

COVID Open Data

Rapport

Open Data / Données Ouvertes

Kawtar HAMDI
Lisa ESTEBE
Salomé CHEVAILLER
Stéphane SUN

Sommaire

Sommaire	1
Introduction	2
I - Qu'est ce que l'Open Data ?	3
Principes de l'Open Data	3
L'Interopérabilité	5
Historique	6
Cadre d'utilisation	7
La législation	7
Les licences	9
II - Utilisation Générale	11
Les Gouvernements	11
Exemples courants	12
Classement des données	13
Impact sur la société	14
III - Santé	16
Utilisation en santé	16
Un suivi personnel des patients	16
Créer de nouveaux écosystèmes de santé	17
Contribuer à l'avancée de la recherche sur les maladies rares	17
L'accélération de la recherche sur le cancer	18
COVID	19
Conclusion	21
Bibliographie	22

Introduction

L'arrivée de l'informatique marque le début de l'automatisation du traitement et de la gestion des données.

Des quantités massives de données sont produites et collectées en permanence. Celles-ci pouvant être sous différentes formes, provenir de différentes sources et concerner différents sujets. Elles sont exploitées pour différentes fins notamment la prévision, la simulation, la modélisation ...

Ces données peuvent être générées passivement ou volontairement et certaines d'entre-elles sont considérées comme étant "publiques", c'est-à-dire accessibles à tous. Elles sont qualifiées comme étant "ouvertes".

Les enjeux auxquels contribuent les données dont les données ouvertes obligent les gouvernements ainsi que les acteurs de ce domaine, à s'accorder sur leur utilisation afin d'éviter des dérives ou abus.

I - Qu'est ce que l'Open Data ?

Principes de l'Open Data

Dans cette partie, nous aborderons la notion d'Open Data, traduite en français par “données ouvertes”, ainsi que ses différents principes.

L'Open Data fait référence à des données accessibles, utilisables et partageables par n'importe quelle personne, sans conditions techniques, juridiques ou financières. Leurs contenus peuvent être diffusés par différents acteurs, comme des services publics, des associations ou encore des entreprises privées. Culturellement, on peut associer l'Open Data à une philosophie du partage du savoir avec le plus grand nombre de personnes possible. Politiquement, l'Open Data suit une démarche de transparence et d'implication des citoyens dans la société qu'ils partagent¹.

D'après l'Open Knowledge Foundation, les principales caractéristiques de l'Open Data sont la disponibilité, la réutilisation et la distribution, ainsi que la participation universelle. Le critère de la disponibilité indique que les données doivent être entièrement accessibles et moyennant un coût de reproduction raisonnable, c'est-à-dire pouvant être téléchargeables sur Internet de préférence². Ces données doivent aussi être disponibles sous une forme modifiable et plutôt pratique. Le critère de la réutilisation et de la distribution indique que les données doivent être partagées suivant des conditions qui permettent leur réutilisation et leur redistribution, tout en pouvant être mélangées avec d'autres données. Et enfin, pour respecter le critère de la participation universelle, il faut que les données puissent être utilisées, réutilisées et redistribuées par tout le monde, sans aucune discrimination envers des domaines d'activités, des personnes ou des groupes³.

L'Open Data a trois principaux objectifs. Le premier étant d'améliorer le fonctionnement de la démocratie en permettant la transparence, la concertation et l'ouverture à de nouveaux points de vue. Le second objectif est d'améliorer l'efficacité de l'action publique. Et, en permettant le partage de données entre utilisateurs et réutilisateurs qui s'en servent dans des services à forte valeur ajoutée, cela remplit le troisième objectif de l'Open Data qui est de proposer de nouvelles ressources pour l'innovation économique et sociale⁴.

¹ Futura, « Open data ».

² « What Is Open? »

³ L, « Open Data définition ».

⁴ « LA DEMARCHE ».

Pour qu'une donnée soit considérée comme "ouverte", elle doit suivre 10 critères. Ces critères ont été établis en 2010 par la Sunlight Foundation, une organisation non partisane américaine en faveur du gouvernement ouvert, et en voici la liste :

- 1) Complète
- 2) Primaire
- 3) Opportune
- 4) Accessible
- 5) Exploitable
- 6) Non-discriminatoire
- 7) Non-propriétaire
- 8) Libre de droits
- 9) Permanente
- 10) Gratuite

Toujours en 2010, l'informaticien britannique Tim Berners-Lee, principal inventeur du World Wide Web, a proposé une échelle allant de 1 à 5 étoiles évaluant la qualité des données ouvertes.

Cette échelle prend en compte les données numériques de base mais peut également être adaptée à des photos, vidéos, rapports, etc. qui peuvent être mis en ligne⁵.

★	Données non filtrées (éventuellement dégradées) par exemple mises en ligne avec n'importe quel format
★ ★	Données disponibles de manière structurée (ex : données tabulaires en CSV, XML, Excel, RDR)
★ ★ ★	Données librement exploitables - juridiquement (Cf. licences), - techniquement (dans des formats non-propriétaires, pas sous Excel notamment)
★ ★ ★ ★	Données identifiées par des URL (avec date de mise à jour) afin que l'on puisse « pointer » un lien vers elles (et les retrouver éventuellement mises à jour)
★ ★ ★ ★ ★	Données liées à d'autres données, pour les contextualiser et enrichir

Attention à ne pas confondre les données ouvertes avec les données fermées.

⁵ « Données ouvertes ».

Comme vu précédemment, les données ouvertes reposent sur trois critères qui sont leur accessibilité, leur disponibilité dans un format lisible, et leur possibilité d'être utilisées et partagées par quiconque à des fins commerciales ou non.

Les données fermées, quant à elles, restreignent l'accès à certaines informations. Par exemple, elles peuvent être disponibles que pour certaines personnes au sein d'une organisation, être brevetées ou exclusives, être semi-limitées pour certains groupes de personnes. Elles peuvent également avoir des frais de licence afin d'y accéder, ou encore être très difficilement accessibles.

Toutes les données nécessitant la sécurité sont donc fermées, comme peuvent l'être les informations relatives à la santé recueillies par un hôpital ou une compagnie d'assurance⁶.

