

# **Défis de la réutilisation des données ouvertes et stratégies alternatives**

**Séminaire Méthodologie de  
Sciences Po**

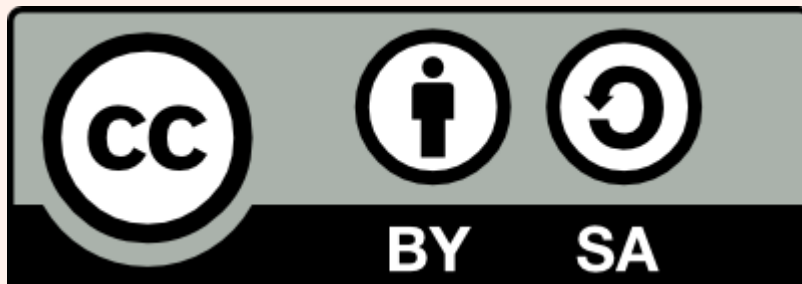
**Samuel Goëta, Dataactivist / Sciences  
Po Aix**

**2020-10-21**

Ces slides en ligne : <http://dataactivist.coop/metsem>

Sources : <https://github.com/dataactivist/metsem>

Les productions de Dataactivist sont librement réutilisables selon les termes de la licence [Creative Commons 4.0 BY-SA](https://creativecommons.org/licenses/by-sa/4.0/).



# Objectifs du séminaire

- Partager mon expérience issue de mon travail avec des réutilisateurs et des producteurs
- Présenter les principales difficultés de la réutilisation des données ouvertes
- Découvrir des stratégies alternatives pour obtenir des données quand elles n'existent pas, ne sont pas ouvertes ou ne sont pas exploitables

# Quelques mots sur Dataactivist

- Dataactivist est une société coopérative **pure player de l'open data** créée en 2016, par Samuel Goëta et Joël Gombin.
- Se positionnant sur **toutes les étapes du travail d'ouverture des données**, Dataactivist travaille tant avec les producteurs de données qu'avec les réutilisateurs et participe à l'appropriation des données par chacun.
- Nous appliquons nos propres **valeurs** : nous sommes une coopérative ; nos supports de formation et nos contenus sont librement réutilisables, publiés en licence Creative Commons.
- Une approche issue de la recherche : voir notamment **<https://dataactivist.coop/these>**.
- Nous animons la communauté **#TeamOpenData**.

# Quelques mots sur Dataactivist

Nos métiers :

- Conseil sur les **stratégies d'ouverture de données** : nous aidons les organisations dans la conception et la mise en œuvre de leur stratégie d'ouverture de données.
- **Accompagnement dans la réutilisation de données ouvertes** : nous aidons les organisations à utiliser les données au quotidien.
- **Sensibilisation et formation à la donnée** : nous formons à la culture générale des données, nous enseignons les grands principes et bonnes pratiques de l'open data et nous introduisons à la data science.
- **Médiation de données** : nous organisons des hackathons, des open data camps, des expéditions de données...

# Dataactivist : une entreprise née de la recherche

TheseSG\_finalHAL.pdf (page 232 sur 266) v

données ouvertes sans compétences préalables. C'est aussi le sens de mon prochain projet qui donnera une suite pratique à cette thèse. Avec Joël Gombin, politologue spécialiste de l'utilisation de données électorales, nous montons dataactivi.st, une société coopérative qui proposera des activités de conseil, de formation et de médiation sur l'ouverture des données et qui mettra en pratique ce qu'elle prêche en publiant sous licence libre le contenu des formations qu'elle produit. À travers dataactivi.st, nous allons essayer de nous adresser en particulier à des associations et des militants pour voir quelles formes pourraient prendre le « dataactivisme », à travers par exemple la critique des données publiées, la production de nouvelles en réponse aux données ouvertes de l'administration ou le militantisme pour l'ouverture de données spécifiques. Par rapport aux pratiques « statactivistes » (Bruno, Didier & Prévieux, 2014), nous voulons essayer de comprendre ce qui change lorsque des militants sont équipés de données. Nous nous lançons officiellement le 23 septembre 2016 à Aix-en-Provence en marge de la Data Literacy Conference organisée par la FING, une conférence qui part du principe que « les données sont désormais une affaire trop importante pour être laissées entre les mains des spécialistes et les considère comme un élément de la "littératie" qui, au même titre que la lecture et l'écriture, peut être accessible

Extrait de la conclusion de ma thèse : <https://dataactivist.coop/these>

# Open data : quelques éléments de contexte

# L'open data : une notion aux multiples facettes

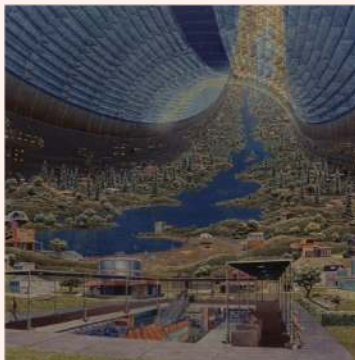


XV.  
La société a le droit de demander compte à tout agent public de son administration.

La transparence



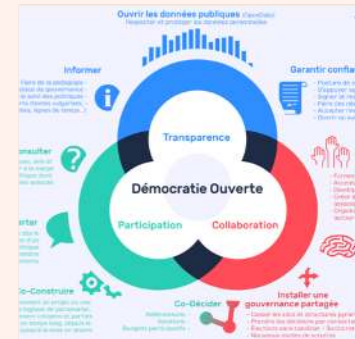
Le partage des données en science



La libre circulation de l'information

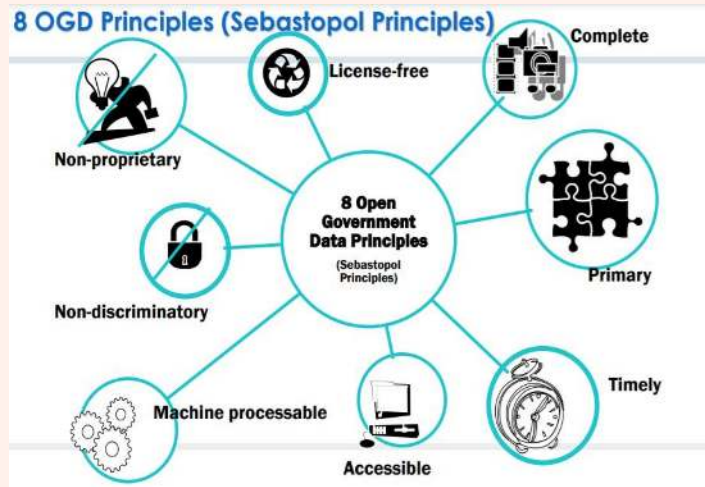


Les industries de la donnée



La transformation des administrations

# Des principes techniques qui se sont imposés



"Un régime fermé pourrait divulguer de grandes quantités de données conformes à ces huit exigences [...] Une diffusion électronique des déclarations de propagande faites par les dirigeants politiques de la Corée du Nord, par exemple, pourrait satisfaire à ces huit exigences."

