

Ma fresque idéale

Qu'est ce que l'open data ?

A quoi ça sert ?

Comment on fait ?

Ma fresque idéale (libre)

Lexique

- **Donnée** : Un élément brut (ex : -18°C), auquel on ajoute du sens pour obtenir une information (il fait froid), qui apporte une connaissance (à -18°C je vais avoir des engelures), grâce à laquelle j'aurais une compétence pour prendre une décision (j'enfile des gants).
- **Donnée ouverte** : Une donnée, souvent produite par des acteurs publics, librement accessible, modifiables, réutilisable dans un langage accessible aux ordinateurs
- **Code source** : Ensemble d'instructions écrites dans un langage de programmation informatique. Il permet d'écrire des programmes destinés aux ordinateurs.
- **Algorithme** : Description d'une suite d'étapes (par ex: un calcul) permettant d'obtenir un résultat à partir d'éléments fournis (données) en entrée et des différentes règles (critères d'attribution).
- **Droit d'accès** : Pour que le droit d'accès s'exerce, il faut que le document demandé soit de nature administrative et que l'administration l'a bien en sa possession. Il s'applique également aux documents reçus par les administrations de personnes privées.
- **Les exclusions au droit d'accès** : Sont exclus du droit d'accès les documents relevant du secret défense, secret médical, secret des affaires, de la protection de la vie privée, s'il porte préjudice à une personne physique par jugement de valeur ou appréciation. Il convient alors d'occulter les mentions protégées.
- **CRPA** : Code des relations entre le public et l'administration. Il régit les relations entre le public et l'administration.

Lexique

- **CADA** : Commission d'Accès aux Documents Administratifs (1978). Elle veille à la liberté d'accès aux documents administratifs et aux archives publiques. Elle peut être saisie par les personnes physiques ou morales qui ont eu un refus d'accès à des documents auprès de l'administration sollicités initialement.
- **LRN** : Loi pour une République Numérique. Dite loi Lemaire (2016). Elle introduit l'ouverture par défaut des données publiques pour les administrations d'État, les collectivités locales de plus de 3 500 habitants, les établissements publics et les organismes privés chargés d'un service public, à l'exception des entités de moins de 50 agents ou salariés.
- **RGPD** : Règlement Général sur la Protection des Données (2016). Règlement européen qui encadre la collecte, le traitement, l'utilisation et la circulation des données à caractère personnel (DCP). Il s'applique aux organismes privés et publics.
- **Donnée à caractère personnel** : Une donnée personnelle est "toute information se rapportant à une personne physique identifiée ou identifiable directement ou indirectement" (CNIL). Ex : nom, prénom, identifiant, biométrie, génétique, voix, image,...
- **DPO** : Délégué-e à la protection des données. Personne en charge d'appliquer les procédures pour être en accord avec le RGPD au sein de sa structure/son entreprise. Cette personne doit avoir les moyens matériels et organisationnels, les ressources et le positionnement adéquat pour mener à bien cette charge.
- **Anonymisation** : Traitement des données personnelles qui vise à rendre impossible toute réidentification de la personne. Elle est irréversible (technique destructive d'informations). Elle permet de conserver les données sans limitation.
- **Pseudonymisation** : Traitement des données personnelles qui vise à ne plus attribuer les données à une personne physique identifiée. Elle consiste à remplacer les données directement identifiantes (ex : nom, prénom) par des données indirectement identifiantes (un alias, un numéro). Elle est réversible.