L'Interopérabilité

Nous allons maintenant définir ce qu'est l'interopérabilité et voir en quoi c'est une notion directement liée à l'Open Data.

L'interopérabilité est la capacité de mélanger différents ensembles de données. Elle permet donc à différents composants de fonctionner ensemble, créant ainsi des systèmes larges et complexes (chose n'étant pas possible sans interopérabilité).

Pour l'Open Data, il est nécessaire que les données puissent être mélangées librement. L'interopérabilité est donc essentielle pour pouvoir tirer des bénéfices de cette ouverture des données. Grâce à cela, il sera possible de développer plus de produits et services et de qualité bien meilleure.

Il est aussi nécessaire que les données partagées utilisent un langage de programmation commun ou qu'il existe un élément qui constitue un programme faisant l'intermédiaire entre les informations. C'est pour cette raison que le World Wide Web Consortium (W3C) vante l'application de standards de l'Open Data, permettant ainsi de répondre à ce besoin de communication.

⁶ Team, « What Is Open Data? »

De ce fait, il est nécessaire que certains critères soient respectés lors de la publication de données, comme la certification de la provenance de la donnée, l'indication des métadonnées liées, c'est-à-dire la date et l'heure de la création, ainsi que la garantie de la qualité de l'information avec le nom de l'auteur si possible³.

Historique

Si nous avons accès aujourd'hui à une majorité de rapports, de recherches, scientifiques, cela n'a pas toujours été le cas. En 1942 Robert Merton publiait un essai portant sur la sociologie des sciences dans lequel il définit le but de la science comme suit : « L'objectif institutionnel de la science est l'extension de la connaissance certifiée [*the institutional goal of science is the extension of certified knowledge*] ». En effet, pour nos sociétés modernes dont le fonctionnement nécessite la production de connaissances certifiées, la « science » n'est autre que l'activité collaborative capable d'assurer cette production. Mais encore faut-il accéder à toutes ces connaissances.

Dans les années 1970, Merton, considéré alors comme le fondateur de la sociologie des sciences, avance l'idée que la recherche doit être accessible à tous. Selon lui, chaque chercheur doit contribuer au « pot commun » et abandonner les droits de propriété intellectuelle pour permettre aux connaissances de circuler.

Avec l'essor des revues en 1990, apparaît le mouvement Open Science, il reflète l'intention initiale de la science en soutenant la transparence et la collaboration dans la recherche et la communication scientifique. Ce mouvement fait suite au constat que les recherches scientifiques sont payées par les fonds publics, et donc les contribuables ne devraient pas être restreints à ces résultats. En découle une large demande d'accès aux publications scientifiques. Les auteurs et les revues se mettent dès lors à adopter les Creative Commons, licence permettant à n'importe qui de lire et d'utiliser librement les publications. Aujourd'hui, nous sommes toujours dans cette transition, il existe une multitude de revues libres d'accès, d'autres demeurent payantes comme Elsevier ou Springer.

Il faudra attendre 1995 pour que le terme d'Open Data apparaisse pour la première fois dans une revue scientifique américaine. Les données Open Data sont basées sur l'idée que non seulement les résultats et les rapports de recherches devraient être ouverts, mais également les données sous-jacentes qui les soutiennent. Elinor Ostrom, prix Nobel d'économie, identifie l'Open Data comme étant une nouvelle sorte de « bien public » et estime que ses données sont un enrichissement pour notre société.

Avec les progrès technologiques et l'informatique, la capacité de produire et de partager des données en quantité s'est largement développée. A notre époque où le volume de données générées quotidiennement est considérable, la demande d'une capacité de stockage de données toujours plus importante ne cesse de croître.

On ne compte plus les outils capables d'accéder à toutes ces données, plates-formes numériques, capteurs sans fil, applications de réalité virtuelle et téléphones portables nous placent l'information à portée de clic. Véritable phénomène mondial, la tendance à l'Open Data appuie les mouvements innovants dans l'analyse de données incluant le Big Data, l'intelligence artificielle et l'apprentissage automatique.

On ne peut que constater une demande de données « ouvertes par défaut », les gouvernements proposent d'ailleurs de plus en plus de données ouvertes sur leurs sites Web, les chiffres concernant l'évolution du Covid-19 sont par exemple accessibles sur le site du gouvernement français. Le désir, la demande, et avec eux les attentes, pour les données ouvertes, s'amplifient chaque jour, l'Open Data devient la nouvelle norme⁷.

Cadre d'utilisation

La législation

La réglementation juridique concernant les données ouvertes étant différente dans chaque pays, nous allons ici principalement évoquer le cadre juridique français.

Les enjeux et l'influence que peuvent avoir les données ouvertes, obligent à définir un cadre législatif à l'utilisation ainsi qu'à la définition des données ouvertes⁸. Les données concernées sont principalement celles issues des services publics.

Le gouvernement est en effet tenu de partager ses données sur ses activités et ses travaux. Ce droit d'accès aux informations est inscrit dans la Déclaration des droits de l'homme et du citoyen de 1789 par les articles suivants⁹ :

Article 14 :

Tous les citoyens ont le droit de constater, par eux-mêmes ou par leurs représentants, la nécessité de la contribution publique, de la consentir librement, d'en suivre l'emploi, et d'en déterminer la quotité, l'assiette, le recouvrement et la durée.

Article 15 :

⁷ *Reaping the benefits of Open Data in public health*, P Huston , VL Edge, E Bernier

⁸ « *Projet_Open_Datalab-quelles_obligations_rglementaires-v1.1.pdf* ».

⁹ « Déclaration des Droits de l'Homme et du Citoyen de 1789 | Conseil constitutionnel ».

La société a le droit de demander compte à tout agent public de son administration.

Hormis certaines exceptions, tout document produit ou reçu par l'administration française dans le cadre de sa mission doit donc être obligatoirement publié¹⁰.

Cependant, l'accès au public de ces informations peut entrer en conflit avec le Règlement Général sur la Protection des Données (RGPD) ayant pour objectif de protéger toute information pouvant être reliée à une personne physique. Afin de concilier l'ouverture des données et la protection des données personnelles, plusieurs lois ont été mises en place¹¹.

