

Les données ouvertes : un bien commun

Peut-on articuler données et biens communs? En quelques lignes, pour débattre de cette publicisation

Source : <https://www.ritimo.org/Les-donnees-ouvertes-un-bien-commun>

L'open data, ou l'ouverture des données publiques, consiste en la mise à disposition proactive de données issues principalement des administrations publiques, afin de favoriser la création de valeur par les citoyens et les entreprises, une plus grande transparence de l'action publique ou encore la participation citoyenne. Juridiquement, l'ouverture exige que les données soient accompagnées de licences dites ouvertes permettant à quiconque d'utiliser les données à la seule condition d'indiquer leur source et, éventuellement, de partager les améliorations avec la même licence. Puisque l'ouverture des données a vocation à devenir une pratique courante dans les administrations, assiste-t-on à l'émergence d'un bien commun numérique ? Quelles sont les communautés qui exigent leur ouverture et défendent leur gestion en bien commun ?

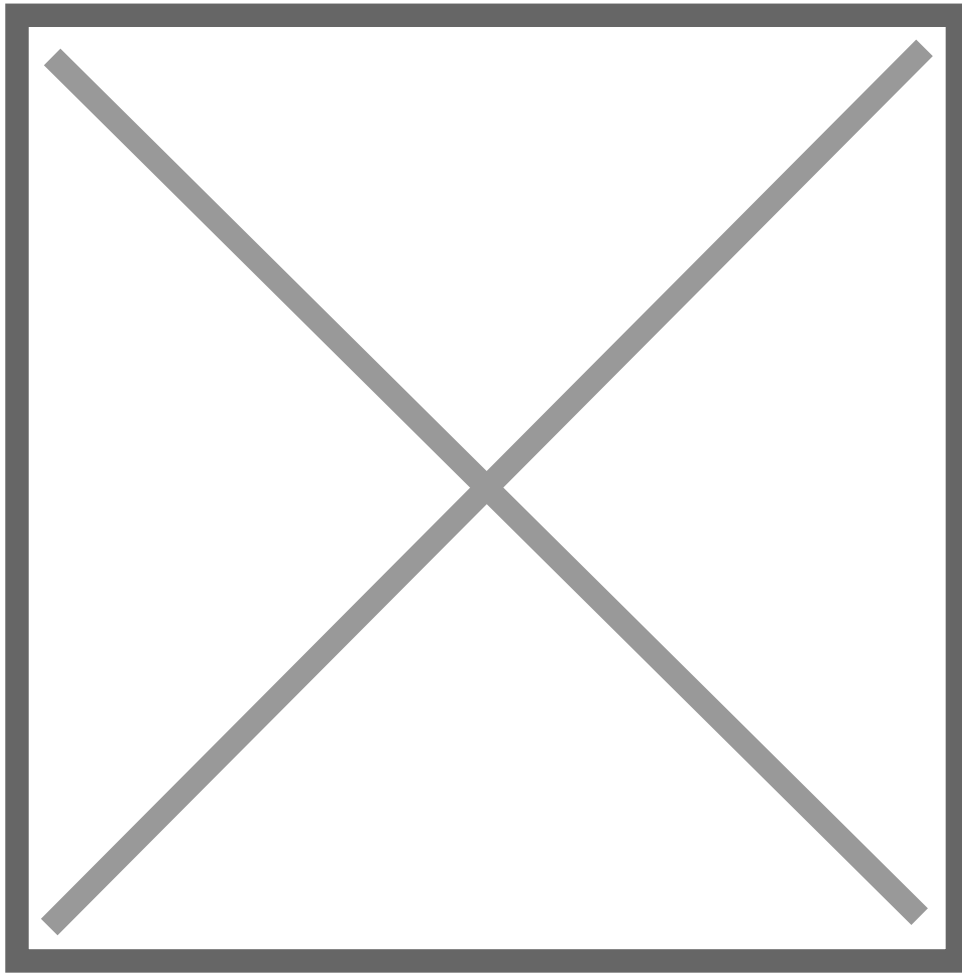
Exiger l'ouverture des données publiques

