



HAL
open science

Données ouvertes (Open Data)

Melanie Dulong de Rosnay

► **To cite this version:**

Melanie Dulong de Rosnay. Données ouvertes (Open Data). Dictionnaire des biens communs, PUF, pp.403-407, 2017, Quadriga, 978-2130654117. halshs-01575643

HAL Id: halshs-01575643

<https://shs.hal.science/halshs-01575643v1>

Submitted on 21 Aug 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.

Melanie Dulong de Rosnay, “Données ouvertes (Open Data)”, in Marie Cornu, Judith Rochfeld et Fabienne Orsi (dir.), *Dictionnaire des biens communs*, Presses Universitaires de France, Quadrige, Paris, 2017, p.403-407.

PREPRINT, DECEMBRE 2016

OPEN DATA - OUVERTURE DES DONNEES, V. Base de données, *Creative Commons*, Domaine public, Données publiques, Licence ouverte, *Open Access* (ouverture de l'accès)

Définition de la notion. L'*Open Data* ou ouverture des données désigne une donnée numérique ouverte, ou plus largement le mouvement techno-politique à l'origine de l'ouverture des données produites par le secteur public principalement, mais aussi privé. Pour être reconnues comme ouvertes, les données doivent être rendues disponibles dans un format standard ouvert et permettre la reproduction et la réutilisation libre et gratuite par tous. Des agences gouvernementales, des associations et des organisations non-gouvernementales participent à la définition du concept et à la production de plateformes et d'outils de traitement (fouille ou visualisation par exemple) permettant la réutilisation et la manipulation de ces données. Seules un minimum de conditions sont acceptables pour que les données soient reconnues comme ouvertes, notamment les obligations de leur attribuer une source, de partager à l'identique* et de non-altération. Sont principalement concernées les données publiques et les données issues de la recherche scientifique, considérées comme des biens communs informationnels et dont l'accès ouvert peut être mis en place par des lois ou des politiques institutionnelles volontaristes. D'ailleurs, la loi française du 7 Octobre 2016 pour une « République numérique » consacre la notion en organisant l'ouverture de l'accès aux données publiques, visant les informations figurant dans des documents administratifs qu'elles détiennent, incluant les bases de données qu'elles produisent ou qu'elles reçoivent, et distinguant les données d'intérêt général et les données de références. La loi vise également les données scientifiques qu'elle définit comme les données issues d'une activité de recherche.

Historique, notions voisines et fondements. Le droit d'accès aux données publiques (appelées également informations du secteur public ou documents administratifs) et la liberté d'information ont pour fondement juridique la transparence et la responsabilité des gouvernements envers les citoyens. L'article 15 de la Déclaration des Droits de l'Homme et du Citoyen de 1789 proclame ainsi que « *la Société a le droit de demander compte à tout Agent public de son administration* ». Aux États-Unis le *Freedom of Information Act* (FOIA, loi pour la liberté d'information) de 1966, en Suède une loi de 1776, et en France le titre de la loi de 1978 intitulé « de la liberté d'accès aux documents administratifs et de la réutilisation des informations publiques », codifiée code des relations entre le public et l'administration et réformée par la loi pour une République Numérique du 7 Octobre 2016, régulent l'accès aux documents administratifs des agences gouvernementales. Le mouvement *Open Data* est en essor depuis les années 2000, stimulé par les capacités techniques de reproduction, de distribution et de traitement des données numériques. Il s'est accéléré dans l'Union européenne avec les directives 2003/98/EC et 2013/37/EU et la création, dans les années 2010, de plateformes nationales de type data.gov destinées à héberger des jeux de données publiques ouvertes. Aux droits fondamentaux et à l'impératif démocratique viennent s'ajouter des justifications économiques et politiques (Uhlir, 2004), adossées à un discours techniciste et néolibéral : la transparence devrait promouvoir la participation des citoyens aux décisions publiques et la responsabilité des gouvernants ; le contribuable finançant la production

de données publiques et scientifiques devrait pouvoir accéder aux résultats ; l'innovation des entreprises, associations et citoyens, et donc la croissance économique et l'emploi, seraient stimulées par la disponibilité de données ouvertes, sur la base desquelles des services et des applications pourront se développer ; enfin, les politiques publiques et les décisions des citoyens devraient se fonder sur des données et des preuves scientifiques, l'avancée des connaissances et des découvertes scientifiques étant accélérée et facilitée par le croisement et la fouille de bases de données, qui pour ce faire doivent donc être disponibles et ouvertes juridiquement et techniquement.