Au niveau national, plusieurs lois se sont succédées depuis 1978 afin d'amener progressivement l'ouverture des données publiques¹² et de l'intégrer au sein de la société. La mise à disposition des informations publiques par le gouvernement par la possibilité d'y avoir accès numériquement facilement pour tous date cependant de 2011 avec la création de la mission Etalab¹³, ayant pour objectif de centraliser et de coordonner la conception et la mise en œuvre de la stratégie de l'État dans le domaine des données.

- 1978 - [Loi n°78-753 du 17 juillet 1978, dite "loi CADA"](#) : création du droit d'accès aux documents administratifs
- 2005 - [Ordonnance n°2005-650 du 6 juin 2005 sur la réutilisation de l'information publique](#) (transposition de la directive européenne de 2003) : création du droit de réutiliser l'information publique
- 2011 - [Création d'une mission « Etalab »](#) chargée de la création d'un portail unique interministériel des données publiques
- 2015 - [Loi relative à la gratuité et aux modalités de la réutilisation des informations du secteur public](#) : création du droit de réutiliser librement les données publiques
- 2016 - [Loi pour une République numérique](#) : consécration du principe de l'open data par défaut.
- 2015: [Loi pour la croissance, l'activité et l'égalité des chances économiques](#) : ouverture en open data des données de transport
- 2015 - [Loi sur la Nouvelle organisation territoriale de la République](#) : publication en open data des données des collectivités publiques de plus de 3500 habitants
- 2016 - [Loi pour la Modernisation de notre système de santé](#) : publication en open data des données de santé

¹⁰ « Guide pratique de la publication en ligne et de la réutilisation des données publiques ("Open Data") ».

¹¹ « Les collectivités territoriales et l'open data ».

¹² « Chronologie de l'open data | guides.etalab.gouv.fr ».

¹³ Décret n° 2011-194 du 21 février 2011 portant création d'une mission « Etalab » chargée de la création d'un portail unique interministériel des données publiques.

L'une des lois notables, la loi Lemaire, datée du 7 octobre 2016¹⁴, marque le début de la promotion du numérique en France et a notamment pour objectif de favoriser l'ouverture des données en précisant le cadre juridique des données et en précisant aussi le rapport de ces données auprès du public¹⁵.

Les informations à caractère personnel ne peuvent donc pas être considérées comme étant des informations publiques.

Plusieurs lois sont apparues depuis, afin d'étendre l'ouverture des données vers d'autres domaines autres que celui de l'administration française.

Du point de vue européen, les premières directives d'ouverture datent de 2003 mais les premières obligations de transparence des États datent de 2007 par l'obligation de publier certaines données¹⁶.

- 2003 - [Directive européenne 2003/98/CE, dite PSI](#) : ensemble de règles concernant la réutilisation des données et documents détenus par les organismes des États membres de l'Union européenne
- 2007 - [Directive européenne INSPIRE](#) : obligation de publier en open data les données environnementales et géographiques
- 2013 - [Directive 2013/37/UE modifiant la directive 2003/98/CE](#) : encadrement du droit de redevance accordé aux administrations
- 2018 - [Directive 2019/1024/UE concernant les données ouvertes et la réutilisation des informations du secteur public](#) : inclusion des données des entreprises investies d'une mission de service public dans le champ de l'open data.

La troisième directive datant du 20 juin 2019¹⁷ permet d'inclure les informations fournies par des entreprises participant à des missions d'intérêt public, comme étant des informations publiques, ce qui permet leur libre utilisation.

Les licences

Afin d'encadrer les projets utilisant les données ouvertes, les fournisseurs de ces jeux de données peuvent restreindre les données sous différentes licences ayant chacune diverses spécificités.

La licence ouverte, créée à l'initiative du gouvernement français par Etalab permet de réutiliser les données de n'importe quelle manière, tant qu'elle cite sa source. Elle permet donc notamment de :

¹⁴ LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1).

¹⁵ « Chapitre Ier : Etendue du droit de réutilisation (Articles L321-1 à L321-3) - Légifrance ».

¹⁶ Directive 2007/2/CE du Parlement européen et du Conseil du 14 mars 2007 établissant une infrastructure d'information géographique dans la Communauté européenne (INSPIRE).

¹⁷ Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public.

- Partager (copier, distribuer, utiliser)
- Adapter (modifier, construire, transformer)
- Réutiliser à des fins commerciales
- Produire à partir de la base de données¹⁸.

Elle est la licence offrant le plus de libertés concernant les possibilités qu'elle procure et a été créée spécifiquement pour supporter les projets utilisant les données ouvertes dans le cadre des missions du service public¹⁹.

La licence ODbL (Open Database Licence) est quant à elle issue de l'Open Knowledge Foundation. Elle offre les mêmes libertés que la licence ouverte mais a cependant certaines restrictions supplémentaires²⁰ :

- Lorsque l'on utilise le jeu de données d'une source OBdL et qu'on décide de l'adapter, il faut fournir la nouvelle version adaptée sous OBdL
- La redistribution d'un jeu de données adapté sous OBdL doit obligatoirement se faire aussi sous OBdL

¹⁸ « ETALAB-Licence-Ouverte-v2.0.pdf ».

¹⁹ « Licence Ouverte / Open Licence ».

²⁰ « ODC Open Database License v1.0 | Software Package Data Exchange (SPDX) ».

II - Utilisation Générale

Les Gouvernements

Dans tous les pays du monde, les gouvernements collectent toutes sortes d'informations sur les personnes, les événements, les institutions et l'environnement, qui sont facilement stockables dans des bases de données en ligne.

Au départ, ces informations n'étaient disponibles que par certaines agences du gouvernement mais elles sont aujourd'hui disponibles au grand public pour différentes raisons. Premièrement, la transparence, qui garantit que les activités du gouvernement sont accessibles par les citoyens. De plus, cela permet de dégager une valeur sociale et commerciale puisque les citoyens peuvent utiliser ces données de manière innovante créant une utilité supplémentaire. Pour finir, les citoyens peuvent ainsi avoir accès aux données sur lesquelles se fondent les grandes décisions gouvernementales. On peut alors parler de gouvernance participative²¹.

