



HAL
open science

Les facettes de l'Open Data : émergence, fondements et travail en coulisses

Jérôme Denis, Samuel Goëta

► To cite this version:

Jérôme Denis, Samuel Goëta. Les facettes de l'Open Data : émergence, fondements et travail en coulisses. Pierre-Michel Menger; Simon Paye. Big data et traçabilité numérique. Les sciences sociales face à la quantification massive des individus, , pp.121-138, 2017. hal-01622273

HAL Id: hal-01622273

<https://minesparis-psl.hal.science/hal-01622273>

Submitted on 24 Oct 2017

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.

Les facettes de l'Open Data : émergence, fondements et travail en coulisses

Jérôme Denis

Centre de sociologie de l'innovation
i3 CNRS - Mines ParisTech, PSL Research University

Samuel Goëta

Dataactivist

2017. In : P. Menger & S. Paye (Eds.), *Big data et traçabilité numérique. Les sciences sociales face à la quantification massive des individus*. Paris: Collège de France, p. 121-138.

Résumé

Dans ce chapitre, nous revenons sur l'émergence des politiques d'*open data* en mettant en lumière les grands principes (de la transparence jusqu'à la modernisation de l'administration, en passant par la libre circulation de l'information) que différentes initiatives ont progressivement stabilisés pour faire de l'ouverture des données publiques un enjeu international. Nous montrons ensuite, à partir d'une enquête ethnographique menée dans plusieurs institutions françaises ce que cette ouverture implique concrètement : un travail délicat qui demeure largement invisible et représente le coût caché des principes fondateurs de l'*open data*.

Mots Clés

open data, données ouvertes, transparence, travail invisible, administration

Les efforts se sont concentrés sur le travail politique et technique de mise en place de projets d'*open data*, mais pas assez sur l'analyse de ces mouvements discursifs et matériels et sur leurs conséquences. Il en résulte que nous manquons de travaux étudiant de projets d'*open data* en actes, de description des assemblages qui les entourent et leur donnent forme, et de manières désordonnées, contingentes, et relationnelles qu'ils ont de se déployer¹. (Kitchin, 2014, p. 66)

Open data et *big data* sont deux termes qui sont entrés dans le vocabulaire des débats sur le numérique entre 2008 et 2009. Le site *data.gov* est lancé en mai 2009 par le président Barack Obama un an après que le magazine *Wired* se soit enthousiasmé, dans son édition de mai 2008, pour les promesses de l'exploitation des données massives à « l'âge du pétaoctet ». Bien qu'*open data* et *big data* soient apparus quasi simultanément et que le discours commun les confonde souvent, les deux termes désignent des pratiques différentes de diffusion et d'exploitation des données.

L'ouverture des données publiques (Open Data) caractérise les pratiques proactives de publication de données produites dans le cadre d'une mission de service public et ne contenant pas d'informations personnelles. Aujourd'hui, elle est devenue une priorité des gouvernements au point que les chefs d'État réunis lors du G8 de 2013 en Irlande du Nord ont adopté une charte sur l'Open Data qui déclare que les données publiques de leurs administrations doivent désormais être librement réutilisables par défaut. Si jusqu'ici les administrations devaient légalement partager ces informations lorsque le public les demandait, les politiques d'Open Data incitent désormais les agents à publier les données d'eux-mêmes. Gestion de la dépense publique, contrôle de l'environnement, efficacité des transports, transparence des industries extractives, régulation du marché des taxis : dans des situations très variées, l'Open Data est présenté comme une solution à une multitude de problèmes.

Les données publiées par les gouvernements se prêtent rarement au qualificatif de *big data* ou « données volumineuses », « données massives » ; dans la plupart des cas, ce sont ce que Rob Kitchin appelle au contraire des *small data* (Kitchin, 2014). D'une part, du fait de leur volume : le plus gros fichier publié sur le portail *open data* français, *data.gouv.fr*, tient sur un disque DVD et la majorité des données n'y dépassent pas le mégaoctet. D'autre part, ces données font l'objet de mises à jour périodiques, annuelles ou mensuelles, exceptionnellement quotidiennes ou en temps réel. À l'inverse, les bases de données massives sont actualisées en permanence. Autre particularité, les données du Big Data ne sont quasiment jamais exploitables gratuitement et librement en intégralité alors que les données ouvertes (*open data*) peuvent être intégralement utilisées par tous. On peut toutefois citer les données ouvertes et massives des instituts météorologiques des États-Unis et de Finlande, mais les cas similaires sont exceptionnels. Malgré ces différences majeures, les données ouvertes sont une des sources des services du Big Data qui peuvent les croiser avec d'autres bases de données.

L'Open Data se distingue aussi du Big Data par les modalités de son émergence. Cette forme d'action publique répond à des aspirations variées : plus grande transparence de

¹ Les traductions, sauf mention contraire, sont des auteurs de ce chapitre.

l'État, croissance économique de l'industrie de la donnée, transformation des administrations, participation du public ou encore liberté de l'information. Cet impératif de partage et de mise à disposition au public n'est par ailleurs pas neutre, il engendre une évolution des pratiques de production et de diffusion de données au sein même des administrations. Leur publication en vue d'une réutilisation par des tiers n'est jamais purement mécanique. Elle repose sur une série de transformations qui reconfigurent en partie le travail des agents administratifs.

Dans ce chapitre, nous proposons de revenir sur les circonstances de l'émergence des mouvements, des pratiques et des politiques aujourd'hui rassemblés sous le terme d'*open data*. Dans un premier temps, nous rappellerons le contexte historique des premières initiatives d'ouverture des données publiques. Cette approche nous permettra d'identifier les promesses et les aspirations qui ont été associées à cette notion aux multiples facettes. Nous nous appuyons ensuite sur une ethnographie de l'ouverture de données publiques dans des administrations locales et nationales en France² afin de comprendre le travail effectué dans les « coulisses » de l'Open Data. Nous décrivons les principales opérations qui permettent concrètement l'ouverture des données publiques.