Yu & Robinson (2012)  
« **The New Ambiguity of Open Government** »

# Loi pour une République Numérique : l'ouverture des données par défaut

Promulguée le 7 octobre 2016, après 4 ans d'annonces et de consultation, le titre 1 de la **loi pour une République Numérique** porte sur l'ouverture des données publiques.



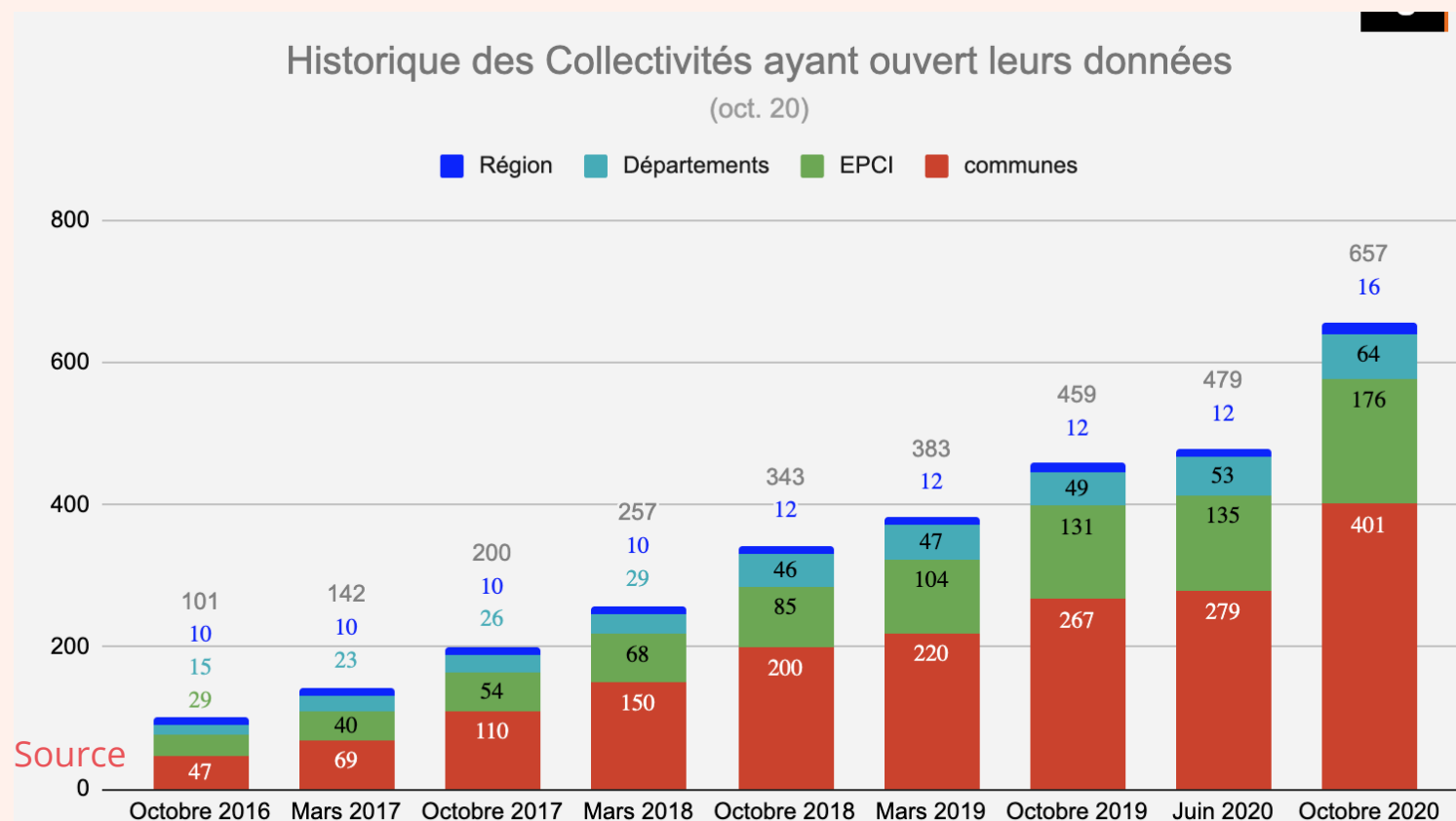
La loi impose un principe d'**ouverture des données par principe** qui ne fait pas l'objet de sanctions à toutes les administrations, les entreprises délégataires d'une mission de service public et les **collectivités locales de plus de 3500 habitants et 50 agents**.

L'ouverture concerne un **périmètre très large** :

- les documents communiqués suite à des demandes CADA
- les "bases de données"
- les données "dont la publication présente un intérêt économique, social, sanitaire ou environnemental". Rares sont donc les données publiques qui ne sont pas concernées par cette obligation d'ouverture.

# Une loi qui peine à être appliquée

Selon l'observatoire open data des territoires, seulement 10% des collectivités concernées par la loi pour une République numérique ont ouvert au moins un jeu de données.



# Le long chemin de l'ouverture d'un jeu de données



- Un travail invisible
- Un travail rarement pris en compte dans les missions des agents
- Opérer une "inversion" de l'infrastructure (Bowker, 1994 ; Star, 1999)

# 4 défis de la réutilisation des données ouvertes

# Défi 1 : la découvrabilité des données

**Découvrabilité** : capacité d'un contenu à sortir du lot

ou plutôt la **trouvabilité** : capacité pour un usager de tomber sur le bon contenu

La découverte des données est une condition préalable pour que les données ouvertes atteignent leur potentiel et la plupart des données sont actuellement difficiles à trouver.