Les données gouvernementales permettent donc une meilleure transparence et un contrôle démocratique accru. Les citoyens se sentent ainsi impliqués puisqu'ils ont davantage de pouvoir, augmentant ainsi le taux de participation. Ces données ont aussi propulsé l'innovation et les entreprises peuvent créer et améliorer des produits et services privés³.

En France, la plateforme data.gouv.fr a été créée en 2011, permettant ainsi aux services publics de publier des données qui peuvent être enrichies, modifiées et interprétées par la société civile afin de produire ensemble des informations d'intérêt général.

Pour ce faire, il suffit de s'inscrire gratuitement sur la plateforme sous son identité de personne physique ou de personne morale. Cette plateforme est le premier site au monde à proposer ce type de service²².

Ainsi, la France fait partie des sept pays les mieux classés dans le Global Open Data Index 2016 (GODI), géré par l'Open Knowledge Network, qui évalue l'ouverture des données concernant par exemple les budgets gouvernementaux, la qualité de l'air, les cartes nationales, les dépenses publiques, les résultats des élections, les prévisions météorologiques, etc²¹.

²¹ D'Agostino et al., « Open data and public health ».

²² « L'ouverture des données publiques ».

Exemples courants

L'Open Data est partout autour de nous, dans tous les domaines de notre vie quotidienne. De la signalisation de bouchons ou du retard d'un train, jusqu'au suivi en temps réel de notre livreur de pizza.

Les données ouvertes nous aident à prendre des décisions. Où trouver des toilettes au Danemark ? findtoilet.dk liste les toilettes publiques du pays. Quand sortir prendre l'air aux Pays-Bas ? Il est possible de recevoir un message si la qualité de l'air de son quartier se dégrade. Où promener son chien à New York ? Il suffit de consulter la liste des parcs pour chiens de la ville.

La mairie de Paris possède son propre portail de données ouvertes où elle référence notamment la localisation des Vélib et AutoLib.

Utilisées massivement par les gouvernements, ces données permettent aux citoyens de s'investir dans la vie de leur ville, de leur pays, de se sentir plus concernés. Au Danemark il est possible de suivre l'activité du parlement et les processus de création de lois via le site internet folketsting.dk. Les données ouvertes gouvernementales permettent une meilleure transparence, conférant plus de pouvoir aux individus. *Tax Tree* en Finlande et *Where does my money go* en Grande Bretagne renseignent sur l'utilisation de l'argent des Taxes.

L'Open Data c'est aussi de l'économie comme pour Google Traduction qui utilise les documents de l'Union Européenne traduits en plusieurs langues pour développer ses algorithmes de traduction et améliorer ainsi son service.

Elles donnent aux entreprises la possibilité d'améliorer leurs services privés, leurs produits, et s'érigent en véritable propulseur de l'innovation. Les services gouvernementaux deviennent de plus en plus efficaces.

En combinant différentes sources de données massives de nouvelles connaissances peuvent être générées, comme c'est le cas pour la santé. C'est en croisant les données sur les morts du choléra et l'emplacement des sources d'eau que le Dr. Snow a découvert la relation entre la pollution de l'eau potable et le choléra à Londres au XIX^{ème} siècle.³ Imaginons au XXI^{ème} siècle, avec toutes les données accessibles grâce à l'Open Data, les découvertes qui pourraient naître.

Classement des données

Au vu des enjeux ainsi que de l'impact de l'Open Data, plusieurs classements sur l'accessibilité des données en fonction des pays ont été établis.

Ces classements sont particulièrement importants pour les démocraties, car elles attestent de la transparence des gouvernements, mais aussi car elles permettent aux populations de contribuer et participer aux différents à la société en débattant, critiquant ou proposant des actions à partir de l'analyse de ces données²³.

Ces classements classent les pays en fonction de l'accessibilité des données ouvertes. Parmi ces différents classements, 3 sont considérés comme étant les plus pertinents :

L'Open Knowledge Foundation avec le Global Open Data Index qui se base sur les critères suivants :

- Le format des données publiées
- L'implication du gouvernement dans la publication et collecte des données
- La manière dont sont collectées les données
- La précision et la diversité des données
- L'intervalle de réactualisation des données

Ce classement prend en compte tout type de données, peu importe leur origine gouvernementale.

Cependant il ne juge pas ni ne prend en compte la manière dont ces données sont utilisées, ni l'impact qu'elles ont sur leur environnement. Il ne prend pas non plus en compte la qualité des données fournies²⁴.

Ce classement se focalise donc principalement sur l'origine et le partage des données.

L'OCDE (Organisation de Coopération et de Développement Économique) quant à elle base son classement sur l'accessibilité des données ainsi que l'utilité et ré-usabilité des données gouvernementales. Elle ne prend cependant en compte que les données sur des thèmes particuliers :

- L'économie
- La finance
- Les transports
- La société
- La géographie
- L'environnement

²³ Ruijter, Grimmelikhuijsen, et Meijer, « Open Data for Democracy ».

²⁴ « Methodology - Global Open Data Index ».

Les pays sont donc principalement classés à partir de la transparence des données disponibles ainsi que leur impact dans leur pays²⁵.

Ce classement se focalise donc essentiellement sur l'utilisation des données gouvernementales.

La World Wide Web Foundation établit le classement OpenData Barometer et ses critères se basent sur :

- L'accessibilité des données
- L'utilisation de ces données
- L'impact des données sur la société (politique, économique, social)

Et prend cependant en compte uniquement les données issues des gouvernements, des citoyens et de la société civile ainsi que du monde de l'entrepreneuriat²⁶.

Son classement n'est donc pas uniquement restreint aux données gouvernementales, mais il se focalise cependant sur le rôle joué par ces données au sein de la société.

Bien que ces trois classements soient tous différents et se focalisent sur des notions différentes, historiquement les même pays se retrouvent à la tête de ces classements.