Du droit d'accès au droit de réutilisation des données publiques

Il convient d'abord de revenir sur les origines de ce qui est progressivement devenu une véritable injonction à ouvrir les données publiques. Le mouvement que l'on désigne aujourd'hui sous le terme d'Open Data est ancré dans des concepts, des pratiques et des contraintes réglementaires qui anticipent largement son avènement. Déjà, en 1789, la Déclaration des droits de l'homme et du citoyen stipule dans son article 15 que « la Société a le droit de demander compte à tout Agent public de son administration » faisant de l'accès à l'information publique un des fondements de la démocratie naissante (Vismann, 2008 ; Kafka, 2012). Une racine plus proche des politiques d'Open Data réside dans la notion d'Open Government apparue à la suite de la Seconde Guerre mondiale aux États-Unis (Yu & Robinson, 2012). Portée en grande partie par des journalistes, l'« ouverture des gouvernements » désigne ici la révélation par l'État des secrets de son fonctionnement. L'émergence de ce terme, devenu rapidement un slogan, doit être comprise au regard de la critique de l'opacité de l'armée américaine lors de la Guerre froide et particulièrement pendant la Guerre du Vietnam (Yu & Robinson, 2012). Autour de la notion de « gouvernement ouvert », les journalistes tentaient alors d'obtenir un droit d'accès à l'information publique, un « droit de savoir », qui est devenu progressivement la norme dans la plupart des démocraties

² Cette ethnographie rassemble plusieurs types de matériaux, 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 analysée.

occidentales. En 1966, le *Freedom of Information Act* signé par le président Lyndon B. Johnson donne à chaque citoyen américain le droit d'exiger de l'administration les informations publiques qu'elle détient. En France, la loi CADA du nom de la Commission d'accès aux documents administratifs, établit en 1978 un droit similaire qui permet à chaque citoyen d'exiger d'une administration la publication des informations publiques. L'exercice de ce droit s'effectue par le dépôt d'une requête officielle auprès de correspondants en charge de leur traitement, la commission pouvant arbitrer en cas de litige. Il est important de préciser que les données concernées par ces dispositions sont des informations publiques, produites dans le cadre d'une mission de service public ; les données personnelles ou individuelles ne sont pas concernées.

Le terme *open data* a émergé dans un autre domaine. Il est apparu en 1995 dans un rapport de l'Académie des sciences américaine suggérant le partage libre des données environnementales collectées par satellite³. Dans des sciences comme la botanique, la génétique ou l'astronomie, les pratiques de partage de données sont devenues monnaie courante, associées au développement de grandes infrastructures assurant la diffusion libre des données dans les communautés scientifiques (Bowker *et al.*, 2010 ; Edwards *et al.*, 2011 ; Strasser, 2012). Cette politique vise à compenser le coût élevé de la collecte des données et à faciliter leur contrôle par les pairs.

Mais la diffusion du terme d'*open data* dans le vocabulaire des débats autour des politiques numériques ne s'installe qu'à l'issue de la rencontre de Sebastopol, tenue en 2007 en Californie, rassemblant des activistes du numérique comme Aaron Swartz, le fondateur de Reddit et militant de l'accès libre aux publications scientifiques, Lawrence Lessig, l'avocat créateur des licences Creative Commons, ou encore Tim O'Reilly, à l'origine de la notion de « Web 2.0 ». Cette rencontre vise à définir des principes essentiellement techniques de la diffusion de données, dans le but de faciliter leur réutilisation par des machines. Ces principes exigent la publication volontaire des données dès leur production, telles que collectées et sans modification ultérieure. Tandis que la notion d'Open Government portait sur la publication de données sensibles concernant des secrets de l'État, la rencontre de Sebastopol impose des principes techniques au processus de diffusion sans en interroger le « contenu ».

Cette rencontre a exercé une grande influence sur l'équipe en charge du numérique du président Obama. Le jour de son entrée à la Maison-Blanche, il signe un mémorandum sur l'*open government* qui exige que chaque administration publie un plan d'action sur la transparence, la participation et la collaboration avec la société civile. Cette initiative a abouti à la publication d'une série de sites qui permettent aux citoyens de participer et d'accéder à l'information publique, le plus connu étant data.gov, premier portail de diffusion de données gouvernementales ouvertes. Il s'ensuit, par mimétisme, une prolifération de portails qui diffusent des données publiques, notamment au Royaume-Uni et en France où sont inaugurés data.gov.uk en 2009 et data.gouv.fr en 2011. Ces portails, qui regroupent des jeux de données extrêmement variés, s'inspirent des principes édictés à Sebastopol, dont ils stabilisent la définition par la pratique.

³ National Academy of Sciences (1995), « On the full and open exchange of scientific data », <http://www.nap.edu/readingroom.php?book=exch&page=summary.html> (dernière consultation le 22 mars 2017).

Les critères de définition d'une donnée ouverte, établis à Sebastopol et consolidés par des organismes comme l'Open Knowledge Foundation (OKFN) ou la Sunlight Foundation, peuvent aujourd'hui se résumer à quatre points principaux. Premièrement, une donnée est réputée ouverte si « chacun peut l'utiliser, la réutiliser et la redistribuer aux seules conditions de citer la source et/ou de partager à l'identique » selon la définition de l'OKFN⁴. Cette dernière condition impose que la donnée reste ouverte dès lors qu'elle est réutilisée publiquement, pour préserver son caractère de bien commun. Deuxièmement, d'un point de vue technique, une donnée ouverte doit être lisible par les machines (*machine-readable*) et exploitable pour un traitement automatisé. Cette condition invite à publier les données dans des formats ouverts, dont les spécifications n'appartiennent pas à un acteur, pour éviter une dépendance des utilisateurs à l'égard d'un logiciel particulier. Elle impose également de ne pas publier de fichiers au format PDF pour éviter une extraction complexe des données. Troisièmement, à la différence des informations diffusées dans le cadre du « droit de savoir », les données sont publiées volontairement par les administrations sur des portails qui facilitent leur utilisation par les usagers. Enfin, les données doivent être « brutes », c'est-à-dire diffusées sans retraitement, idéalement avec le plus haut niveau de détail, afin de réduire les asymétries d'information entre l'administration et le public.

Cette brève description de l'émergence de la notion d'*open data* relativise l'impression de nouveauté couramment associée à ce mouvement. Les politiques d'ouverture de données s'inscrivent dans la continuité de pratiques et de dispositions anciennes qui imposent la diffusion de l'information publique, complétées par des principes techniques qui favorisent le traitement automatisé des données. L'émergence de ces principes a été rendue possible par la convergence de plusieurs mouvements autour d'un concept porteur de multiples espoirs.

Les multiples facettes de la demande d'ouverture des données

Les injonctions à l'ouverture des données se sont réclamées d'une multitude de mouvements et de principes. Concernant la demande d'ouverture des données publiques, cinq facettes peuvent être identifiées. Leur mise en évidence permet de dépasser l'apparente uniformité des acteurs qui se sont réunis autour du terme *open data*.

La transparence

Les politiques d'ouverture de données publiques sont couramment associées à la notion de transparence. Dans la continuité d'une série de mesures qui imposent à l'État de diffuser des informations concernant son fonctionnement, les politiques d'Open Data s'appuient sur l'idée que les institutions pourraient être transparentes et maîtrisables de l'extérieur. Dans la lignée de l'article 15 de la Déclaration des droits de l'homme et du citoyen, les constitutions démocratiques ont progressivement imposé aux gouvernements de rendre des comptes sur leur fonctionnement. Le droit de chaque

⁴ Open Knowledge Foundation, « The Open Definition », accessible en ligne : <http://opendefinition.org/> (dernière consultation le 22 mars 2017).

citoyen à exiger des informations sur la conduite des affaires publiques, devenu la norme dans la plupart des démocraties, a suscité l'émergence du concept d'*accountability* (redevabilité) (Kafka, 2012). Depuis les années 1980, l'*accountability* désigne aussi la prolifération de dispositifs et de pratiques qui exigent que des comptes soient rendus et qui sont désormais tellement répandus que certains considèrent qu'ils constituent une véritable « culture de l'audit » (Porter, 1996 ; Strathern, 2000).