Rappelons d'abord que l'*open data* s'inscrit dans des pratiques anciennes de diffusion et de partage de l'information. L'accès à l'information détenue par l'État fait partie des revendications centrales de la Révolution Française. La Déclaration des droits de l'homme et du citoyen de 1789 stipule dans son article 15 que « la société a le droit de demander compte à tout agent public de son administration. » Après la Seconde Guerre mondiale, cette exigence de redevabilité (*accountability*) des agents publics a connu un nouvel essor avec l'émergence du concept d'*open government* qui exige que les citoyens aient accès aux secrets de l'État. Faisant face aux critiques de l'opacité de l'armée lors de la guerre du Vietnam, le président Johnson adopte en 1966 le *Freedom of Information Act* (FOIA) qui donne à tout citoyen le droit d'exiger les informations dont dispose une administration, sous réserve de certaines conditions liées à la confidentialité et à la sécurité nationale. Les dispositions du FOIA ont été reprises dans la plupart des démocraties et constituent un droit fondamental quant à la transparence de l'action publique.

Les premières pratiques de partage volontaire de données ont eu lieu dans le domaine des sciences. Avec le développement des réseaux informatiques et des capacités de calcul, le partage de données scientifiques a connu un essor considérable. Un exemple particulièrement illustratif est le projet de décodage du génome, qui a nécessité le partage de données entre laboratoires car le séquençage exigeait des infrastructures de recherche spécialisées. Bruno Strasser, dans son étude du partage des données génétiques [1], souligne la vigueur du débat sur la nécessité du partage

des données lors de la création de la première base de données du génome, GenBank. En effet, selon lui, le partage des données scientifiques s'inscrit dans une « économie morale » où les chercheurs ne diffusent leurs données que si un échange de capitaux symboliques compense leur partage. Il cite le cas de GenBank qui est parvenue à s'imposer comme la principale base de données génétiques parce qu'elle ne considérerait pas les données comme sa propriété, contrairement au projet concurrent, l'*Atlas of Protein Sequence*. Depuis 1990, les scientifiques doivent publier sur Genbank les articles utilisant des données sur le génome. Ces pratiques scientifiques qui inaugurent le partage de données à grande échelle n'étaient pas encore qualifiées d'*open data*. C'est en 1995 que le terme apparaît dans un rapport de la *National Academy of Science* des États-Unis, intitulé « *On the Full and Open Exchange of Scientific Data* » et dans lequel l'académie réclame le partage des données recueillies par les satellites.

Alors que les pratiques d'ouverture de données sont déjà établies dans les sciences, l'*open data* en tant que revendication politique fait son apparition au Royaume-Uni. En 2006, le *Guardian* publie une tribune exigeant l'ouverture des données détenues par l'État britannique. Intitulée « *Give us back our crown jewels* », la tribune revendique en particulier l'ouverture des données de l'*Ordnance Survey*, l'institut géographique du Royaume-Uni, dont l'utilisation est soumise au paiement d'une redevance. L'argument, qui deviendra par la suite un fondement central de l'*open data*, consiste à revendiquer l'ouverture de ces données en vertu du fait qu'elles sont produites à partir de l'argent des contribuables. En 2007, le mouvement *open data* entre dans le débat politique aux États-Unis, à la suite d'une réunion qui se tient à Sebastopol, en Californie, et qui regroupe des activistes numériques bien connus comme Lawrence Lessig, Tim O'Reilly et Aaron Swartz, en vue de formuler des principes susceptibles d'être repris par les candidats à l'élection présidentielle américaine. Ces principes exigent la libération des données publiques dans leur intégralité dès leur production et telles qu'elles sont collectées, assortie à une licence ouverte permettant leur réutilisation par quiconque. Cette demande de données « brutes » vise l'émergence d'une nouvelle ère de la transparence en réduisant les asymétries d'information entre l'administration et les citoyens. Elle vise également l'émergence d'une vague d'innovation via la réutilisation des données par les développeurs et les entreprises ainsi qu'un meilleur usage des ressources publiques. L'injonction à l'ouverture des données publiques a engendré une prolifération de portails qui diffusent ces données partout dans le monde avec, au niveau national, le lancement de data.gov aux États-Unis en 2008, data.gov.uk en Royaume-Uni en 2009 et data.gouv.fr en 2011. En 2013, les principes de l'*open data* sont repris par le G8 qui, dans sa charte sur l'*open data*, se fixe les mêmes objectifs de transparence et de création économique, et établit par ailleurs que l'ouverture des données deviendra la pratique par défaut des administrations des huit pays signataires.