La France par exemple a toujours fait partie des pays les mieux classés. Pour l'édition des classements de 2019 par exemple, elle a été dans le top 4, tous classements confondus.

Les pays en tête de classement sont essentiellement des pays démocratiques. Cela s'explique notamment par le fait que ces classements accordent une certaine importance aux données issues des gouvernements mais aussi de l'importance des impacts qu'ont les données ouvertes.

Impact sur la société

Si l'Open Data apporte beaucoup à nos sociétés, certains sont angoissés par cette transparence. Dans un premier temps, il est essentiel de différencier Open Data et données accessibles au public. Si certes tout le monde accède aux données publiques, toutes les données accessibles au public ne sont pas ouvertes³.

Il est indéniable que l'Open Data nous accompagne dans notre vie de tous les jours et la facilite. Les données ouvertes rythment nos gouvernements, notre économie, notre culture, notre système d'éducation ou encore notre système de santé.

²⁵ « Open Government Data - OECD ».

²⁶ « ODB-4thEdition-Methodology.pdf ».

L'impact de l'Open Data sur les recherches est considérable. Mais si elles gagnent en popularité et en nécessité, les données ouvertes ne sont pas encore généralisées. La raison principale étant que les chercheurs ne savent pas comment le partage de leurs travaux affectera leur carrière. Pourtant, il est établi que les recherches dites ouvertes sont plus de fois citées, les articles en libre accès ont un avantage de 36 à 172%, et bénéficient d'une plus grande couverture médiatique, ceux en libre accès sont consultés près du double de fois. Ce qui mène à des opportunités de financement, des opportunités d'emploi et la mise en évidence de potentiels collaborateurs. Il est prouvé que les pratiques de recherche ouvertes apportent des avantages significatifs aux chercheurs par rapport aux pratiques fermées plus traditionnelles²⁷.

Les preuves que le partage ouvert d'articles et de données est bénéfique pour les chercheurs sont bien réelles. Plus encore, ce partage est profitable aux recherches elles-mêmes. En effet, avec la possibilité de croiser de multiples données, avec une telle quantité de ressources à portée de main, des découvertes naissent.

Les données ouvertes ont d'ores et déjà permis d'étendre considérablement le savoir et les connaissances. Elle favorise également la croissance économique. Les entreprises développent de plus en plus de solutions et de logiciels basés sur ce principe. Elles sont aussi bénéfiques aux chercheurs ainsi qu'à la recherche.

C'est donc tout naturellement que l'on a pu constater la mise en commun de données concernant la Covid-19 via des plateformes collaboratives créées au début de la crise du Covid ou des plateformes déjà existantes telle que Just One Giant Lab, dont l'utilisation a explosé avec l'arrivée de la pandémie.

Mais ce partage est encore jugé insuffisant, la communauté scientifique, désireuse d'une science plus transparente et partageuse, a lancé une pétition au début de la crise, réclamant une plus grande ouverture à l'Open Science. Démarche soutenue par les tenants de l'approche One Health qui affirme l'interdépendance des santés animales, humaines et des écosystèmes. La Covid 19 étant une zoonose, infection transmise par un animal dont, de surcroît, l'écosystème a été dégradé par l'activité humaine²⁸.

²⁷ McKiernan et al., « How open science helps researchers succeed ».

²⁸ « *L'Open Science* », mouvement qui préconise une science plus partageuse.

III - Santé

Utilisation en santé

Dans cette partie, nous allons parler des différentes utilisations de l'Open Data dans la santé car l'espace DATA permet la mise à disposition de tableaux de données dynamiques abordant les thèmes de la santé et de la protection sociale. De plus, grâce à l'Open Data nous pouvons suivre l'état des patients ainsi que la recherche médicale sur les maladies rares.

Un suivi personnel des patients

La loi de la santé du 26 janvier 2016 a organisé un Système National des Données de Santé (SNDS)²⁹ qui ouvre à un accès des différents de données, à la fois Big Data et Open Data. L'usage de ces données s'avère stratégique, notamment pour la protection des libertés et droits des personnes, pour la recherche biomédicale, pour la gestion des risques et pour le financement de la santé et de l'Assurance Maladie³⁰.

[C'est l'article 193 de la loi de modernisation de notre système de santé du 26 janvier 2016 qui instaure le Système national des données de santé (SNDS). Le Chapitre V de la loi s'intitule : Créer les conditions d'un accès ouvert aux données de santé.]

Ce système a pour objectif de favoriser les études, les recherches ou l'évaluation présentant un caractère d'intérêt public. Cette notion n'est pas définie par le texte fondateur du SNDS. Le législateur entend rendre accessible des données recueillies par les services publics et les différents organismes de l'assurance maladie. Et selon les travaux parlementaires, il faut y entendre un intérêt général ou un bénéfice collectif.

[Dossiers Solidarité et Santé n° 64, Données de santé anonymat et risques de ré-identification, juillet 2015.

Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, modifiée par la loi N°2018-493 du 20 juin 2018 relative à la protection des données personnelles.]

²⁹ « Open Data en santé ».

³⁰ « Qu'est-ce que le SNDS ? | SNDS ».

Créer de nouveaux écosystèmes de santé

La création du SNDS est instaurée par l'[article 193](#) de la loi de modernisation de notre système de santé de janvier 2016. Cet article décrit notamment³¹ :

- les composantes du SNDS ;
- les finalités autorisées de traitement des données issues du SNDS ;
- les conditions de sécurité de traitement de ces données.

Le but de cette création est de rassembler des bases de données en santé déjà existantes, afin de faciliter la gestion médicale du patient et améliorer la performance des soins et des traitements médicaux.

Contribuer à l'avancée de la recherche sur les maladies rares

La diversité des pathologies rares et leurs premiers symptômes - parfois bénins - rendent difficile l'établissement d'un diagnostic rapide et pertinent, ainsi que l'orientation du malade vers un organisme de soin spécialisé.

C'est quoi la Banque Nationale des Données de Maladies Rares (BNDMR) ?