Glossaire

- **Déclaration des droits de l'homme et du citoyen** : L'article 15 de la Déclaration des Droits de l'Homme et du Citoyen de 1789 prévoit que « la Société a le droit de demander compte à tout Agent public de son administration. » Il a constitué le fondement de la transparence administrative et a conduit à la création d'institutions comme les Archives Nationales ou la Cour des Comptes.
- **Mise en données du monde** : La mise en données du monde ou "datafication" désigne le phénomène dans lequel presque tous les aspects du monde et de la vie humaine laissent une trace sous la forme de données numériques. Capteurs, smartphones, procédures numériques : le volume de données a explosé au cours des 20 dernières années.
- **Statistique** : La statistique désigne la discipline qui étudie des phénomènes à travers la collecte, le traitement et l'analyse de données. À la fois outil de savoir et outil de pouvoir selon le sociologue Alain Desrosières, elle a permis d'exprimer sous une forme numérique ce qui, auparavant, était exprimé par des mots. La statistique repose aujourd'hui en France sur l'Insee et les 16 services statistiques ministériels (SSM).
- **Données publiques ouvertes par défaut** : Le premier principe de la charte internationale de l'open data demande que l'ouverture devienne la norme : les données publiques devraient être ouvertes, sauf exception légale, sans que le public ait à les demander. En France, ce principe s'est traduit dans la Loi pour une République Numérique de 2016 pour toutes les administrations et collectivités de plus de 3500 habitants et 50 agents.
- **Données à jour et complètes** : Le deuxième principe de la charte internationale de l'open data réclame que les données soient publiées sous leur forme complète, dans le plus fort niveau de détail possible. Elles doivent être mises à disposition dès que possible et, lorsque c'est pertinent, en temps réel.

Glossaire

- **Données accessibles et utilisables** : Le troisième principe de la charte internationale de l'open data demande que les données soient exploitables par les machines afin de faciliter leur traitement. Les données doivent être dans un format ouvert, qui n'est pas la propriété d'une organisation et fait l'objet d'une gouvernance ouverte. Enfin, les données doivent être faciles à trouver sur des portails et mises à disposition avec une licence ouverte autorisant largement les usages les plus larges.
- **Des données comparables et interopérables** : Le quatrième principe de la charte internationale de l'open data demande que les données soient facilement comparables entre les territoires et à travers le temps. Concrètement, la charte encourage la documentation des données et incite à l'usage de standards de données qui normalisent les champs et les valeurs de chaque fichier décrivant le même objet.
- **Pour une gouvernance et un engagement citoyen** : Dans son cinquième principe, la charte internationale défend l'idée que l'open data améliore la gouvernance des institutions et favorise la participation citoyenne. La reddition des comptes avec des données complètes et détaillées permet aux citoyens de refaire les calculs, de proposer d'autres manières d'analyser un phénomène et parfois de contester une décision publique.
- **Pour un développement inclusif et l'innovation** : Dans son sixième principe, la charte internationale de l'open data suggère que la société civile et le secteur privé peuvent libérer la créativité dès lors que les données sont ouvertes. L'innovation par la réutilisation des données ouvertes jouera un rôle important face aux défis du dérèglement climatique et de la montée des inégalités.
- **Pour la modernisation** : Les politiques d'open data reposent aussi sur leur capacité à provoquer des changements dans le fonctionnement de l'administration : transversalité, désilotage, innovation ouverte... En France, le Ministère de la transformation et de la fonction publique assure la mise en œuvre de l'ouverture des données et l'envisage comme un moyen de faire émerger un service public "augmenté" par la collaboration avec la société civile.

Glossaire

- **Suivi de la pandémie** : Les médiateurs de données ont joué un rôle important dans le suivi de la pandémie. Par ex., le site CovidTracker, développé par Guillaume Rozier, a aidé les citoyens à suivre l'épidémie de Covid-19 avec des tableaux de bord interactifs et simples à comprendre et à utiliser. Fin 2020, le site a enregistré près de 15 millions de visites mensuelles. D'autres médiateurs comme Météo Covid ou Our World in Data ont aidé massivement le public à comprendre les données.
- **Suivre les finances locales** : La publication par les collectivités territoriales de leurs données financières (budget, subventions attribuées...) répond au besoin de transparence de la vie publique. Depuis février 2020, l'Observatoire des Finances et de la Gestion publique Locales (OFGL) met en open data des données relatives aux finances et à la gestion des collectivités locales françaises.
- **Détection des entreprises en difficulté** : Signaux Faibles est une start-up d'État qui utilise des données publiques et privées afin de détecter les signes avant-coureurs de difficultés des entreprises. Le but est d'intervenir au plus vite pour soutenir l'entreprise. Grâce à des algorithmes, les agents peuvent plus facilement et rapidement identifier les entreprises en difficulté.
- **Transparence des produits alimentaires** : Open Food Facts est une base de données collaborative sur les produits alimentaires née en France en 2012. 9000 contributeurs dans 200 pays y ont renseigné des données sur plus de 2,5 millions de produits. En plus de l'application mobile Open Food Facts, ces données ont aussi servi à la création de Yuka, utilisé par 21 millions de personnes, et à concevoir le système d'étiquetage nutritionnel Nutriscore.
- **Planification des itinéraires en transports** : L'impact de l'open data est particulièrement fort dans le domaine de la planification d'itinéraires en transports publics. Outre le géant Google Maps, plusieurs applications utilisent les données ouvertes et standardisées sur les horaires de transport. On peut citer les applications CityMapper, ou encore Moovit récemment rachetée par Intel pour 900 millions de dollars qui revendique 1.5 milliard d'utilisateurs dans 3500 villes réparties dans 112 pays.