Acteurs et bénéficiaires de l'ouverture des données publiques, image extraite de la vidéo « l'open data à la loupe » de l'association Libertic

Source Kiosque Paris, Licence : Art Libre

Une ressource qui n'est pas gérée comme un bien commun

L'injonction à l'ouverture des données publiques s'appuie sur l'argument selon lequel les données produites par les administrations et financées par l'argent public constituent un bien commun qui doit être partagé avec la société. Mais les données sont-elles gérées comme un bien commun ? Existe-t-il des mécanismes de gouvernance partagée qui permettraient à la société de gérer les données ouvertes comme une ressource commune ?

En exigeant des données « brutes », les militants de l'ouverture des données ont attiré l'attention sur des fichiers qui n'étaient jusqu'alors pas partagés et servaient au travail quotidien des agents de l'administration. Leur publication est donc inédite pour les administrations et cela amène naturellement des résistances et des questionnements. La plupart des données publiées sur les portails *open data* sont des fichiers administratifs et non des données statistiques produites spécifiquement pour créer un savoir général sur la société. Ces données sont en effet produites

pour répondre à la mission d'un service administratif et non en vue de devenir une ressource susceptible d'être utilisée par des citoyens. La sélection des données qui vont être publiées sur ces portails *open data* fait aussi l'objet de négociations entre les personnes en charge d'ouvrir les données et leurs producteurs. Ce travail de sélection, que nous avons observé lors d'une enquête dans des administrations françaises ayant libéré des données publiques [2], va à l'encontre des principes fondamentaux de l'*open data*, définis en 2007 à Sébastopol, qui exigent l'ouverture de toutes les données publiques dès leur production. Dans la pratique, les données sont au contraire sélectionnées par l'administration selon des critères variés comme la facilité techniques de leur extraction, l'intérêt pour des ré-utilisateurs potentiels ou encore la publication d'une donnée similaire sur d'autres portails *open data*. De plus, la question du degré de sensibilité des données, c'est-à-dire du risque que leur ouverture représenterait pour le public mais également pour l'administration qui les produit, est aussi souvent cruciale. Enfin, les données sont travaillées en vue de leur publication ; elles ne sont pas brutes au sens d'inaltérées. Les producteurs de données les éditent pour enlever des éléments problématiques, nettoyer leur mise en forme ou encore effacer les commentaires, acronymes et colonnes qui n'ont pas d'intérêt hors de leur contexte de production.

Toutes ces étapes qui sont cruciales dans la construction des données publiques échappent au regard de leurs utilisateurs. La production, la sélection et la modification des données publiques sont donc invisibles aux citoyens alors même que c'est dans les coulisses de ces administrations que se fabrique la transparence de l'État. Les citoyens n'ont pas non plus la possibilité de participer à la gouvernance de ces données en contrôlant leur production et leur diffusion. Pourtant, une telle régulation existe pour les données des organismes statistiques dont la mission est de produire un savoir général sur la société. En France, les statistiques publiques font l'objet depuis 1972 d'une gouvernance partagée par un organisme de concertation, le Conseil National de la Statistique (CNIS), composé de représentants des institutions, des syndicats, des organisations patronales, de la recherche et de la société civile. Le CNIS peut ainsi délibérer sur la nécessité de la production, de la diffusion ou encore sur les catégories en vigueur concernant les données produites par les organismes statistiques. Une telle gouvernance partagée n'existe pas encore pour les données des administrations dont la gestion échappe au contrôle des citoyens.