<https://index.okfn.org/insights/>

# Un exemple de problème de découvrabilité

The image shows a screenshot of a tweet from Jules Grandin (@JulesGrandin) dated November 29, 2017. The tweet discusses the difficulty of finding data on the RER (Parisian regional rail network) and includes a screenshot of the 'Data.gouv.fr' website. The website screenshot shows a search for 'comptage-voiture' (vehicle counting) and a list of results, including 'comptage-voiture' and 'comptage-voiture-2017'. The tweet also shows engagement metrics: 58 retweets and 156 likes.

**Samuel Goëta** @samgoeta

**Jules Grandin** @JulesGrandin **Abonné**

Rechercher des données : une déception en 4 actes

Comptage voyageurs rER

Textes Images Maps Vidéos

Résultats (0,26 secondes)

comptage-voiture sur RER - Data.gouv.fr  
gouv.fr/fr/datasets/comptage-voiture-2017  
contient le comptage voyageur effectué

12:50 - 29 nov. 2017

58 Retweets 156 J'aime

7 58 156

Angellina Jolie 21 Tweets  
Bertrand Cantat 1243 Tweets  
David Silva 1309 Tweets  
Affaire Maelys 1416 Tweets  
Hit Sale

Éditer le profil

# Un exemple de données introuvables

Ce jeu de données contient plus de 6000 marchés publics attribués par l'Etat. Mais, malgré mon **signalement**, tout est fait pour le rendre introuvable !

The screenshot shows the data.gouv.fr interface. At the top, there is a search bar with the text 'Recherche' and a 'Thématiques' dropdown menu. Below the search bar, the dataset title 'AIFE\_DE\_44785462100045' is displayed. A note states 'Ce jeu de données provient d'un service public certifié' and 'creation catalogue de donnees essentielles pour la place de marché:44785462100045'. A 'Ressources' section lists four XML files with their sizes and 'TÉLÉCHARGER' buttons. On the right, a 'Producteur' section identifies 'aife' as the 'Agence pour l'Informatique Financière de l'Etat' and provides contact options like 'VOIR LE PROFIL', 'CONTACTER', and 'SUIVRE'. An 'Informations' section lists the license as 'License Not Specified' and dates from 2018 to 2019.

# Autre exemple de problèmes de découvrabilité

Sur data.gouv.fr, une requête "coronavirus" ne renvoie vers aucun jeu de données publié par Santé Publique France. Il faut rechercher "covid" pour les trouver, du nom de la maladie causée par le SARS COV-2.

The screenshot shows the data.gouv.fr search results for the query "coronavirus". The page header includes the French Republic logo and the text "data.gouv.fr" and "Plateforme ouverte". Below the header, there are navigation links: "Données", "Réutilisations", "Organisations", "Tableau de bord", and "Documentation". A search bar contains the text "coronavirus". Below the search bar, there are three tabs: "27 jeux de données", "54 réutilisations", and "2 organisations". The search results are displayed in a list format. The first result is titled "Données relatives à l'épidémie de COVID-19 en France : vue d'ensemble" and is published by "etalab". The second result is titled "CoronaVirus (Covid19) - Evolution par pays et dans le monde (maj quotidienne) (coronavirus)" and is published by "P". The third result is titled "Coronavirus - Commerçants parisiens - Livraison à domicile" and is published by "Mairie de Paris".

data.gouv.fr Plateforme ouverte

Données Réutilisations Organisations Tableau de bord Documentation

coronavirus

27 jeux de données 54 réutilisations 2 organisations

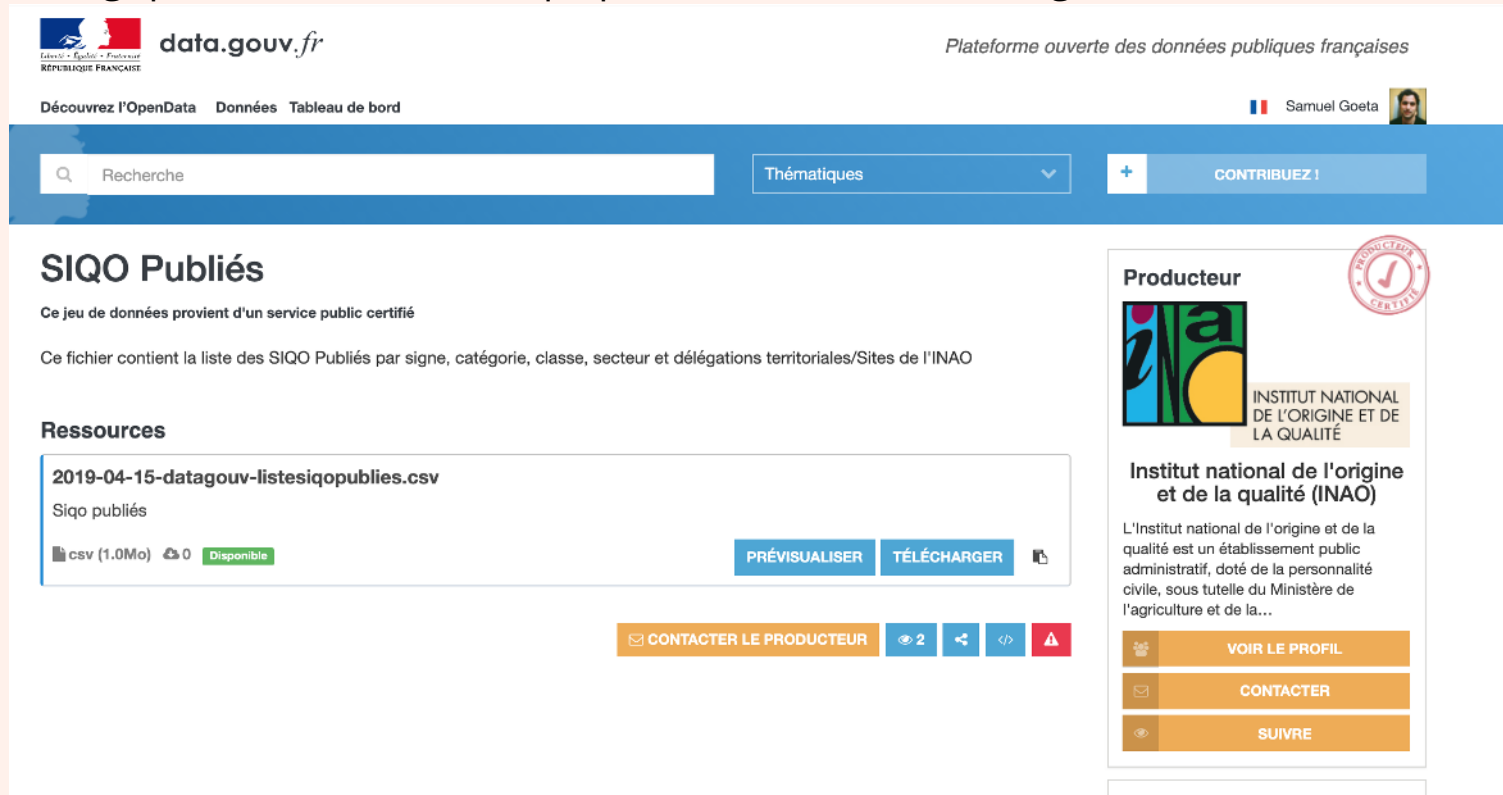
**Données relatives à l'épidémie de COVID-19 en France : vue d'ensemble**  
Tableau de bord de suivi de l'épidémie Les données sont exposées sur l'onglet vue d'ensemble du tableau de bord de suivi de l'épidémie disponible sur gouvernement.fr. Ce dernier présente depuis le 28 mars 2020 les données relatives à l'épidémie de  
Quotidienne France Département français 0 0