Glossaire

- **Carte de la pollution de l'air près des écoles** : Les données ouvertes facilitent la mobilisation citoyenne. L'association Respire en Ile-de-France a créé en 2018 une carte de la pollution de l'air extérieur aux abords des écoles, crèches et halte garderie. Combinant la géolocalisation des établissements et des mesures de la qualité de l'air, l'association a dénoncé des niveaux très préoccupants de pollution à proximité des grands axes.
- **Enquêtes journalistiques** : Le data journalism est un mouvement qui renouvelle le journalisme par l'analyse et la visualisation de données. Par exemple, des journalistes de FranceInfo ont montré que, malgré la parité dans les investitures du RN aux législatives de 2022 (40% de femmes), seulement 35 femmes étaient investies dans les 100 circonscriptions "gagnables". Résultat : il y a 37% de femmes dans le groupe RN à l'Assemblée.
- **Licences homologuées** : La loi pour une République numérique a limité par décret les licences (texte juridique qui définit les droits et devoirs des usagers) :
 - la Licence Ouverte 2.0 d'Etalab qui impose la mention de la source et de la date de mise à jour
 - la licence ODbL (Open Database License) qui impose, en plus, le partage des modifications avec la même licence dans une logique de "pot commun"
- **Droit d'accès aux documents administratifs** : Prévu par loi CADA de 1978, aujourd'hui codifié dans le code des relations entre le public et l'administration, le droit d'accès concerne tous les documents sous toute forme produits ou reçus dans le cadre d'une mission de service public. Les personnes publiques ou privées sont tenues de les communiquer au public sauf si cela porte atteinte à un secret protégé. L'open data consiste à publier de manière volontaire les données sans attendre une requête.
- **Portage politique par les élus** : Le rapport au Premier ministre du député Bothorel sur la politique de la donnée a montré que les politiques liées aux données manquent souvent de portage hiérarchique dans les directions métiers. Les cadres de la fonction publique ne sont pas assez informés sur le potentiel des données publiques et les moyens à y consacrer. Dans les collectivités, le portage par les élus est une condition essentielle de réussite des projets d'open data.

Glossaire

- **Format ouvert** : La loi pour une République Numérique consacre l'utilisation de formats ouverts afin de faciliter la réutilisation des données par les machines. L'administration est dorénavant tenue, lorsqu'elle communique un document administratif de le mettre à disposition du citoyen "dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé".
- **Service public de la donnée** : Créé par la loi pour une République numérique, le service public de la donnée vise à faciliter la réutilisation des jeux de données de référence à plus fort impact économique et social (ex. : Base Sirene, Base Adresse Nationale, Code Officiel Géographique, Plan cadastral informatisé...) Les producteurs de ces données s'engagent sur la qualité, la disponibilité, la mise à jour et le traitement des signalements des usagers.
- **Secret statistique** : Le secret statistique interdit la communication de données issues de traitements à finalités statistique dès lors qu'elles sont individuelles ou susceptibles d'identifier des personnes. Le secret statistique est opposable à toute réquisition judiciaire et toute utilisation des données à des fins de décision à l'égard des personnes concernées comme le contrôle fiscal.
- **Secrets protégés** : Certains documents administratifs (et donc des données) ne peuvent être communiqués lorsque leur contenu porte atteinte aux secrets suivants définis par la loi : défense nationale, conduite de la politique extérieure, monnaie et le crédit public, secret médical, vie privée, secrets industriels et commerciaux, secret des affaires...
- **Anonymisation des données** : Le droit d'accès ne s'applique pas pour les informations à caractère personnel relevant de la vie privée. Il faut anonymiser les données et s'assurer que les individus ne sont pas réidentifiables par croisement ou caractéristiques trop spécifiques. Leur ouverture sans anonymisation n'est permise que si une disposition législative le prévoit, si les personnes y ont consenti ou si les documents relèvent des catégories prévues par la loi (article D312-3-1 du CRPA). Pour aller plus loin, voir le RGPD, Règlement général sur la protection des données.