Ressources et acteurs concernés. Les données ouvertes produites par les institutions publiques sont diverses et variées : statistiques et données sur le transport, l'environnement, le cadastre, le logement, la météo, l'éducation, l'emploi, la culture, le commerce, les dépenses publiques, etc. Des données ouvertes peuvent également être produites par les pairs sur la base de biens communs : c'est le cas des données géographiques du projet *Open Street Map* ou des données scientifiques du génome humain (*Personal Genome Project*). Les premières banques de données mises à la disposition des citoyens en France ont été les données juridiques (lois et décisions de justice) avec l'ouverture du service Légifrance en 1999. Les plateformes ouvertes des données publiques belges et française — data.gov.be et data.gouv.fr mis à disposition par Etalab, une mission du gouvernement dédiée à l'ouverture des données et aux développements des outils nécessaires — sont en ligne depuis 2011. Etalab a choisi de développer une licence ad hoc, la Licence Ouverte, qui comprend une clause de compatibilité externe avec la licence Creative Commons* Attribution. Pour les données scientifiques, des institutions, notamment européennes, ont développé des portails pour le dépôt des données ouvertes de la recherche (*Dataverse* ou *OpenAIRE* pour les projets européens) ou patrimoniales (*Europeana* pour les œuvres mais aussi les métadonnées produites par les bibliothèques et les musées).

Les acteurs de l'*Open Data* sont des agences gouvernementales, des chercheurs et des associations (*Open Data Institute*, *Open Knowledge Foundation*, Regards Citoyens, *Open Government Partnership*) qui vont promouvoir la libération et la réutilisation des données en facilitant la mise à disposition ouverte juridique mais aussi technique, avec des formats et standards ouverts permettant de lier, structurer et enrichir les données. Elles produisent des recommandations, critères et définitions de ce que constitue une donnée ouverte. Selon l'*Open Definition* (un standard qui propose des critères minimum pour définir l'open), une donnée sera ouverte si elle peut être utilisée, modifiée et partagée librement et gratuitement par quiconque pour tout propos et avec pour seules conditions acceptables l'attribution, l'intégrité, le partage à l'identique, la notice de la licence, ou le format de la source.

Mise en œuvre juridique : lois sur le droit d'auteur, législations sur les données publiques et licences ouvertes. Trois méthodes juridiques, mutuellement non exclusives, conduisent à la disponibilité des données publiques : leur placement d'emblée dans le *domaine public** par la loi; des lois spécifiques sur l'ouverture des données publiques ; et, enfin, l'utilisation de *licences ouvertes** pour leur distribution sur des plateformes dédiées, selon des conditions oscillant entre la dédicace contractuelle dans un *domaine public volontaire* (les Pays-Bas utilisent CC0) et l'utilisation de clauses de *partage à l'identique*. Les législations sur le droit d'auteur et/ou les informations publiques et les politiques et pratiques varient grandement en la matière (Jasserand & Hugenholtz, 2011).

Au titre du premier modèle, les États peuvent faire le choix de lever tout droit d'auteur sur les

œuvres et données produites par leurs agents publics, les plaçant de facto dans le *domaine public* et levant ainsi toute limitation juridique à l'accès et à la réutilisation. La section 105 de la loi sur le *Copyright* aux États-Unis prévoit que les œuvres créées par des agents fédéraux ne bénéficient pas de protection dans le cadre domestique. En revanche, le *Copyright* peut s'appliquer quand les données sont utilisées en dehors du cadre domestique, à l'étranger, et une exception concerne les données de référence standard produites par le Secrétariat du Commerce (*Standard Reference Data Act*, 1968). Dans le même ordre d'idées, en vertu de l'ancien article 1235 du Code civil du Mexique de 1928, le gouvernement ne pouvait pas obtenir de droit d'auteur ; la loi sur le droit d'auteur de 2013 lui applique l'extension de la durée de protection de 100 ans. De nombreuses législations européennes (Allemagne, Belgique, Espagne, France, Pologne, Suède, etc. ; van Eechoud, 2011) exemptent de droit d'auteur les textes et actes officiels, un sous-ensemble des données publiques.