L'injonction à l'ouverture des données publiques s'inscrit directement dans la veine de ces dispositifs et dans le cadre de la mise en place d'objectifs et d'indicateurs généralisés. La publication de données ouvertes est ainsi souvent considérée comme une forme supérieure de transparence, réputée plus objective que les procédures de révélation qui l'ont précédée (Birchall, 2014). Dispensées d'un biais narratif ou interprétatif, les données « brutes » sont associées à une objectivité qui configure les politiques d'Open Data comme incarnant le renouveau voire le paroxysme de la transparence de l'État (Goëta, 2015). Par ailleurs, Evelyn Ruppert (2012) explique la dimension performative de ces politiques d'ouverture. L'utilisation des données ouvertes et leur circulation à travers des médiations sociotechniques met en œuvre (*enact*) la transparence publique, en faisant notamment émerger de véritables « publics des données » (*data publics*) qui sont institués par les dispositifs d'*open data* en inspecteurs du fonctionnement de l'État.

La libre circulation de l'information

La demande de données ouvertes s'inscrit également dans un mouvement plus large qui considère que la diffusion de l'information est impérative pour le maintien des sociétés démocratiques. L'influence de la pensée cybernétique est ici déterminante. En postulant que l'information est le « ciment de la société » indispensable à la perpétuation de la civilisation, la cybernétique défend en effet le modèle des sociétés « ouvertes » qui font reculer localement ce que ces théoriciens appellent l'« entropie », la menace d'un chaos inéluctable qu'annoncent les principes physiques de la thermodynamique (Breton, 2004 ; Lafontaine, 2004 ; Triclot, 2008). L'ouverture des données publiques intervient par ailleurs à un moment où les données sont conçues comme une matière première (Ribes & Jackson, 2013) indispensable à la création de valeur ajoutée. Le schéma de la pyramide du savoir (Ackoff, 1989), admis couramment malgré son réductionnisme, place ainsi les données comme le fondement de toute forme d'information (Rowley, 2007 ; Frické, 2009 ; Floridi, 2011). Alors que les données sont désormais indispensables pour analyser un phénomène, leur diffusion est vue comme un vecteur de démocratisation de l'information et de l'expertise.

Les mouvements liés au logiciel libre ont aussi largement milité pour la liberté de circulation de l'information. Sébastien Broca, dans *Utopie du logiciel libre* (2013), revient sur les fondements intellectuels de ces mouvements et soutient que « la libre circulation de l'information est la valeur la plus communément associée au combat du Libre, que ce soit par les universitaires ayant étudié le sujet ou par les hackers eux-mêmes. » Des partisans du logiciel libre se sont en effet engagés pour que la connaissance et les œuvres bénéficient des mêmes libertés d'utilisation, de modification, de copie et de redistribution que le code informatique des logiciels qu'ils utilisent. Dans la continuité de luttes contre les entraves à la circulation de l'information, des « libristes » ont participé à la définition des principes de l'Open Data en défendant notamment l'usage

de formats aux spécifications ouvertes et de licences dites *copyleft* qui empêchent légalement ce que la littérature autour des biens communs appelle l'« enclosure » des données. En France, « Regards citoyens », la principale association en faveur de l'ouverture des données publiques, utilise exclusivement des logiciels libres et comprend parmi ses membres fondateurs des militants de l'April, une association de défense du libre. Dans la revendication de la circulation de l'information, l'ouverture des données publiques apporte aussi une nouvelle exigence : celle de pouvoir réutiliser des données brutes et non des agrégats déjà traités par leurs producteurs.

La demande de données brutes

Les politiques d'ouverture des données publiques sont liées à l'entrée dans le langage courant de la notion de « données brutes ». Celle-ci est issue des sciences expérimentales et désigne les mesures telles qu'elles sont produites par les instruments. L'adjectif *brut* qualifie des données qui n'ont pas fait l'objet d'un traitement humain ou informatisé pour corriger ou « nettoyer » ces mesures (Walford, 2013). Dans *Science and Technology Studies*, Bowker (2008) et Gitelman (2013) ont dénoncé « un oxymore et une mauvaise idée » qui sous-entend que la production de l'information serait transparente et naturelle. Car, de fait, l'étude de la production de données scientifiques rend compte de l'impossibilité d'identifier un moment où la donnée aurait connu un état brut, sans traitement (Ribes et Jackson, 2013). Les « données » sont toujours « obtenues » (Latour, 1993, p. 188), elles sont le fruit de choix, de théories et d'interventions et ne sont jamais pré-interprétatives comme le laisse penser l'expression « données brutes ».

La revendication d'accéder aux données brutes va aussi de pair avec l'émergence des *data-driven sciences*^{*}, comme l'astronomie, la botanique ou la génétique, où le partage et la réutilisation de grandes bases de données sont devenus des pratiques courantes (Strasser, 2012). Dans un contexte où les résultats scientifiques sont régulièrement contestés, publier les données avant leur traitement et leur agrégation devient une pratique courante pour assurer la transparence des procédures des chercheurs. En climatologie, où des controverses ont mis en doute la réalité du changement climatique, les chercheurs ont dû répondre aux demandes des climatosceptiques qui ont exercé leur droit légal d'accès à l'information (*freedom of information*) pour obtenir les données brutes (Edwards, 2010).

Concernant les données publiques des administrations, la revendication d'accès aux données brutes émerge en 2007 suite à un article de Rufus Pollock⁵, fondateur du réseau international Open Knowledge qui milite pour l'ouverture du savoir. Dans un billet de blog intitulé « Give us the data raw and give us the data now », il critique les sites web publics qui se concentrent sur la production d'une belle interface pour présenter les données. Il explique que les interfaces deviennent rapidement obsolètes ; selon lui, les administrations devraient d'abord publier les données brutes pour que le public ou l'institution créent les interfaces qui vont répondre aux besoins des usagers. Cette exigence s'est répandue véritablement en 2009 après une intervention de Tim

⁵ Pollock R. (2007), « Give us the data raw, and give it to us now », *The Open Knowledge Foundation Blog*, <http://blog.okfn.org/2007/11/07/give-us-the-data-raw-and-give-it-to-us-now/> (consulté pour la dernière fois le 27 février 2017).