C'est une base de donnée nationale sécurisée mise en place sous l'impulsion du ministre de la santé qui permet de rassembler un set de données médicales pour tous patients atteints de maladies rares dans le but de³² :

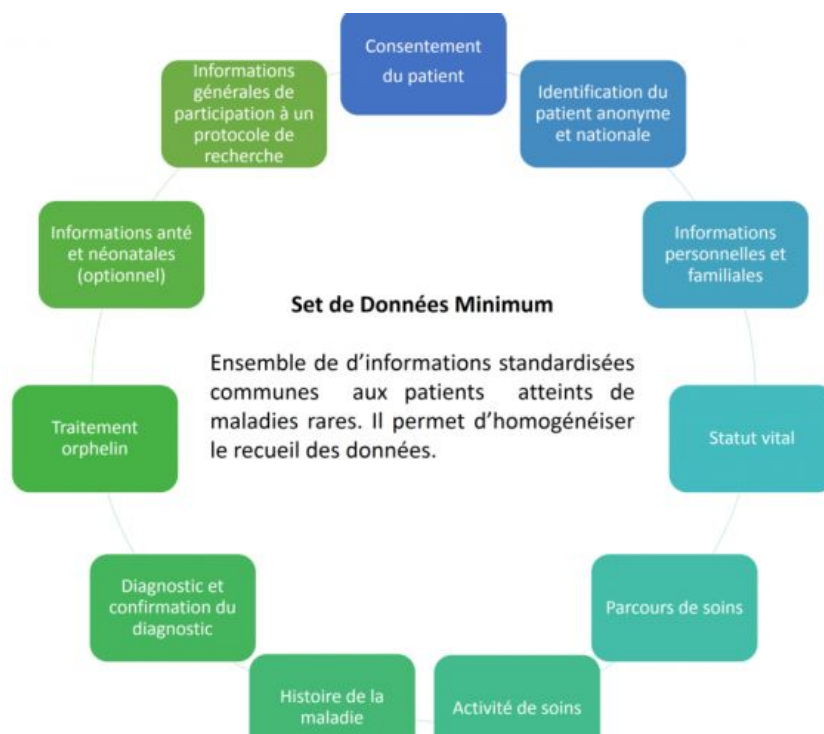
- Réaliser des analyses statistiques et études épidémiologiques (sur l'incidence et prévalence, le parcours de soins;)
- Suivre la file active
- Faciliter l'identification des patients éligibles pour la mise en œuvre d'essais cliniques
- Exploiter les données recueillies par l'hôpital (monocentrique), pour PIRAMIG notamment.

Pour chaque patient, un Set de Données Minimums (SDM) est collecté qui documente la prise en charge et l'état de santé des patients qui sont suivis dans les centres experts français³³.

³¹ LOI n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé (1).

³² « BNDMR ».

³³ « Le Set de Données Minimal National ».



L'accélération de la recherche sur le cancer

Récemment, on trouve des machines learning et deep learning qui aident à bien comprendre le cancer et à définir son évolution. Parlons maintenant de INCA, l'institut national du cancer, qui est une agence d'expertise sanitaire et scientifique en cancérologie afin de lutter contre le cancer. On trouve sur leur site une rubrique pour les données du cancer avec un accès direct et simple. Cette rubrique permet l'amélioration du partage et la valorisation des différentes informations.

Les informations mises à disposition, validées par les producteurs de données, présentent une vision transversale du cancer et de la cancérologie, en traitant de l'épidémiologie, des facteurs de risques, des dépistages, des soins, de la prise en charge du patient, en attachant une attention particulière à la vie après le cancer et de la recherche.

Ces indicateurs sont regroupés dans des fiches de synthèse, elles-mêmes rassemblées dans des thématiques et des sous-thématiques, afin de faciliter la recherche et la lisibilité de ces informations pour l'internaute. Les principales thématiques sont l'épidémiologie, la prévention et les facteurs de risques, le dépistage, les soins, l'après cancer et la recherche et les innovations en cancérologie.³⁴

³⁴ « Institut National du Cancer - data.gouv.fr ».

COVID

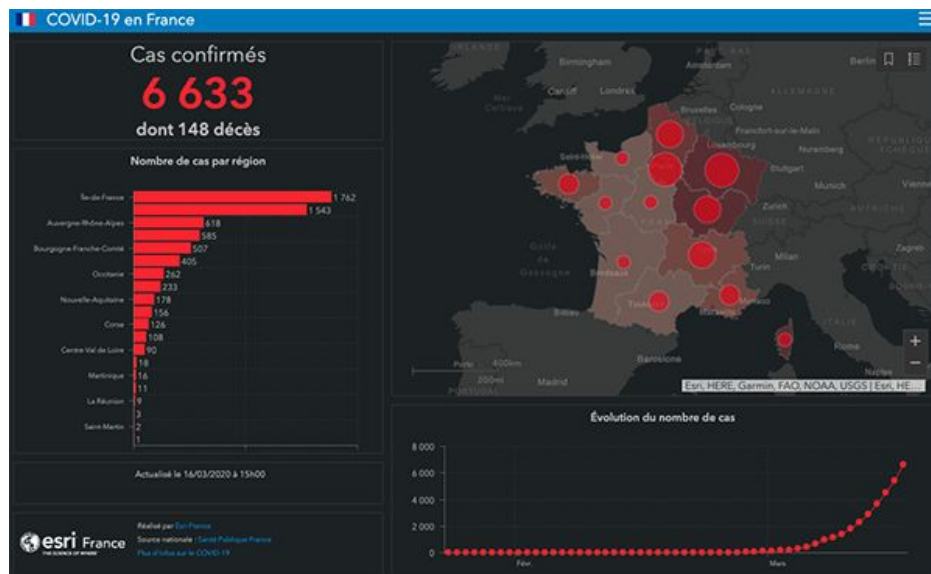
L'Open Data alliée essentielle dans la lutte contre le COVID-19 :

Depuis l'épidémie, l'accès à différentes informations reste un fait essentiel afin de lutter contre cette maladie.

Plusieurs informations sont fournies par les différents pays et disponibles pour le public :

- Nombre de cas par pays
- Nombre de personnes décédées
- Nombre de places disponibles dans les hôpitaux
- Flux des voyageurs
- Résultats des tests de vaccination
- ...