Les commoneurs [3] : producteurs et gestionnaires de données comme biens communs

En l'absence d'une gouvernance effective et redevable des données publiques, des commoneurs se sont saisis des données publiques pour préserver leur caractère de biens communs. L'*Open Knowledge Foundation*, un réseau mondial militant pour l'ouverture des contenus et des données, a publié en 2009 la licence *Open Database License* (ODbL) qui applique le principe du *copyleft* aux données ouvertes, ce qui était auparavant impossible car la licence *Creative Commons* ne s'appliquait pas aux bases de données. La licence ODbL exige le partage avec la même licence des données réutilisées, préservant ainsi leur caractère de biens communs. Des administrations comme celle de la ville de Paris ont utilisé cette licence pour éviter une éventuelle *enclosure* de leurs données, craignant que leur exploitation par des acteurs privés ne bénéficie pas à la collectivité en retour. Des commoneurs ont aussi contribué à faire émerger les données publiques

comme un bien commun en republiant leurs améliorations. En France, l'association Regards Citoyens extrait automatiquement les données issues des sites de l'Assemblée Nationale et du Sénat pour les exploiter sur son observatoire citoyen de l'activité parlementaire, nosdeputes.fr et nossenateurs.fr. Ces données sont republiées en *open data* dans des formats exploitables par les machines et ont permis le développement de nombreux autres outils de surveillance de l'activité parlementaire. De manière similaire, le site *OpenCorporates* regroupe les données des registres des entreprises de 75 juridictions dans le monde et permet à quiconque de les exploiter, même pour des usages commerciaux, tant que la clause de partage à l'identique garantissant le caractère de biens communs des données est respectée.

Par ailleurs, des bases de données gérées comme un bien commun sont aussi produites par les citoyens. Le cas le plus connu est celui d'OpenStreetMap (OSM), une base de données géographiques mondiale qui a été créée en 2004 par un chercheur britannique à la suite du refus de l'*Ordnance Survey* de partager ses données gratuitement. OSM repose sur les contributions des citoyens qui éditent le « Wikipedia de la carte ». La base de données couvre désormais la plupart des pays du monde et concurrence le géant Google Maps. En 2013, OSM a célébré son millionième contributeur. Dans les sciences, TeleBotanica est un réseau de botanistes, professionnels et amateurs, qui créent collaborativement une base de données partagée sur la faune et la flore partout dans le monde. On peut aussi citer OpenFoodFacts qui permet à chacun de scanner les données nutritionnelles présentes sur les emballages des aliments, ou encore OpenMeteoData pour le partage libre de données météorologiques. Tous ces projets reposant sur la participation des citoyens à la création de bases de données sous licence libre complètent les données publiques, et ils nous permettent d'envisager une gouvernance partagée de ces ressources numériques essentielles.

Notes

[1] Bruno J Strasser, "The Experimenter's Museum GenBank, Natural History, and the Moral Economies of Biomedicine," *Isis*, 2011, vol. 102, no. 1, p. 60-96.

[2] Jérôme Denis et Samuel Goëta. « La fabrique des données brutes. Le travail en coulisses de l'open data », in *Penser l'écosystème des données. Les Enjeux Scientifiques et Politiques des données numériques*, dir. Mabi Plantin Monnoyer- Smith, éditions FMSH, à paraître. Accessible en ligne : <http://halshs.archives-ouvertes.fr/halshs-00990771>

[3] Les commoneurs sont des individus qui produisent ou entretiennent des biens communs dont ils participent collectivement à la gestion. (voir David Bolier, *La Renaissance des communs. Pour une société de coopération et de partage*, Traduit de l'américain par Olivier Petitjean, C&F éditions, 2014)