Berners-Lee, le fondateur du Web. Lors d'une conférence TED dont la vidéo dépasse le million de vues⁶, il demande au public présent de crier « nous voulons des données brutes ! » Il y évoque l'existence de données « inaltérées », qu'il suffirait de réclamer pour qu'elles soient ouvertes à tous. Cette conférence a eu une grande influence sur l'émergence de l'injonction politique d'ouverture des données, envisagée comme un processus simple qui n'a pas à répondre à des critères de qualité, les données étant disponibles à l'état brut.

L'industrie de la donnée

Le quatrième aspect qui caractérise l'émergence des politiques d'*open data* a trait à l'avènement d'une industrie de l'information qui a considéré l'accès gratuit aux données publiques comme un « avantage compétitif » dans son développement (Kitchin, 2014). Les promesses économiques de l'ouverture des données sont liées à l'émergence d'une industrie de l'information. Cette industrie est particulièrement développée aux États-Unis où aucune information publique fédérale ne peut être protégée par le droit d'auteur depuis 1988 (Ronai, 1997). Dans les années 2000, la Commission européenne s'est inspirée du cadre réglementaire états-unien et a évalué la valeur économique de la diffusion gratuite ou à un coût marginal des données publiques à plusieurs dizaines de milliards d'euros par an dans l'Union. En 2003, elle a adopté la directive *Public Sector Information* (PSI) pour inciter les États membres à diffuser gratuitement leurs données. La Commission européenne multiplie les études sur le potentiel économique de la libération des données, évaluant jusqu'à 200 milliards d'euros par an la valeur de leur circulation optimale dans les pays de l'Union (Vickery, 2011). Ces promesses de croissance justifient les efforts financiers et humains (suppression des redevances, développement de portails, organisation d'événements) consentis par l'État pour ouvrir ses données. L'État planifie un retour sur investissement dans la taxation des revenus de la réutilisation dont il est prévu que le montant dépasserait largement le coût de l'ouverture.

Ces prévisions de croissance se sont transformées en revendication à la suite d'une campagne du quotidien britannique *The Guardian* en 2006 intitulée « Rendez-nous les bijoux de la couronne » qui exigeait la gratuité des données « payées par les impôts ». Parmi ces « bijoux », les données de l'Institut géographique britannique *Ordnance Survey* ont été perçues comme les plus prometteuses en termes de croissance. La tribune insistait sur le fait qu'aucun des services de cartographie numérique comme Google Maps n'était britannique du fait des fortes redevances exigées pour utiliser les données de l'*Ordnance Survey*. Le *Guardian* évoquait alors la différence avec la conception américaine :

Aux États-Unis, on considère que les données collectées avec l'argent du contribuable devraient lui être restituées gratuitement. Et une étude détaillée montre que l'attitude fermée du Royaume-Uni envers ses données signifie que nous perdons des opportunités commerciales et même que nous freinons la

⁶ Berners-Lee T. (2009), « The next Web », *TED*, http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html (consulté pour la dernière fois le 27 février 2017).

recherche scientifique dans des domaines comme le changement climatique⁷.

Cet extrait montre bien que la promotion d'une ouverture des données publiques au Royaume-Uni s'est formée autour de conceptions assez éloignées des autres facettes : nulle part il n'est fait référence aux notions de transparence ou de circulation de l'information.

La modernisation des administrations

Au-delà des promesses de croissance, les politiques d'*open data* reposent aussi sur leur capacité à provoquer des changements dans les procédures utilisées par l'administration et ainsi à accélérer sa « modernisation ». Les projets d'ouverture des données publiques sont soutenus par les services en charge de la transformation du service public. Ces derniers prévoient que la diffusion des données permettra aux agents de bénéficier des informations des autres services. La réutilisation des données produites dans d'autres divisions de l'organisation rendrait possible la fin du « travail en silos », expression désignant le cloisonnement des services et l'insuffisance des pratiques de partage d'informations. L'ouverture des données publiques offrirait aussi la possibilité de créer des services innovants en utilisant des données qui n'étaient jusqu'alors pas connues. Les administrations en charge de l'efficacité du service public anticipent aussi que les producteurs vont devenir redevables auprès de leurs collègues de la qualité des données.

Cette modernisation concerne aussi le public lui-même. L'ouverture des données encouragerait la participation des citoyens à l'élaboration des politiques par la réduction des asymétries d'information. En disposant des données, idéalement dans leur intégralité et collectées selon les principes de Sebastopol, les politiques d'*open data* faciliteraient l'apparition d'une citoyenneté informée au fondement de politiques publiques d'inspiration libérale (Barry, 2001). La disponibilité des données publiques permettrait l'émergence de contre-pouvoirs qui pourraient avoir recours à celles-ci dans leurs activités militantes. D'autres modalités d'action collective ont vu le jour récemment, comme le « statactivisme » de certains acteurs qui se mobilisent pour que les données, les catégories et leurs indicateurs prennent en compte des situations jusqu'alors occultées ou ignorées (Bruno, Didier & Prévieux, 2014). Les services gouvernementaux en charge de la transformation de l'administration soutiennent les politiques d'*open data* pour favoriser une meilleure prise en compte des avis des citoyens. Plus largement, l'ouverture de l'information stratégique au public est perçue comme une source d'efficacité et d'innovation des organisations. Les politiques d'*open data* s'inscrivent dans la continuité des travaux des théoriciens des organisations qui encouragent une diffusion de l'information. En particulier, les travaux d'Henry Chesbrough (2006), à l'origine de la notion d'innovation ouverte, ont répandu l'idée qu'ouvrir l'information stratégique peut constituer une source d'innovation bien supérieure au régime classique du confinement des ressources en interne.

⁷ Charles et Cross (2006), « Give us back our crown jewels », *The Guardian*, 9 mars 2006, <https://www.theguardian.com/technology/2006/mar/09/education.epublic> (dernière consultation : 3 avril 2017).

Ces bénéfices internes et externes positionnent les politiques d'*open data* comme moteur de la transformation du service public. La Banque mondiale encourage ainsi financièrement les pays en développement à ouvrir leurs données à la fois pour faire émerger une société civile informée, mais aussi pour « moderniser » le fonctionnement des administrations. Cette « modernisation » s'inscrit aussi dans le contexte de réduction des dépenses de l'État, où l'ouverture des données financières permettrait un contrôle de l'action publique qui dépasse les instances dédiées et l'étende à la société civile dans son ensemble. Des critiques ont associé l'ouverture des données publiques à la privatisation du service public (Bates, 2012). Une étude au Royaume-Uni a montré que les élus locaux ont réduit la question de l'ouverture des données à l'obligation de publication du détail des dépenses publiques de plus de 500 £ imposée par le gouvernement Cameron, oubliant les autres bénéfices attendus de la politique *open data* (Halonen, 2012).