**CoronaVirus (Covid19) - Evolution par pays et dans le monde (maj quotidienne) (coronavirus)**  
[EDIT 12/09/2020] Vous retrouverez désormais dans les fichiers ci-dessous les 30 derniers jours, trop de personnes ne respectent pas la demande de ne pas récupérer trop souvent le jeu de données (aucun intérêt de récupérer toutes les minutes)  
2020 Quotidienne Monde Pays 20 6

**Coronavirus - Commerçants parisiens - Livraison à domicile**  
Liste des commerçants proposant un service de livraison à domicile Livraison à domicile : Les commerçants parisiens mobilisés : Vous êtes commerçant parisien et proposez de la livraison à domicile ? Dites-le nous pour permettre aux Parisiens isolés ou  
Inconnu 1 0

# Autre exemple de problèmes de découvrabilité

Ce fichier contient tous les signes officiels de la qualité et de l'origine (SIQO) : les AOC et AOP, IGP, Label rouge, spécialité traditionnelle garantie, Agriculture biologique... Très utile mais qui peut le trouver sur data.gouv.fr ?

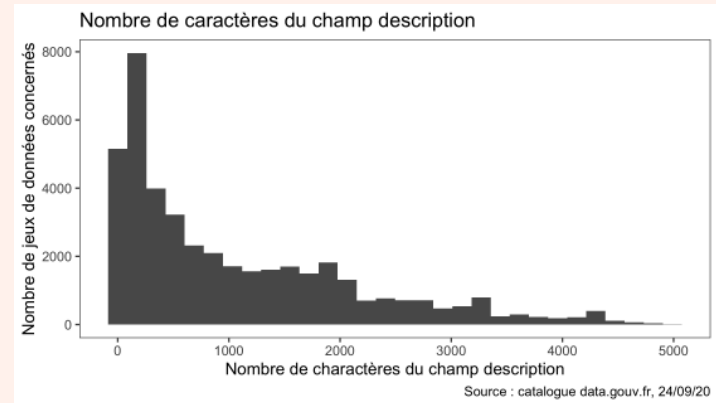


The screenshot shows the data.gouv.fr interface. At the top, the logo for data.gouv.fr is visible, along with the text "Plateforme ouverte des données publiques françaises". Below the logo, there are navigation links: "Découvrez l'OpenData", "Données", and "Tableau de bord". A search bar with the text "Recherche" and a dropdown menu for "Thématiques" are present. A "CONTRIBUEZ !" button is also visible. The main content area is titled "SIQO Publiés" and includes the text "Ce jeu de données provient d'un service public certifié" and "Ce fichier contient la liste des SIQO Publiés par signe, catégorie, classe, secteur et délégations territoriales/Sites de l'INAO". Below this, there is a "Ressources" section with a file listing: "2019-04-15-datagouv-listesiqopublies.csv" with "Siqu publiés" and "csv (1.0Mo) 0 Disponible". There are buttons for "PRÉVISUALISER" and "TÉLÉCHARGER". At the bottom of the file listing, there is a "CONTACTER LE PRODUCTEUR" button and social media icons. On the right side, there is a profile for the "Producteur" "Institut national de l'origine et de la qualité (INAO)", featuring the INAO logo and a "PRODUCEUR CERTIFIÉ" seal. The profile text states: "L'Institut national de l'origine et de la qualité est un établissement public administratif, doté de la personnalité civile, sous tutelle du Ministère de l'agriculture et de la...". Below the profile text are three buttons: "VOIR LE PROFIL", "CONTACTER", and "SUIVRE".

## Défi 2 : la documentation des données

Beaucoup de données sont accompagnées d'une description très réduite.

Sans ces informations, il est très difficile pour un usager de comprendre ce que contient le jeu de données, ses conditions de production des données et ses limites.



Sur data.gouv.fr, **22% des jeux de données** ont une description de moins de 180 caractères (un sms)

**59% ont moins de 1000** caractères (une demi-page)

# Un exemple de problème de documentation

Ce jeu de données sur les élections présidentielles de 2002, très réutilisé, comporte un champ description plus court que le titre du jeu de données.

