690.
- Goëta, Samuel, *Instaurer des données, instaurer des publics : une enquête sociologique dans les coulisses de l'open data*, Thèse de doctorat en sociologie, Paris, Télécom ParisTech, 2016.
- Gourgues, Guillaume, « Penser la participation publique comme une politique de l'offre : une hypothèse heuristique », *Quaderni* 2012, 79, p. 5-12.
- Lascombes, Pierre, Le Gales, Patrick (dir.), *Gouverner par les instruments*, Paris, Presses de Sciences Po 2005.
- Ruppert, Evelyn, « Doing the Transparent State: open government data as performance indicators », in Mugler, Johanna, Park, Sung-Joon (dir.), *A World of Indicators: The production of knowledge and justice in an interconnected world*, Cambridge, Cambridge University Press, 2012, p. 51-78.
- Thévenot, Laurent, « Les investissements de forme », in Thévenot, Laurent (dir.), *Conventions Économiques*, Paris, Presses Universitaires de France, 1986, p. 21-71