Mais la progressive émergence d'une injonction à ouvrir des données dans les administrations publiques et la mise en œuvre de politiques d'*open data* dans de nombreux pays laissent paradoxalement dans l'ombre la question des données elles-mêmes et celle des conditions de leur « ouverture ». De quelles données parle-t-on exactement lorsqu'on défend la nécessité de libérer les données publiques ? Quelles sont ces ressources que certains dépeignent comme le pétrole du XXI^e siècle, qu'il suffirait de mettre à la disposition de la population dans leur état « brut » pour qu'elle y trouve des moyens de produire de la transparence et de l'innovation ? Plus qu'un angle mort, les données constituent une évidence des politiques de l'*open data*. Insister pour que soit mis en œuvre un programme d'ouverture de données suppose que des données soient déjà présentes dans les administrations publiques, prêtes à être « libérées ». Or, dès que les programmes prennent chair dans les administrations, cette évidence ne tient plus. L'ouverture des données n'a rien de mécanique et la mise en œuvre concrète d'une politique d'*open data* passe par une série d'opérations qui témoignent de l'épaisseur sociotechnique des processus d'ouverture.

Le travail d'ouverture des données

Sans prétendre à l'exhaustivité, nous proposons d'exposer ici cinq étapes principales qui ponctuent cette ouverture progressive. Ces opérations montrent que les données, loin d'être des ressources toutes prêtes dont la diffusion pourrait s'effectuer à coût quasi nul, sont travaillées dans le processus même de leur ouverture. Elles sont d'autant plus délicates qu'elles sont portées par des collectifs de travail hétérogènes. Dans la plupart des cas étudiés, les chargés de projet sont rattachés à des services de communication et étaient au départ peu sensibilisés à la gestion des données administratives. Par ailleurs, au sein des services internes, ces politiques ajoutaient de nouvelles tâches et de nouvelles missions à des personnes qui ne travaillaient pas à la diffusion publique des données qu'elles produisaient ou qu'elles manipulaient. Les principaux acteurs de ces projets ne sont donc pas initialement des « petites mains » de l'information (Denis & Pontille, 2012) : leur activité déborde largement les seules production et maintenance de données et leur statut au sein des organisations concernées ne les vouaient pas à rester en coulisses d'un service destiné à des usagers extérieurs. En revanche, une partie des opérations qu'ils réalisent constituent

bien un « travail invisible de l'information » (Denis, 2011), sur lequel reposent concrètement les politiques d'*open data* et leurs promesses de transparence, d'innovation, et de modernisation publique.

Identifier

Les moments qui initient les premiers pas d'un programme d'*open data* témoignent clairement du statut ambigu, et plus complexe qu'on ne l'imagine de prime abord, des données. Non seulement le périmètre des données potentiellement candidates à l'ouverture est débattu pour chaque projet concret, mais la connaissance même de leur existence, de leur nature et de leur emplacement est également problématique. Avant la question « comment allons-nous ouvrir tel ou tel jeu de données ? », de nombreux acteurs se confrontent à une interrogation plus générique : « de quelles données disposons-nous ? »

Les premiers pas d'un programme d'*open data* prennent ainsi la forme d'une exploration, une enquête menée par les personnes qui en ont la charge. Le terme d'*exploration* n'est pas anodin : cette enquête n'est pas un recensement. L'identification des données s'effectue progressivement, au fil d'échanges avec les services internes. Et le processus se nourrit lui-même, faisant émerger de nouvelles pistes au fil de son avancement. Selon les situations que nous avons étudiées, la sollicitation de ces services prend des formes variées : appel à initiatives internes, contact direct, premières idées de données pertinentes dessinées par l'équipe *open data* puis proposées aux services concernés, voire propositions faites par des associations de citoyens ou des développeurs qui ont eu l'occasion de les formuler.

L'inventaire, loin d'être mécanique, est un travail collectif : il engage des réunions, voire des négociations, qui articulent des enjeux hétérogènes, où se mêlent la concurrence entre les collectivités, l'intérêt imaginé du public pour tel ou tel type d'informations, la facilité ou au contraire les difficultés techniques liées à la diffusion des données ou à leur caractère plus ou moins « sensible ». Les données ne sont donc pas sélectionnées sur la base de critères simples, définis à l'avance : leur exploration est progressive et collective. Elle passe par une co-construction qui montre que l'identification même des données qui vont être ouvertes participe pleinement du processus d'ouverture.

On descend, on descend jusqu'au plus petit dénominateur commun pour qu'on puisse identifier vraiment toutes les données. Et ce qui est fou, c'est qu'à partir de ces trente rendez-vous, à chaque fois que je les rencontre, ils m'identifient cinq autres personnes qu'il faudrait que je vois donc, en gros, c'est un peu exponentiel. (Chef de projet open data d'une collectivité locale)

L'identification initie donc une opération générative (Law, 2009), un processus d'instauration, qui engendre un périmètre de données qui sont désignées non seulement comme « ouvertes » (ouvrables, dans un premier temps), mais aussi comme « données » tout court.

Cette première opération a également des conséquences organisationnelles. L'identification concerne autant des données que des services qui vont en devenir les référents. La mise en place de l'Open Data « travaille l'organisation », pour reprendre les termes de Cochoy, Garrel et de Terssac (1998), redistribuant certaines cartes, attribuant des rôles nouveaux et des responsabilités inédites.

Extraire

L'identification n'est qu'une première étape qui, dans la plupart des cas, ne suffit pas à « mettre la main » sur les données, notamment parce qu'une grande partie d'entre elles est littéralement enfouie dans des bases de données.

Le propre des bases de données relationnelles est d'offrir à leurs usagers des accès spécifiques, des « vues », répondant à leurs préoccupations professionnelles (Dagiral et Peerbaye, 2013). Ces « vues utilisateurs » évitent à tout un chacun de se préoccuper de la manière dont les données sont effectivement organisées dans les disques durs des machines (Castelle, 2013). Mais ce qui est pensé comme une source de confort et de facilité d'usages du point de vue des concepteurs des bases de données apparaît comme une contrainte forte pour les responsables d'un programme d'*open data*. Ouvrir les données suppose de passer outre les « vues utilisateurs » de chaque métier afin d'entrer en contact direct avec les données. Mais rares sont les bases qui disposent de fonctionnalités d'extraction automatique qui permettraient de récolter les données indépendamment de leur mise en forme logicielle :

Il faut bien comprendre, c'est qu'au départ pour la plupart des systèmes des applications qu'on a chez nous qu'on a acheté, elles ne sont pas du tout conçues pour faire de l'open data. Donc, c'est compliqué. On est obligé, nous, de développer des moulinettes, des tas de choses pour pouvoir sortir des données proprement. (Gestionnaire de bases de données de transport)

L'extraction des données repose ainsi sur une autre forme d'exploration, qui passe par le développement d'outils *ad hoc* (des « moulinettes ») pour atteindre la « vue physique » des bases de données, outils d'autant plus complexes à élaborer qu'ils concernent des parcs informatiques hétérogènes, au sein desquels les logiciels ne sont que rarement harmonisés et où cohabitent des générations différentes d'outils.