Glossaire

- **Données d'intérêt général** : La loi pour une République numérique a introduit la notion de données d'intérêt général. Elle désigne des données détenues par des entreprises qui sont particulièrement importantes pour l'action publique. Cette loi contraint les délégataires d'un service public qui doivent fournir les données aux pouvoirs publics. Sous certaines conditions, le MEFR peut aussi imposer aux entreprises la transmission de certaines données à des fins d'enquêtes statistiques.
- **Données géographiques** : "La directive Inspire de 2007 impose aux autorités publiques de recenser les données géographiques, de créer des métadonnées normées et de suivre des standards européens. Le rapport Faure-Muntian en 2018 a rappelé l'importance de données géographiques de qualité pour que l'Etat exerce ses missions et sa souveraineté en matière de défense, de sécurité, de fiscalité, de protection de l'environnement ou de transport."
- **Documentation des données** : L'insuffisance de la documentation est un défaut majeur des données ouvertes. Sur data.gouv.fr, 22 % des jeux de données ont une description de moins de 180 caractères (soit un SMS) et 59 % moins de 1 000 caractères (une demi-page). C'est généralement insuffisant pour que les réutilisateurs comprennent le contenu et les limites des données.
- **Acculturation interne et externe** : L'utilité des données est mal comprise, et reste identifiée comme un sujet purement "technique". L'acculturation des citoyens et des agents est un élément indispensable dans une stratégie data. Pour y répondre, le mouvement de la littératie de données considère que la compréhension et l'usage des données doivent devenir une compétence accessible à tous comme la lecture et l'écriture.
- **Remontée du cycle de vie de la donnée** : Il est indispensable de prendre en compte les pratiques des réutilisateurs en amont de la production des jeux de données. Pour ce faire, une réflexion sur leur structure, sur le format des fichiers ou sur la documentation doit être menée systématiquement pour faciliter l'appropriation des données par des acteurs tiers.

Glossaire

- **Périmètre de l'ouverture des données par défaut** : La Loi pour une République Numérique impose un principe d'ouverture des données par défaut à toutes les administrations de plus de 50 agents et, pour les collectivités locales, de plus de 3500 habitants. Il concerne un périmètre très large comprenant les documents communiqués suite à des demandes CADA, les "bases de données" et les données "dont la publication présente un intérêt économique, social, sanitaire ou environnemental".
- **Données brutes** : Il est essentiel de d'ouvrir les données brutes, c'est-à-dire qui n'ont pas été modifiées, agrégées ou manipulées. Cela peut demander un travail de "brutification des données" pour les rendre lisibles par des machines. Aux réutilisateurs après, d'effectuer les calculs et agrégations qu'ils souhaitent pour les rendre intelligibles par des humains. Par exemple, les fichiers Excel avec des lignes fusionnées ou des totaux devront être transformés en fichiers .csv structurés, avec un enregistrement par ligne.
- **Communauté(s) Open Data** : Animer une communauté de réutilisateurs (intégrer les commentaires des réutilisateurs dans les processus métiers, organiser des ateliers, ...) est important. Cela peut aussi prendre la forme d'une communauté interne qui permet aux personnels en charge de l'open data de rencontrer leurs pairs et de valoriser leur travail. Il existe une communauté française de l'Open Data. Le forum #TeamOpenData est un des lieux d'échange de celle-ci.
- **Socle Commun des données locales** : Le socle commun des données locales est une liste des données normalisées que les collectivités territoriales doivent ouvrir en priorité (délibérations, subventions, menus des cantines...). Ces données sont adossées à des schémas de données qui décrivent les champs et valeurs de chaque fichier. Favorisant la réutilisation des données et leur interopérabilité, ces schémas sont consultables sur le portail national schema.data.gouv.fr.
- **Portail open data** : Un portail open data est une plateforme technique sur laquelle une administration publie ses données. Elle permet de présenter les données, les mettre en scène dans des visualisations, les partager dans des formats ouverts. En France, le portail data.gouv.fr centralise les données produites par l'administration, mais aussi par le secteur privé lorsqu'elles sont d'intérêt public. Parfois, un portail peut aussi servir à partager des données en interne à la collectivité de manière fermée.