Dans l'Union européenne, et au titre d'un deuxième mécanisme, les États sont dotés de législations sur les données publiques, harmonisées par la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 sur la réutilisation des informations du secteur public révisée par la directive 2013/37/EU. Le droit de l'Union européenne fixe un socle minimum d'obligations (Janssen & Dumortier, 2003 ; van Eechoud, 2014) concernant la disponibilité pour l'accès et la réutilisation des documents et informations produites et détenues par les organismes du secteur public. En France, la liberté d'accès aux documents administratifs est organisée par le code des relations entre le public et l'administration issue de la loi de 1978 du même nom et prévoit l'accès *via* la Commission d'accès aux documents administratifs (CADA). En Belgique, c'est l'article 32 de la Constitution qui prévoit que « *Chacun a le droit de consulter chaque document administratif et de s'en faire remettre copie* ». Cependant, de très nombreuses exceptions grèvent le principe général d'accès et de libre réutilisation (voir la dernière section de cette notice, *infra*). De plus, ces législations portent sur la communication des données suite à des demandes effectuées par des requérants (« *accès passif* », van Eechoud, 2011) et n'imposent aux institutions publiques aucune obligation de publier les données (« *accès actif* », van Eechoud, 2011).

Un troisième mécanisme permet aux États de mettre à disposition des jeux de données de manière pro-active, sans attendre une saisine citoyenne. Les plateformes gouvernementales du type *data.gov* distribuent ainsi les données publiques sous différents types de licences ouvertes, qui vont permettre la réutilisation, même à des fins commerciales selon certaines conditions. L'Australie, les États-Unis, le Brésil ou le Burkina Faso utilisent une licence *Creative Commons** Attribution, tandis que le gouvernement néerlandais diffuse ses données sous *CC0*, le protocole *Creative Commons* de dédicace anticipée dans le *domaine public*. La France, la Grande-Bretagne et l'Italie ont fait le choix de développer leurs propres licences ouvertes gouvernementales, tout en les décrétant compatibles avec les licences *Creative Commons* Attribution afin de permettre l'interopérabilité avec d'autres jeux de données. L'utilisation de licences standard diminue les coûts d'information et de transaction, mais impose une obligation contractuelle d'attribution assez lourde et ne permet pas d'insérer des clauses spécifiques, comme l'offre de garanties ou l'interdiction d'altérer les données (van Eechoud, 2011). Certaines institutions utilisent une licence qui comprend une clause de *partage à l'identique* afin de maintenir les éventuelles bases de données dérivées sous le même régime. Plusieurs associations dont l'*Open Knowledge Foundation* ont développé des licences ouvertes destinées spécifiquement au droit *sui generis* des producteurs de bases de données*. L'*Open Database Licence (ODBL)* choisie par la Ville de Paris pour son portail *Open Data* a pour objectif de conserver les versions modifiées ouvertes et

donc dans le *commun*. Dans la pratique, appliquer une clause de *partage à l'identique* à l'un des 164 jeux de données concernés peut devenir problématique car cela soulève la question de la qualification de ce que sera une « base de données collaborative », partie d'un ensemble de bases de données à laquelle la condition de viralité ne s'appliquera pas, et une « base de données modifiée », qui devra être partagée selon les mêmes conditions, avec l'ensemble des modifications apportées, des différences ou de la méthode. Ces définitions, provenant du droit d'auteur et de ses œuvres collectives ou dérivées, imposent de lourdes obligations aux éventuels ré-utilisateurs, et vont plus loin que les lois sur l'ouverture des données publiques.