C'est une deuxième étape de la progressive instauration des données qui se joue dans le travail d'extraction de données. Cette étape ne se limite évidemment pas à des aspects « purement » informatiques, mais elle dépend aussi des relations que les administrations entretiennent avec leurs prestataires techniques. Selon les termes des contrats, et la plus ou moins bonne volonté de ces prestataires, les vues physiques (les chemins d'accès vers les données « elles-mêmes ») sont plus ou moins masquées, et les bricolages pour y parvenir plus ou moins assimilables à des détournements, voire à des ruptures contractuelles. L'idée que les données publiques seraient des ressources dormantes qui ne demanderaient qu'à être libérées pour être exploitées, des « *commodities* » (Ribes *et al.*, 2013), est donc mise à mal non seulement par le coût que représente le travail d'extraction, mais aussi par l'ambiguïté de l'agencement

sociotechnique dont les données sont dépendantes. Certains prestataires techniques qui vendent et entretiennent les systèmes d'information des institutions publiques sont propriétaires des chemins d'accès et des dispositifs de stockage de leurs bases de données : l'inaccessibilité des données est au cœur de leur modèle économique.

Nettoyer

Une fois les données identifiées et effectivement accessibles, elles passent par une étape cruciale : leur nettoyage. En sciences, ce type d'opérations est bien connu. Nombreux sont les travaux en *science and technology studies* qui ont insisté sur cette opération cruciale (Edwards, 2010 ; Walford, 2013). Si certaines tâches en sont proches, le nettoyage mis en œuvre dans les programmes d'*open data* ne répond toutefois pas aux mêmes préoccupations qu'en sciences : il n'est pas question ici de corriger les parasites générés par les instruments et le contexte de la mesure (en effaçant le bruit et les artefacts), mais plutôt de « mettre en qualité » des données en se confrontant à leurs usages précédents.

Cela passe par le repérage et la correction d'erreurs, par l'élimination de lignes vides, l'absence de données pouvant créer des incompréhensions, voire gripper le mécanisme des logiciels dédiés à les traiter. Il s'agit aussi pour les personnes chargées de cette étape d'harmoniser des identifiants, c'est-à-dire de travailler à l'articulation de jeux de données :

Typiquement, sur les jeux de données des élections : entre les derniers fichiers des dernières élections, et puis les vieux trucs, les fichiers n'étaient pas présentés pareil. C'était des choses très bêtes mais il y avait des fois le titre de colonne qui était soit le nom du candidat soit le nom de son parti ou alors les deux, et j'ai essayé d'uniformiser tout ça pour que tous les fichiers se ressemblent et soient structurés pareil. (Chargé de projet *open data* dans une intercommunalité)

Quel est le sens de ces opérations de nettoyage ? Faut-il en conclure que les données avec lesquelles travaillent les administrations sont de mauvaise qualité, « sales », et que les programmes d'*open data* sont de bonnes occasions pour, enfin, les améliorer ? La question est un peu plus complexe. Comme l'ont montré Garfinkel et Bittner (1967) dans un article célèbre, la question de la qualité n'est pas une propriété intrinsèque des données, et l'on trouve de nombreuses « bonnes raisons organisationnelles » à l'existence de « mauvaises » données. C'est aussi ce que montre Lampland (2010) à propos des pratiques comptables qui reposent en partie sur la manipulation de « *false numbers* » qui n'ont pas de conséquences particulières tant qu'ils ne sont pas saisis par des dispositifs d'audit qui en transforment le sens et la portée, les rendant problématiques. Les « mauvaises » données des uns peuvent ainsi être les « bonnes » données des autres. C'est le cas par exemple dans les transports, où les horaires utilisés dans certaines bases affichaient « 25 h 10 » pour « 01 h 10 » dans la nuit. De telles valeurs qui permettent de lever les ambiguïtés dans le cadre de la coordination du travail n'ont pas de sens en dehors de ce contexte.

Le nettoyage mis en œuvre dans le cadre des politiques d'*open data* ne témoigne donc pas d'une mauvaise qualité « intrinsèque » des données des administrations, mais plutôt du fait que l'ouverture constitue une épreuve pour les données qui, dans les mains d'autres usagers, pourraient être jugées inadéquates. En faisant migrer les données dans un cadre nouveau, les programmes d'*open data* rendent potentiellement centrales certaines de leurs dimensions qui étaient peu pertinentes dans leurs cadres d'usages initiaux. Des absences jamais remarquées deviennent des manquements, des approximations sans importance deviennent des erreurs, des redondances utiles deviennent des sources de problèmes.

« Brutifier »

Au nettoyage s'ajoute une opération plus curieuse, qui est en quelque sorte son prolongement, et qui témoigne de l'importance du travail sur les données en vue de leur ouverture. Nous l'avons vu, parmi les principes sur lesquels s'appuient les politiques d'*open data*, l'idée que les données ouvertes doivent être « brutes » est centrale. De bonnes données ouvertes sont des données qui ne sont pas retouchées. La première série d'opérations que nous venons de décrire met déjà à mal ce principe, puisqu'elle montre que, au contraire, l'ouverture des données engendre un certain nombre de manipulations essentielles à leur identification, à leur accessibilité et à leur qualité. Mais nous avons appris au fil de l'enquête que les données faisaient également l'objet d'une « brutification ». Certaines opérations visent en effet à les rendre brutes, comme les personnes qui en ont la charge nous l'ont expliqué, assumant ainsi pleinement ce que Geoffrey Bowker a qualifié d'« oxymore » à propos des données brutes en sciences (Bowker, 2000 ; Gitelman, 2013).

Cette « brutification » s'apparente à une forme de nettoyage mais ne vise pas à corriger des « erreurs ». Elle s'articule autour de deux enjeux étroitement liés : l'effacement des traces d'usages dans les données, et leur « délocalisation ». L'effacement des traces d'usage passe par exemple par le masquage de commentaires, ou celui de couleurs dans certaines cellules des tableaux. La délocalisation revient à faire disparaître les marques de l'ancrage des données dans des activités professionnelles spécifiques. Il s'agit par exemple de remplacer les nombreuses abréviations dont on sait qu'elles sont au cœur d'une forme d'efficacité langagière au travail (Fraenkel, 1994), ou encore de traduire des termes techniques dans un langage ordinaire :

C'était un fichier Excel qu'ils avaient mis en forme selon ce dont ils avaient besoin. [...] Donc, c'était vraiment leur fichier de travail. Or, nous, on ne voulait pas ça. Nous, on voulait des données plus brutes c'est-à-dire pas de commentaires, pas de tableaux, pas de mise en forme, juste vraiment les données au jour le jour, statistiques. Moi, je me suis occupée de ce travail-là, rebrutifier les données en fait. (Chargée de projet *open data* dans une intercommunalité)

D'un point de vue général, les opérations de « brutification » visent donc l'intelligibilité des données. Cette mise en intelligibilité passe évidemment par l'élaboration de métadonnées essentielles à tout projet de partage de données (Baker et Bowker, 2007 ; Edwards *et al.*, 2011) : dictionnaire, commentaires dans des documents à part, sont

associés aux données pour que leurs usages soient facilités. Mais elle passe aussi par la transformation des jeux de données eux-mêmes, au sein desquels des termes vont être remplacés, des intitulés simplifiés, d'autres assemblés.