Glossaire

- **Données essentielles de la commande publique (DECP)** : Les données essentielles de la commande publique (DECP) sont les données relatives aux marchés publics et contrats de concession. Elles font l'objet d'une obligation de publication depuis le 1er octobre 2018. La liste des données à publier fait l'objet d'un arrêté spécifique. Ces données peuvent, par exemple, être utilisées pour piloter la politique achat des collectivités (exemple : suivre le nombre de marchés attribués à des TPE et PME locales).
- **Loi d'orientation des mobilités (LOM)** : La loi d'orientation des mobilités (LOM) de 2019 accorde de nouvelles compétences aux collectivités territoriales pour organiser les mobilités sur leur territoire. Elle prévoit l'ouverture des données statiques (arrêts, horaires, tarifs...) et en temps réel (perturbations, disponibilités...) des transports en commun ou à la demande, ainsi que les données des réseaux routiers et de stationnements.
- **Territoires connectés** : Un "territoire intelligent" est un territoire dans lequel, à travers différents outils numériques (dont de nombreux capteurs), des services publics et des politiques publiques sont pilotés par la donnée. Ces évolutions concernent la mobilité (intermodalité, véhicule autonome), la gestion des déchets (optimisation des tournées), le patrimoine (bâtiment connecté), l'énergie ou encore la gestion des risques.
- **Administrateur/administratrice des données** : Si la désignation d'un-e administrateur-trice général-e de la donnée n'est pas obligatoire, plusieurs grandes collectivités en ont nommées. Cette personne conçoit et déploie la gouvernance et la stratégie de la donnée dans l'institution territoriale en coordination avec les différentes directions de la collectivités. Elle anime aussi l'éco-système local de la donnée.
- **Aide à la décision** : L'aide à la décision est l'ensemble des techniques permettant d'effectuer la meilleure prise de décision possible. Différents outils numériques peuvent faciliter la prise de décision comme ceux d'un SI décisionnel. Celui-ci permet de collecter, renforcer, stocker, agréger et restituer les données sur une politique publique et d'éclairer sa mise en œuvre.

Les points à retenir



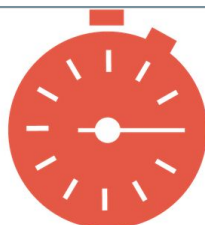
Origines

Déclaration des droits
de l'homme et du
citoyen



Principes

Données publiques
ouvertes par défaut



Principes

Données à jour et
complètes



Principes

Des données
comparables et
interopérables



Principes

Données accessibles et
utilisables



Bénéfices

Pour un
développement inclusif
et l'innovation

Les points à retenir



Bénéfices

Pour une gouvernance
et un engagement
citoyen



Cadre juridique

Périmètre de
l'ouverture des
données par défaut



Cadre juridique

Format ouvert



Bonnes pratiques

Données brutes



Bonnes pratiques

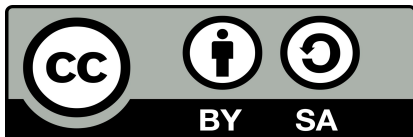
Acculturation interne et
externe



Bénéfices

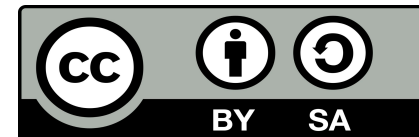
Pour la modernisation

DATAACTIVIST



La fresque des données ouvertes

DATAACTIVIST



La fresque des données ouvertes