The screenshot shows the data.gouv.fr interface for the dataset 'Election présidentielle 2002 - Résultats'. The header includes the logo of the French Republic and the text 'data.gouv.fr Plateforme ouverte des données publiques françaises'. The navigation menu contains 'Données', 'Réutilisations', 'Organisations', 'Tableau de bord', 'Documentation', and 'Connexion / Inscription'. A search bar is present below the navigation. The main content area features the title 'Election présidentielle 2002 - Résultats' and a sub-header 'Ce jeu de données provient d'un service public certifié'. Below this, it states 'Résultats de l'élection présidentielle 2002, 2ème tour, par communes'. A 'Ressources' section displays a resource titled 'Ressource sans nom' with a download button labeled 'TÉLÉCHARGER' and a 'Disponible' status. The 'Producteur' section on the right shows the logo of the 'MINISTÈRE DE L'INTÉRIEUR' and a description of the ministry's role.

# Un autre exemple de problème de documentation

Ici, la documentation est très succincte et ne fait que reprendre le libellés des colonnes. Or, qu'est ce qu'un musée de France ? comment les données sont collectées ? quand vont-elles être mises à jour ?

The screenshot shows the data.gouv.fr interface. At the top, there is the logo of the French Republic and the text 'data.gouv.fr Plateforme ouverte des données publiques françaises'. Below this is a navigation bar with links for 'Données', 'Réutilisations', 'Organisations', 'Tableau de bord', and 'Documentation', along with a 'Connexion / Inscription' button. A search bar is present with the text 'Recherche'. The main content area features the title 'Liste et localisation des Musées de France'. Below the title, it states 'Ce jeu de données provient d'un service public certifié' and includes a 'NEC MERGITUR' badge. The description reads: 'Liste des musées labellisés "Musée de France". Libellés des colonnes : Région ; Département ; Date de retrait de l'appellation ; Date d'appellation ; Identifiant musée (ID) ; Nom du musée ; Coordonnées postale (Adresse, Ville, CP) ; Téléphone ; Fax ; Site web ; Fermeture annuelle ; Période d'ouverture ; Jours nocturnes'. There is a 'Ressources' section with a link to 'Voir aussi : ressources communautaires' and a specific entry 'Liste des musées de France au 31.12.2017'. On the right side, there is a 'Producteur' section with a 'PRODUCEUR CERTIFIÉ' seal and the logo of the 'MINISTÈRE DE LA CULTURE' with the motto 'Liberté, Égalité, Fraternité'. Below this, it describes the mission of the Ministry of Culture.

# Un autre exemple de problème de documentation

Pour ces données, la description est relativement longue **mais...** elle ne fait que reprendre les termes de la licence, après un premier paragraphe très peu informatif.

# Datasheet for Datasets : une approche alternative pour la documentation des données

## Traduction et adaptation du modèle de description des données "Datas"

### Motivations pour la création du jeu de données

- Pourquoi le jeu de données a-t-il été initialement créé ?
- Quelles ont été les utilisations non prévues du jeu de données ?
- Pour quelles autres tâches le jeu de données pourrait-il être utilisé ?
- Quelles sont les utilisations trompeuses du jeu de données ?
- Qui a financé ou soutenu la création du jeu de données ?

### Composition du jeu de données

- Que contient le jeu de données principalement ? Les enregistrements représentent-ils principalement des documents, des personnes, des territoires, des entreprises... ?
- Dispose-t-on d'un schéma décrivant les variables du jeu de données ?
- Que contient chaque champ du jeu de données ?
- Est-ce que le contenu du jeu de données dépend de ressources externes (ex. identifiant SIRET ou lien vers le document...) ? De quelles garanties dispose-t-on concernant la pérennité de ces ressources ?

### Processus de collecte des données

- Comment les données ont été collectées (avec des capteurs, manuellement, par des outils informatiques...) ?
- Qui a assuré le processus de collecte de données (des agents, des bénévoles, des étudiants...) ?
- Quelle a été la période de collecte des données ?
- Les données ont-elles été collectées directement ou inférées à partir d'autres données ?
- Les données ont-elles été collectées sur un échantillon ? Quelle est la population complète ? Selon quelles méthodes ?
- Quelles sont les erreurs connues, les limites, les sources de bruit ou de redondances associées à ces données ?

### Pré-traitement des données

- Comment les données ont-elles nettoyées ou préparées ?
- Les données « brutes » ont-elles été conservées ? Sont-elles diffusées ?
- L'outil de pré-traitement des données est-il disponible ?

### Diffusion du jeu de données

- Les données sont-elles diffusées en ligne ? Selon quelles modalités (sur un portail open data, un site web, une API...) ?
- Si non, les données sont-elles diffusées au cas par cas ? à la demande ?
- Selon quelle licence les données sont-elles diffusées ?
- Des redevances ou des restrictions sont-elles appliquées dans l'accès aux données ?

- Motivations pour la création du jeu de données
- Composition du jeu de données
- Processus de collecte des données
- Pré-traitement des données
- Diffusion du jeu de données
- Maintenance du jeu de données
- Considérations légales et éthiques

## Défi 3 : la qualité des données

Enfin, le troisième défi porte sur la **qualité** des données.

Le rapport de [l'Open Data Barometer](#) signale que :

“les données des gouvernements sont souvent incomplètes, pas actualisées, de mauvaise qualité et fragmentaires” et

“de manière générale, la gestion et la publication des données ouvertes sont fragiles et sujettes à des erreurs multiples.”

△ la qualité des données ne dépend que de l'usage qui en est fait.

La documentation est indispensable pour que le public parvienne à saisir les usages précédents des données et les adapte à de nouveaux usages, souvent inconnus de celles et ceux qui les produisent.

# La qualité : une notion à prendre avec des précautions

Garfinkel et Bittner (1967), deux sociologues confrontés à des difficultés d'exploitation de dossiers médicaux, ont été tentés de dénoncer la « mauvaise qualité » du remplissage de ces dossiers.

Mais ils ont montré que les dossiers médicaux ont « **de bonnes raisons organisationnelles** » d'être « mal » remplis.

Le personnel de l'hôpital collecte des informations pour ses propres missions dans un contexte organisationnel orienté vers un certain type d'actions, le soin des patients, et non la recherche en sciences sociales.

Les dossiers médicaux comme les données sont des **écrits organisationnellement situés**. Ce sont des données métiers, **produites généralement à des fins de gestion**, qui deviennent un objet de savoir et de pouvoir.