La « brutification » montre que le processus d'ouverture ne consiste pas à « libérer » des données qui resteraient primaires, mais à produire des données génériques, à vocation universelle. Cette production passe par une série de transformations par lesquelles des informations et des données « métier » deviennent des données ouvertes.

Formater

Enfin, il faut insister sur une dernière opération – le formatage – qui achève l'instauration des informations initialement identifiées en données ouvertes. Si l'ouverture des données publiques ne connaît pas le même degré de standardisation que le partage des données en sciences, il existe toutefois quelques standards et des formats de données qui sont en voie de consolidation et qui s'imposent dans les bonnes pratiques du secteur (Goëta et Davies, 2016). Parmi ceux-ci, le format CSV (*comma-separated values*) joue un rôle central aujourd'hui, notamment parce qu'il est considéré par la plupart des acteurs du domaine comme un format ouvert, compatible avec la plupart des logiciels tableurs, y compris du grand public. Nous avons ainsi pu entendre à l'occasion de la réunion de lancement du programme *open data* d'une organisation internationale le chargé de projet dire cette phrase qui souligne l'importance de la mise au format dans le processus d'instauration des données ouvertes : « pour moi du brut c'est du CSV ».

Or, la traduction d'un jeu de données en fichier CSV ne va pas de soi. Chaque exportation d'un format propriétaire issu des bases de données originales, ou d'un logiciel tableur, vers le format désiré réserve son lot de problèmes et donne lieu à des ajustements pour que les données initiales ne soient pas corrompues. Dans le meilleur des cas, des opérations en amont de l'exportation permettent de produire des fichiers qui seront compatibles et supporteront mieux le reformatage :

- Quand tu passes [un fichier Excel] en CSV tout saute. Il y a des cellules fusionnées, du gras... Dans certains fichiers les gars ont mis de la couleur et la couleur a une signification, alors, que dans le CSV il n'y pas de couleur. Donc, tu es obligé de créer d'autres colonnes. [...]
- Et tu arrives à comprendre dans tous les cas, tout ce qui est... ?
- Si on ne comprend pas on s'en réfère au producteur de données qui nous a envoyé le fichier. Mais sinon, tu vois, on reçoit aussi des fichiers avec des adresses, ce n'est pas géocodé. Du coup, on géocode. (Chargé de projet *open data* dans une région.)

Par ailleurs, le formatage ne représente pas seulement une énième manipulation technique des données dans le processus qui mène à leur ouverture. Les standards engagent en effet toujours des enjeux politiques étroitement liés à leurs aspects les plus

techniques (Lampland et Star, 2008). Une partie des opérations précédentes décrites ici sont orientées vers cette mise au format des données, elles doivent en tout cas s'articuler avec celle-ci. La question des formats qui vont s'imposer et devenir des standards de l'*open data* est donc cruciale puisqu'ils vont non seulement définir, ou redéfinir en partie, la liste des acteurs institutionnels du domaine, ancrer les politiques d'*open data* dans des principes « de fait », traduits techniquement, mais également organiser le travail de l'information qui assure en coulisse l'existence même des données ouvertes (Goëta, 2014).

Conclusion

Tout au long de ce chapitre, nous avons montré que ce que l'on désigne aujourd'hui sous le terme d'Open Data regroupe des projets d'ouverture de données qui s'inscrivent dans un mouvement relativement récent et nourri de principes variés, inspirés notamment de la cybernétique, du mouvement du logiciel libre et du droit d'accès à l'information publique. Ces projets articulent par ailleurs des principes fort différents, au premier rang desquels la transparence politique, le soutien à l'innovation, et la modernisation des administrations. Au cœur des politiques d'Open Data se trouve toutefois un élément commun, dont l'existence semble évidente aux yeux de tous : les données. Les vocabulaires de la *libération*, de l'*ouverture* et de la *mise à disposition*, associés à l'importance nouvelle prise par les notions de données « brutes » ou « primaires », attribuent une place très particulière à ces entités informationnelles dont l'existence semble aller de soi, et dont la diffusion quasi mécanique est présentée comme un moteur de progrès.

En insistant sur certains aspects pratiques de la mise en œuvre de ces projets, nous avons montré, à rebours de cette vision désincarnée, que l'ouverture des données publiques reposait sur un travail complexe, sensible, par lequel des informations hétérogènes étaient progressivement instaurées en *open data*. Au fil d'une série de transformations techniques, linguistiques et politiques, s'opère ainsi la mue de données « métier » en données « ouvertes ». Nous avons également vu qu'au-delà du coût, généralement caché, que représente un tel travail, le processus d'ouverture a des répercussions organisationnelles qui dépassent les projets de modernisation habituellement envisagés.

L'ancrage historique, l'horizon des motifs qui animent les projets politiques et les activités concrètes qui nourrissent les projets donnent donc à voir la richesse, mais aussi la fragilité, des politiques d'Open Data, dont les sciences sociales ont tout intérêt à étudier la multiplicité plutôt que les grandes lignes communes. Il nous semble qu'une telle démarche serait également bénéfique à l'analyse des projets se réclamant des *big data*, dans un contexte où les débats contemporains tendent justement à se focaliser sur le traitement des données en aval, sur leur utilisation ou sur les enjeux de leur circulation. Au nom de quoi, par qui et dans quelles conditions ces données sont-elles produites ? Quelles transformations subissent-elles pour être associées les unes aux autres et devenir « massives » ? Autant de questions qui permettraient d'éclairer sous un nouvel angle l'assemblage sociotechnique des *big data*, dont tout le monde s'accorde aujourd'hui à souligner le caractère sensible sur les plans cognitif, politique et moral.

Références

Ackoff R. (1989), « From data to wisdom », *Journal of Applied System Analysis*, vol. 15, p. 3-9.

Baker K.S. et Bowker G.C. (2007), « Information ecology: Open system environment for data, memories, and knowing », *Journal of Intelligent Information Systems*, vol. 29, n° 1, p. 127-144.

Barry A. (2001), *Political machines. Governing a Technological Society*, Londres, The Athlone Press.

Bates J. (2012), « “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*, vol. 8, n° 2.

Birchall C. (2014), « Radical transparency? », *Cultural Studies ↔ Critical Methodologies*, vol. 14 n° 1, p. 77-88.

Bowker G.C. (2000), « Biodiversity datadiversity », *Social Studies of Science*, vol. 30, n° 5, p. 643-683.