L'utilisation de l'ensemble de ces données et leur analyse est un atout important pour endiguer la maladie.

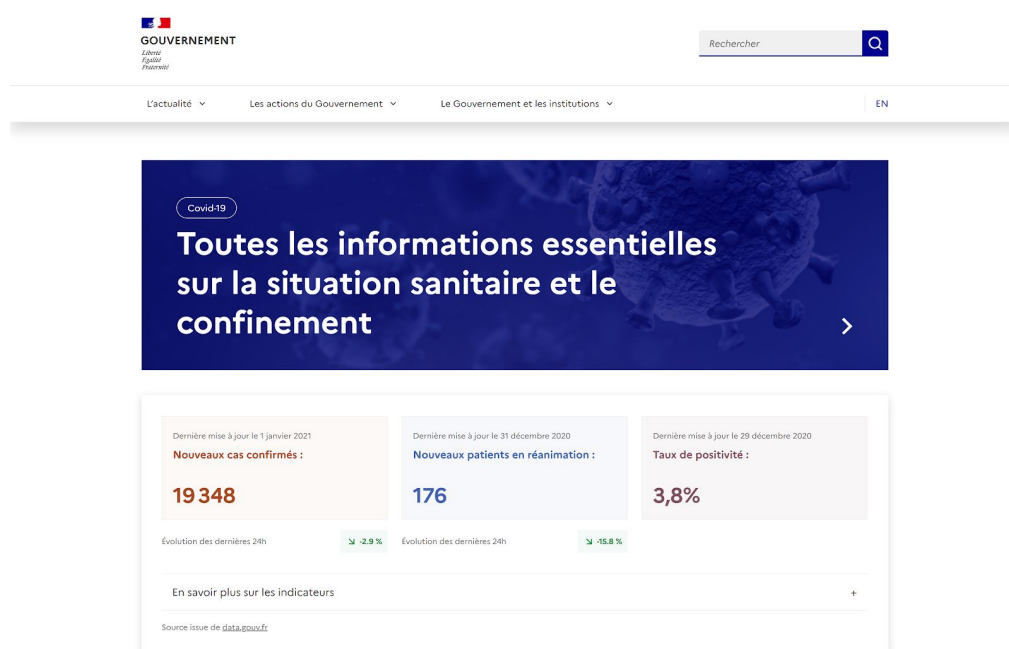


L'accès aux statistiques et aux données brutes sur l'épidémie de coronavirus est un enjeu important de la crise sanitaire actuelle, même si elles sont parfois imparfaites. Depuis plusieurs années, l'Open Data connaît un fort développement : institutions, entreprises et associations dans le monde entier mettent à disposition leurs données, de manière gratuite et ouverte. C'est ainsi que des collectivités locales diffusent l'état-civil ou le menu des cantines, que la SNCF rend publiques les fréquentations des lignes, ou que l'Union Européenne via son agence Eurostat compare, parmi des milliers d'indicateurs, le nombre d'accès à Internet par pays. De nombreux scientifiques partagent également leurs données de cette manière.

En pleine crise du coronavirus, l'Open Data s'est avérée un atout important pour comprendre la propagation du Covid-19 et prendre des décisions. C'est le constat que pose Jean-Marc Lazard, fondateur et président de la société Opendatasoft, qui est un acteur partie prenante de l'Open Data. Opendatasoft est le leader européen des plateformes de mise à disposition de données, et compte de nombreux clients publics et privés comme le ministère de l'éducation nationale, la SNCF, Infolocale, plusieurs villes comme Bruxelles et Rennes.

Pour Jean-Marc Lazard, la mission de l'Open Data est de faire en sorte que l'information circule et cela s'avère particulièrement précieux dans ce contexte de crise sanitaire, à plusieurs niveaux. L'ouverture et la diffusion des données sont utiles aux scientifiques en premier lieu. De la simple comparaison des courbes entre pays aux échanges pointus sur la structure du virus, le partage des connaissances fait avancer la recherche.

Un exemple d'Open Data : le tableau de bord de suivi de l'épidémie de coronavirus en France disponible sur gouvernement.fr présente les données relatives à l'épidémie de COVID-19 en France. Il a été mis en ligne le 28 mars 2020, afin de répondre aux besoins d'information des citoyens et des citoyennes en matière de transparence sur l'évolution de l'épidémie de coronavirus en France.



Ce tableau de bord, dont le code source est libre, a été développé sous l'impulsion d'Etalab et avec la collaboration de la société civile. Il propose une vision consolidée des données officielles disponibles. Plusieurs sources de données viennent alimenter le tableau de bord afin de permettre d'informer au mieux la population en temps réel.

Conclusion

En conclusion, les données ouvertes sont des données publiques disponibles, réutilisables, distribuables et dont la participation est universelle mais qui sont toutefois soumises à des lois. Pour ces données, l'interopérabilité est essentielle afin de tirer des bénéfices de cette ouverture.

Massivement utilisées par les gouvernements, on retrouve l'Open Data partout dans notre quotidien, notamment en santé. Ces données contribuent à la recherche et permettent même de l'accélérer, en santé on les retrouve aussi dans le suivi patient.

Dans le contexte actuel de la pandémie de Coronavirus Covid-19, l'Open Data s'érige en véritable alliée face au virus. Cependant, malgré tout le bénéfice qu'apportent les données ouvertes elles restent parfois imparfaites.