# La circulation des données provoque des frictions

A chaque interface entre deux surfaces, **la friction consomme de l'énergie, produit de la chaleur et use les parties en mouvement.** La métaphore d'Edwards des frictions de données décrit ce qui survient à l'interface entre les « surfaces » de données : les points où les données bougent entre les personnes, les couches sociales, les organisations ou les machines [...] Chaque mouvement de données à travers une interface a un coût en temps, énergie et attention humaine. Chaque interface entre groupes et organisations, ainsi qu'entre les machines, représente un point de résistance où les données peuvent être rendues incompréhensibles, mal interprétées ou perdues. Dans les systèmes sociaux, les frictions de données **consomment de l'énergie et produisent des turbulences et de la chaleur.**



Edwards, Paul N., et al. "Science Friction: Data, Metadata, and Collaboration." Social Studies of Science, vol. 41, no. 5, 2011

# Un exemple de données inutilisables



## Défi 4 : les échanges entre producteurs et réutilisateurs

Pour répondre aux problèmes identifiés par les usagers liés à la découvrabilité, la documentation ou la qualité, on pourrait, plutôt que de les résoudre immédiatement, s'appuyer sur les commentaires et les retours des usagers pour identifier les priorités.

Mais en pratique, la plupart des agents en charge de l'ouverture n'ont pas de temps dédié à traiter ces commentaires et à les résoudre.

Résultat : les échanges se font rares et les usagers n'ont souvent pas de réponse. Sur [data.gouv.fr](https://data.gouv.fr), **40,5% des discussions** n'ont qu'un seul message, la question initiale. Les producteurs ne répondent pas systématiquement.

# **4 stratégies alternatives pour obtenir des données**

# Stratégie 1 : recourir au **droit d'accès**

- Le fondement : la **Déclaration des Droits de l'Homme et du Citoyen de 1789** dans son article 15, "la Société a le droit de demander compte à tout Agent public de son administration."
- Le droit d'accès des citoyens à l'information publique émerge en **1978 avec la loi dite CADA** du nom de la Commission d'Accès aux Documents Administratifs.
- 3e pays au monde après la Suède en 1766 et les Etats-Unis en 1966 avec le Freedom of Information Act (FOIA) en 1966 à accorder un "droit de savoir".



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

# Les conditions d'exercice du droit d'accès

## Une définition très large des documents administratifs

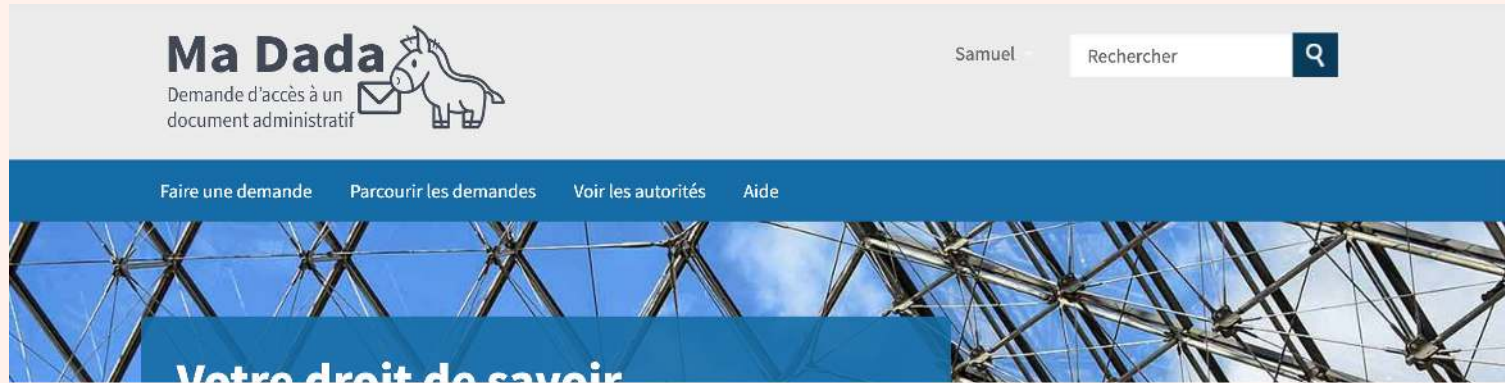
Les documents administratifs peuvent revêtir de nombreuses formes (dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, directives, instructions, circulaires, codes sources, etc.) et adopter tout support (écrit, enregistrement sonore ou visuel, forme numérique ou informatique).

Le droit d'accès s'applique aux documents administratifs produits dans le cadre d'une **mission de service public** ie :

Une mission d'intérêt général sous le contrôle de l'administration et qui est dotée à cette fin de prérogatives de puissance publique est chargée de l'exécution d'un service public.

Il ne s'exerce que si l'administration a effectivement en sa possession le document demandé, que si le document est formellement **achevé**, qu'il existe et qu'il n'est pas **préparatoire à une décision administrative** en cours.

# Madada.fr : faciliter l'exercice du droit d'accès



# Stratégie 2 : **Scraper** les données du web

Le scraping consiste à **extraire les données présentes sur des sites** web généralement dans des pages éparses. Il est courant qu'une organisation diffuse des données sur le web mais ne mette pas à disposition les données derrière ces sites. Or, pour beaucoup de sites publics, ces données auraient vocation à être librement réutilisables par toutes et tous.

Les outils de scraping permettent de **cibler les éléments à extraire sur les pages** (généralement par leurs propriétés dans le code source de la page ou dans la feuille de style), de lister toutes les pages concernées et d'extraire les éléments dans une base de données structurée.