Bowker G.C. (2008), *Memory Practices in the Sciences*, Cambridge (Mass.), MIT Press.

Bowker G.C., Baker K., Millerand F. et Ribes D. (2010), « Toward information infrastructure studies: Ways of knowing in a networked environment », in Hunsinger J., Klastrop L. et Allen M. (dir.), *International Handbook of Internet Research*, Dordrecht, Springer, p. 97-117.

Breton P. (2004), *L’Utopie de la communication : le mythe du “village planétaire”*, Paris, La Découverte.

Broca S. (2013), *Utopie du logiciel libre. Du bricolage informatique à la réinvention sociale*, Neuvy-en-Champagne, Le Passager clandestin.

Bruno I., Didier E. et Prévieux J. (2014), *Statactivisme. Comment lutter avec des nombres*, Paris, Zones.

Castelle M. (2013), « Relational and non-relational models in the entextualization of bureaucracy », *Computational Culture*, n° 3, <http://computationalculture.net/article/relational-and-non-relational-models-in-the-entextualization-of-bureaucracy>.

Chesbrough H. (2006), *Open Business Models: How to thrive in the new innovation landscape*, Cambridge (Mass.), Harvard Business School Publishing.

Cochoy F., Garel J.-P. et de Terssac G. (1998), « Comment l’écrit travaille l’organisation : le cas des normes ISO 9000 », *Revue française de sociologie*, vol. XXXIX, n° 4, p. 673-699.

Dagiral É. et Peerbaye A. (2013), « Voir pour savoir. Concevoir et partager des “vues” à travers une base de données médicales », *Réseaux*, n° 178-179, p. 163-196 (en ligne : <https://www.cairn.info/revue-reseaux-2013-2-page-163.htm>).

Denis J. (2011), « Le travail de l’écrit en coulisses de la relation de service », *Activités*, vol. 8, n° 2, p. 32-52.

Denis J. et Pontille D. (2012), « Travailleurs de l’écrit, matières de l’information », *Revue d’anthropologie des connaissances*, vol. 6, n° 1, p. 1-20.

Edwards P. (2010), *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*, Cambridge (Mass.), MIT Press.

Edwards P., Mayernik M. S., Batcheller A.L., Bowker G.C. et Borgman C.L. (2011), « Science friction: Data, metadata, and collaboration », *Social Studies of Science*, vol. 41, n° 5, p. 667-690.

Floridi L. (2011), *The Philosophy of Information*, Oxford University Press.

Fraenkel B. (1994), « Le style abrégé des écrits de travail », *Cahiers du français contemporain*, n° 1, p. 177-194.

Frické M. (2009), « The knowledge pyramid: A critique of the DIKW », *Journal of Information Science*, vol. 35, n° 2, p. 131-142.

Garfinkel H. et Bittner E. (1967), « "Good" organizational reasons for "bad" clinic records » in Garfinkel H. (dir.), *Studies in ethnomethodology*, Englewood-cliffs, Prentice-Hall, p. 186-207.

Gitelman L. (dir.) (2013), « *Raw Data* » is an Oxymoron, Cambridge, Cambridge (Mass.), MIT Press.

Goëta S. (2015), « L'Open Data : une forme ultime de transparence ? », in Catellani A. et Libaert T., *La communication transparente. L'impératif de transparence dans le discours des organisations*, Louvain-la-Neuve, Presses universitaires de Louvain, p. 49-66.

Goëta, S. et Davies, T. (2016), « The Daily Shaping of State Transparency: Standards, Machine-Readability and the Configuration of Open Government Data Policies », *Science & Technology Studies*, vol. 29, n° 4, p. 10-30.

Halonen A. (2012), *Being open about data. Analysis of the UK open data policies and applicability of open data*, The Finnish Institute in London, rapport disponible en ligne: <http://finnishinstitute.cdn.coucouapp.com/en/articles/48-reports> (dernière consultation le 27 mars 2017).

Kafka B. (2012), *The Demon of Writing. Powers and Failures of Paperwork*, Brooklyn, Zone Books.

Kitchin R. (2014), *The Data Revolution. Big Data, Open Data, Data Infrastructures & their Consequences*, Londres, Sage.

Lafontaine C. (2004), *L'empire cybernétique. Des machines à penser à la pensée machine*, Paris, Seuil.

Lampland M. (2010), « False numbers as formalizing practices », *Social Studies of Science*, vol. 40, n° 3, p. 377-404.

Latour B. (1993), « Le pédofil de boavista, montage photo-philosophique », *Petites leçons de sociologie des sciences*, Paris, La Découverte, p. 171-225.

Latour B. (1993), *La Science en action*, Paris, La Découverte.

Lampland M. et Star S.L. (dir.) (2008), *Standards and their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Cornell, Cornell University Press.

Law J. (2009), « Seeing like a survey », *Cultural Sociology*, vol. 3, n° 2, p. 239-256.

Porter T. (1996), *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*, Princeton, Princeton University Press.

Ribes D. et Jackson S.J. (2013), « Data bite man: The work of sustaining a long-term study », in Gitelman L. (dir.), « *Raw Data* » is an Oxymoron, Cambridge (Mass.), MIT Press, p. 147-166.

Ronai M. (1997), « Données publiques : accès, diffusion, commercialisation », *Problèmes politiques et sociaux*, n° 68, p. 773-774.

Rowley J. (2007), « The wisdom hierarchy: Representations of the DIKW hierarchy », *Journal of Information Science*, vol. 33, n° 2, p. 163-180.

Ruppert E. (2012), « Doing the transparent state: Open government data as performance indicators », in Mugler J et Park S.-J. (dir.), *A World of Indicators: The Production of Knowledge and Justice in an Interconnected World*, Cambridge, Cambridge University Press, p. 51-78.

Strasser B.J. (2012), « Data-driven sciences: From wonder cabinets to electronic databases », *Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 43, n° 1, p. 85-87.

Strathern M. (dir.) (2000), *Audit culture. Anthropological studies in accountability, ethics and the academy*, Londres, Routledge.

Triclot M. (2008), *Le Moment cybernétique. La constitution de la notion d'information*, Seyssel, Champ Vallon.

Vickery G. (2011), *Review of Recent Studies on PSI Re-Use and Related Market*, Paris School of Economics, <https://ec.europa.eu/digital-single-market/en/news/review-recent-studies-psi-reuse-and-related-market-developments> (dernière consultation le 27 mars 2017).

Vismann C. (2008), *Files. Law and Media Technology*, Stanford, Stanford University Press.

Walford A. (2013), *Transforming Data: An Ethnography of Scientific Data from the Brazilian Amazon*, thèse de doctorat (PhD), Université IT de Copenhague.

Yu H. et Robinson D.G. (2012), « The new ambiguity of “open government” ». *UCLA Law Review*, n° 178, p. 178-208.