Enjeux et limites. Les lois sur l'ouverture des données publiques s'articulent difficilement avec d'autres droits comme la protection des données à caractère personnel (Dulong de Rosnay & Janssen, 2014 ; van Eechoud, 2014) et leur portée effective est limitée par de très nombreuses exceptions (imposées aux organismes tenus de générer des recettes, la vie privée, la sécurité nationale, les ressources culturelles, les droits des tiers, etc.). Les licences ouvertes ne pourront pas être appliquées quand l'information publique doit être valorisée financièrement, l'incitation à la marchandisation étant peu compatible avec les principes de l'open data. Enfin, les solutions pour le respect de la vie privée incluent la publication sous une forme agrégée, supposée garantir leur caractère anonyme, mais difficile à garantir dans la pratique.

Certaines institutions imposent le paiement de redevances pour l'utilisation des données publiques, en vertu des lois sur la réutilisation des informations du secteur public qui permettent une tarification au coût marginal, ou à un coût supérieur comme dans le cas des bibliothèques et des musées. Ces redevances peuvent figurer dans les clauses de licences comme celle de l'Agence du Patrimoine Immatériel de l'État (APIE) en France. Basées sur le coût marginal de reproduction, les redevances peuvent constituer d'importantes barrières à la réutilisation des données gouvernementales (Morando & al, 2013). Elles peuvent s'avérer nécessaires pour équilibrer les budgets des institutions qui produisent les données et sont sommés de générer des recettes, même si elles ne correspondent pas aux définitions de données ouvertes agréées par les associations de l'*Open Data* et peuvent être qualifiées d'*enclosure** du *commun* ou de *copyfraud*. Le mouvement de l'*Open Data* est louable et les efforts pour la mise en commun de données disponibles et réutilisables est un bon stimulant pour la démocratie et l'économie. Il convient toutefois de ne pas veiller uniquement à parfaire le degré d'ouverture des données, sous peine de se retrouver avec des portails qui seraient des coquilles vides. On relativisera enfin l'impact des grandes masses de données (*big data*) tant la production de connaissances requiert plus que le traitement de jeux de données rendu possible par leur ouverture, traitement qui n'est pas accessible à tous et peut conduire à renforcer les inégalités.

Mélanie DULONG DE ROSNAY, Chargée de recherche au CNRS, Institut des Sciences de la Communication (CNRS-Paris Sorbonne-UPMC)

Repères bibliographiques :

- DULONG DE ROSNAY M. & JANSSEN K., « Legal and Institutional Challenges for Opening Data across Public Sectors: Towards Common Policy Solutions », *Journal of Theoretical and Applied Electronic Commerce Research (JTAER)* Special Issue “Transparency and Open Data Policies” ISSN 0718–1876, Vol. 9, Issue 3, September 2014, pp. 1-14
- JANSSEN K., & DUMORTIER J., « Towards a European Framework for the Reuse of Public Sector Information: a Long and Winding Road », *International Journal of Law and*

Information Technology, 2003, Vol.11(2), pp.184-201

- JASSERAND C. & HUGENHOLTZ B., *Using Copyright to Promote Access to Information and Creative Content. Public Sector Information* (Part III), 2011. Document WIPO/CR/WK/GE/11/4
- MORANDO F., IEMMA R. & BASSO S., « Is there such a thing as free government data? », *Internet Policy Review*, vol. 2, issue 4, HIIG - Alexander von Humboldt Institute for Internet and Society, 2013
- Open Definition, <http://opendefinition.org/od/2.0/en/>
- van EECHOU M., « Friends or Foes? Creative Commons, Freedom of Information Law and the European Union Framework for Reuse of Public Sector Information », in GUIBAULT L., & ANGELOPOULOS C. (ed.), *Open Content Licensing: From Theory to Practice*, Amsterdam: Amsterdam University Press 2011, p. 169-202
- van EECHOU M., *Position paper on Access to Data*, Deliverable 2.3 of the European Thematic Network on Legal Aspects of Public Sector Information, 2014, 14p