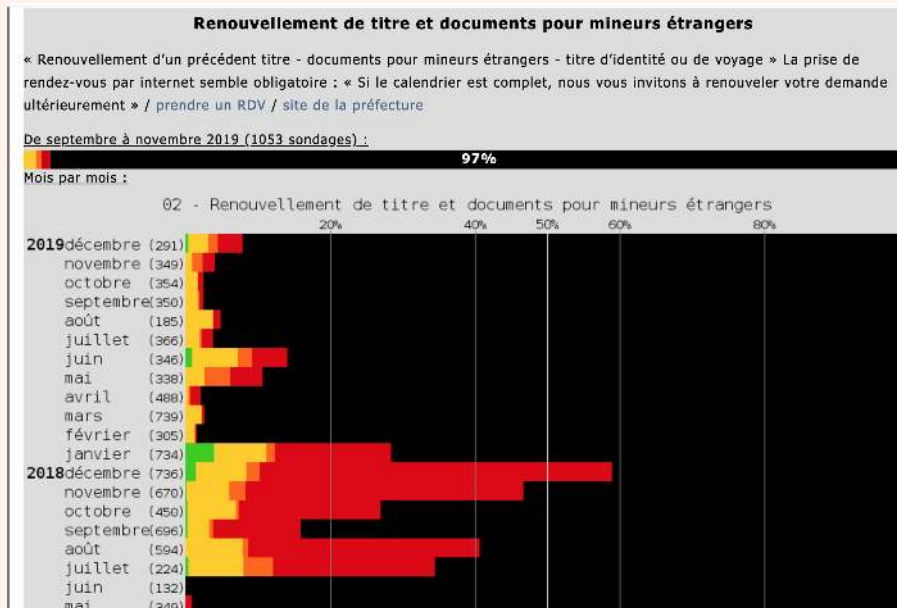
# NosDeputes.fr : un projet citoyen né du scraping

# Le Monde : un projet de scraping

Le journaliste Alexandre Léchenet avait réalisé en 2012 une enquête sur les dépassements d'honoraires des médecins pour Le Monde en extrayant les données du site ameli-direct.fr qui indique les tarifs habituellement exercés par chaque praticien.

# Cimade : à guichets fermés

la Cimade, une association de défense des réfugiés et du droit d'asile, a mis en place le site « A guichets fermés » pour dénoncer l'attente pour les formalités d'immigration devenue « un moyen de domination des personnes étrangères par la préfecture[1]. » Pour produire le site « **A guichets fermés** », l'association a développé des robots qui se rendent toutes les deux heures sur les sites de prise de rendez-vous en préfecture et extraient le délai d'attente pour les principales formalités.



# Stratégie 3 : constituer des données à partir des informations disponibles

La troisième stratégie consiste à rassembler des informations disponibles mais éparses dans une base de données.

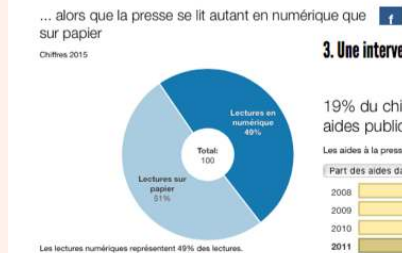
Les documents publiés par les administrations comportent généralement des graphiques, des tableaux ou des indicateurs qui, pris isolément, ne permettent pas de refaire les calculs et de produire de nouvelles analyses.

Par exemple, le SPIIL a constitué une base de données estimant le montant de chacun des dispositifs d'aide à partir de sources très diverses : rapports annuels de performance établis par Bercy, rapport de la Cour des Comptes, budget de l'Etat, questions au gouvernement ou encore des estimations publiées par le SPIIL...

	A	G	H	I	J	K	L	M	
1	Montants en millions d'euros	2009	2010	2011	2012	2013	2014	2015	Sources des données
17	Aide au développement des services en ligne des entreprises de presse	20,2	20,2	18,0	0,0	0,0	0,0	0,0	RAP 2008 à 2013, LFI 2014 et 2015
18	Fonds stratégique pour le développement de la presse	0,0	0,0	0,0	30,4	22,1	22,9	30,5	RAP 2008 à 2014, LFI 2015
20									
21	Autres aides à la presse (montants en millions d'euros)	2009	2010	2011	2012	2013	2014	2015	Sources des données
22	Aide au transport postal de la presse - Partie portée par le programme 150 « Presse » (compté comme aide directe jusque 2013)	107,5	111,0	110,3	108,1	247,6	1,9	0,0	RAP 2008 à 2014
23	Aide au transport postal de la presse - Partie portée par le programme 134 « Développement des entreprises et du tourisme »	158,0	131,0	131,7	132,9	0,0	160,0	130,0	RAP 2008 à 2013, LFI 2014 et 2015
24	Exonération cotisation foncière des entreprises (avant 2009 : papier seulement)	200,0	200,0	200,0	200,0	200,0	200,0	200,0	Rapport public annuel Cour des comptes 2013
26	TVA super réduite (avant 2014 : papier seulement)	190,0	195,0	180,0	250,0	175,0	165,0	170,0	RAP 2008 à 2014, PAP 2015
26	Exonération éco-contribution Ecofisc	29,6	29,6	30,4	31,2	38,4	40,0	41,6	Estimation SPIIL
27	Abattement cotisations sociales journalistes	141,1	141,1	141,1	141,1	141,1	141,1	141,1	Estimation SPIIL
28	Abattement impôt sur le revenu journalistes	60,0	60,0	60,0	60,0	60,0	60,0	60,0	Rapport public annuel Cour des comptes 2013
29	Amorces judiciaires et légales	193,0	193,0	193,0	193,0	193,0	193,0	193,0	DE n° 13513 de M. Zocchetto publiée dans le JO Sénat du 31
30	Délocalisation dons IPG	0,0	0,0	0,0	1,2	1,9	1,9	1,9	Estimation SPIIL
44									

# SPIIL : panorama des aides à la presse

Avec ces données, nous avons conçu des visualisations pour illustrer les messages portés par le SPIIL dans l'analyse de ces données



### 3. Une intervention massive de l'État

19% du chiffre d'affaires de secteur est issu des aides publiques

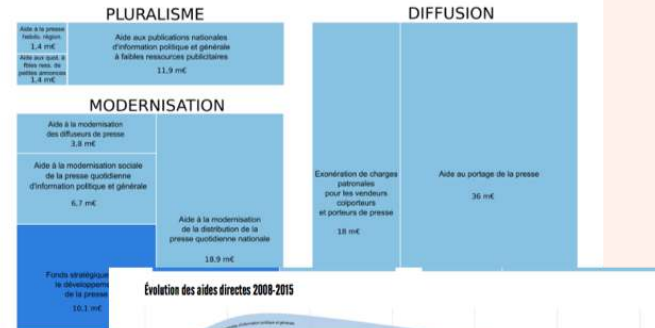
Les aides à la presse se composent des aides directes et des aides indirectes.