Bibliographie

1. Futura. « Open data ». Futura. Consulté le 28 décembre 2020. <https://www.futura-sciences.com/tech/definitions/internet-open-data-15502/>.
2. « What Is Open? » Consulté le 28 décembre 2020. <https://okfn.org>.
3. L, +Bastien. « Open Data définition : qu'est-ce que c'est ? À quoi ça sert ? » *LeBigData.fr* (blog), 5 novembre 2019. <https://www.lebigdata.fr/open-data-definition>.
4. « LA DEMARCHE ». Consulté le 28 décembre 2020. <http://opendata.onisep.fr/49-la-demarche.htm>.
5. « Données ouvertes ». In Wikipédia, 17 novembre 2020. https://fr.wikipedia.org/w/index.php?title=Donn%C3%A9es_ouvertes&oldid=176678198.
6. Team. « What Is Open Data? » Consulté le 28 décembre 2020. <https://www.opendatasoft.com/blog/2017/02/16/what-is-open-data>.
7. Huston, P, VI Edge, et E Bernier. « Reaping the Benefits of Open Data in Public Health ». *Canada Communicable Disease Report* 45, no 10 (3 octobre 2019): 252-56. <https://doi.org/10.14745/ccdr.v45i10a01>.
8. « Projet_Open_Datalab-quelles_obligations_rglementaires-v1.1.pdf ». Consulté le 26 décembre 2020. http://www.opendatalab.fr/images/doc/Projet_Open_Datalab-quelles_obligations_rglementaires-v1.1.pdf.
9. « Déclaration des Droits de l'Homme et du Citoyen de 1789 | Conseil constitutionnel ». Consulté le 26 décembre 2020. <https://www.conseil-constitutionnel.fr/le-bloc-de-constitutionnalite/declaration-des-droits-de-l-homme-et-du-citoyen-de-1789>.
10. « Guide pratique de la publication en ligne et de la réutilisation des données publiques ("Open Data") », s. d., 21.
11. « Les collectivités territoriales et l'open data : concilier ouverture des données et protection des données personnelles | CNIL ». Consulté le 26 décembre 2020. <https://www.cnil.fr/fr/les-collectivites-territoriales-et-lopen-data-concilier-ouverture-des-donnees-et-protection-des>.
12. « Chronologie de l'open data | guides.etalab.gouv.fr ». Consulté le 26 décembre 2020. <https://guides.etalab.gouv.fr/juridique/chronologie/>.

13. Décret n° 2011-194 du 21 février 2011 portant création d'une mission « Etalab » chargée de la création d'un portail unique interministériel des données publiques, 2011-194 § (2011).
14. LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1), 2016-1321 § (2016).
15. « Chapitre Ier : Etendue du droit de réutilisation (Articles L321-1 à L321-3) - Légifrance ». Consulté le 26 décembre 2020. <https://www.legifrance.gouv.fr/codes/id/LEGIARTI000033218992/2016-10-09/>.
16. Directive 2007/2/CE du Parlement européen et du Conseil du 14 mars 2007 établissant une infrastructure d'information géographique dans la Communauté européenne (INSPIRE), Pub. L. No. 32007L0002, OJ L 108 (2007). <http://data.europa.eu/eli/dir/2007/2/oj/fra>.
17. Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public, Pub. L. No. 32019L1024, OJ L 172 (2019). <http://data.europa.eu/eli/dir/2019/1024/oj/fra>.
18. « ETALAB-Licence-Ouverte-v2.0.pdf ». Consulté le 26 décembre 2020. <https://www.etalab.gouv.fr/wp-content/uploads/2017/04/ETALAB-Licence-Ouverte-v2.0.pdf>.
19. Le blog d'Etalab. « Licence Ouverte / Open Licence ». Consulté le 26 décembre 2020. <https://www.etalab.gouv.fr/licence-ouverte-open-licence>.
20. « ODC Open Database License v1.0 | Software Package Data Exchange (SPDX) ». Consulté le 26 décembre 2020. <https://spdx.org/licenses/ODbL-1.0.html>
21. D'Agostino, Marcelo, Noah O. Samuel, Maria Janina Sarol, Federico G. de Cosio, Myrna Marti, Tianyu Luo, Ian Brooks, et Marcos Espinal. « Open data and public health ». *Revista Panamericana de Salud Pública* 42 (2018). <https://doi.org/10.26633/RPSP.2018.66>.
22. Gouvernement.fr. « L'ouverture des données publiques ». Consulté le 28 décembre 2020. <https://www.gouvernement.fr/action/l-ouverture-des-donnees-publiques>.
23. Ruijter, Erna, Stephan Grimmelikhuijsen, et Albert Meijer. « Open Data for Democracy: Developing a Theoretical Framework for Open Data Use ». *Government Information Quarterly* 34, no 1 (janvier 2017): 45-52. <https://doi.org/10.1016/j.giq.2017.01.001>.
24. « Methodology - Global Open Data Index ». Consulté le 27 décembre 2020. <https://index.okfn.org/methodology/>.
25. « Open Government Data - OECD ». Consulté le 27 décembre 2020.

<http://www.oecd.org/gov/digital-government/open-government-data.htm>.

26. « ODB-4thEdition-Methodology.pdf ». Consulté le 27 décembre 2020.
<https://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-Methodology.pdf>.
27. McKiernan, Erin C, Philip E Bourne, C Titus Brown, Stuart Buck, Amye Kenall, Jennifer Lin, Damon McDougall, et al. « How open science helps researchers succeed ». Édité par Peter Rodgers. eLife 5 (7 juillet 2016): e16800. <https://doi.org/10.7554/eLife.16800>.
28. « L'Open Science », mouvement qui préconise une science plus partageuse. Consulté le 2 janvier 2021.
<https://www.nova.fr/podcast/tech-paf/open-science-mouvement-qui-preconise-une-science-plus-partageuse>.
29. « Open Data en santé : où en est-on ? - mc2i : Conseil en transformation numérique ». Consulté le 2 janvier 2021.
<http://www.mc2i.fr/Open-Data-en-sante-ou-en-est-on>.
30. « Qu'est-ce que le SNDS ? | SNDS ». Consulté le 2 janvier 2021.
<https://www.snds.gouv.fr/SNDS/Qu-est-ce-que-le-SNDS>.
31. LOI n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé (1), 2016-41 § (2016).
32. « BNDMR : Base Nationale de Données Maladies Rares ». Consulté le 2 janvier 2021.
<https://www.filfoie.com/recherche/bndmr-base-nationale-de-donnees-maladies-rares-2/>.
33. Banque Nationale de Données Maladies Rares. « Le Set de Données Minimal National », 13 février 2014.
<https://www.bndmr.fr/le-set-de-donnees-minimal/>