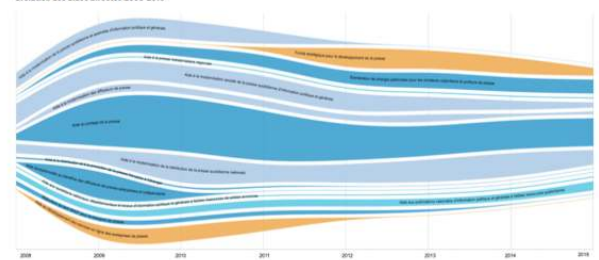
### 1. Des aides directes quasi-exclusivement papier et IPG

Répartition des aides directes 2015

- Aides réservées à la presse papier
- Aides indifférentes au support (papier / numérique)



### Évolution des aides directes 2008-2015



# Transparency France : cartographie de la corruption

Un bénévole de Transparency France, surveille les chroniques judiciaires de la presse locale en France pour détecter les affaires de corruption et met les nouvelles affaires en base de données. Le site [visualiserlacorruption.fr](http://visualiserlacorruption.fr) présente plus de 900 condamnations pour renforcer le plaidoyer de l'association sur le manque de moyens de la justice en France pour lutter contre la corruption.



# Stratégie 4 : faire appel à une foule de contributeurs

On désigne couramment sous le terme de crowdsourcing l'acte de déléguer des actions à un grand nombre d'individus sous la forme d'un appel ouvert à la contribution.

Le crowdsourcing peut aussi servir à des projets d'intérêt général afin de constituer des bases de données ouvertes. Dans l'ensemble des exemples qui suivent, les contributeurs se sont mobilisés pour collecter des données dans des projets à but non lucratif, souvent afin de servir une cause militante ou pour passion.

# Greenpeace : une campagne sur les cantines

Entre septembre 2017 et février 2018, près de 8000 personnes ont collecté des données précieuses sur les menus de cantines des écoles primaires publiques en France : nombre de repas végétariens proposés, fréquence des options végétariennes, origine bio de la viande, intitulés des menus...

Grâce à cette grande enquête, Greenpeace a recueilli plus de 12 000 contributions couvrant plus de 3 200 villes et au moins 60 % de la population des écoliers. Bien qu'elles ne couvrent pas tout le territoire, ces données ont permis de montrer que 69 % des enfants sont obligés de manger de la viande ou du poisson presque tous les jours.

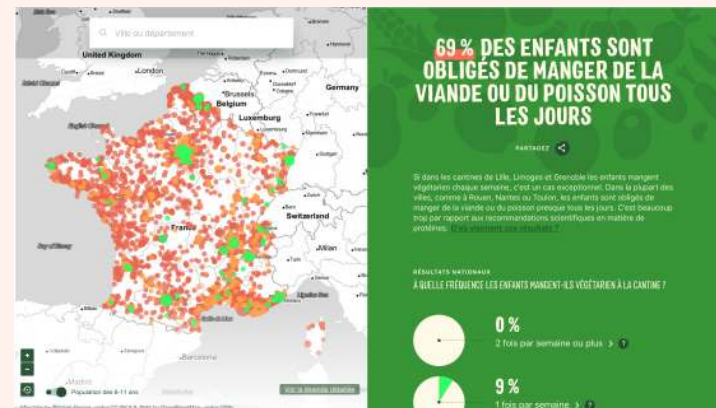
# Greenpeace : une campagne sur les cantines

Pour en savoir plus : lire [cet article](#) sur notre campagne

Outil de contribution :



Site de plaidoyer et de mobilisation



# Regards Citoyens : numériser les déclarations

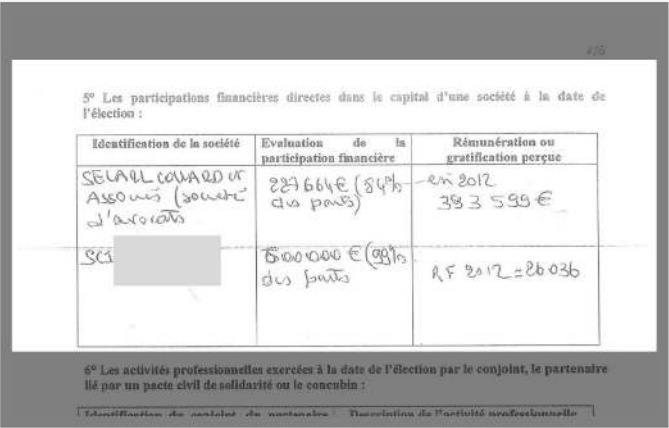
En 2014, les déclarations remplies des 577 députés et 348 sénateurs comportant chacune 12 parties, soit un total de plus de 11 000 extraits de formulaires manuscrits, ont été numérisées en une semaine par près de 8000 contributeurs.

Numérisons les déclarations d'intérêts **Participer** Les contributeurs Consulter les déclarations Sam Goëta

## Déclaration de Gilbert Collard

Partie 6 / 12

### Participations financières



5° Les participations financières directes dans le capital d'une société à la date de l'élection :

Identification de la société	Évaluation de la participation financière	Rémunération ou gratification perçue
SELAH COLLARD et ASSOCIÉS (société d'associés)	227 664 € (84%) en 2012	38 3 599 €
SCI [redacted]	500 000 € (81%) des parts	R.F. 2012 = 26 036

6° Les activités professionnelles exercées à la date de l'élection par le conjoint, le partenaire lié par un pacte civil de solidarité ou le concubin :

[Lien permanent vers cet extrait de déclaration](#)

### Saisir les informations

Merci de saisir ci-dessous une ligne par fonction déclarée par le parlementaire.

Pour supprimer une ligne, cliquez sur le bouton « □ » de la ligne correspondante.  
Si le parlementaire n'a rien saisi ou a indiqué « Néant », cliquez sur le bouton « Valider le formulaire vide ».

Identification de la société	Évaluation de la participation financière	Rémunération ou gratification perçue
Société n°1	Participation n°1	Rémunération n°1 <input type="checkbox"/>
Société n°2	Participation n°2	Rémunération n°2

[Signaler un problème](#) [Valider le formulaire vide](#) [Changer de déclaration](#)

Si vous avez le sentiment que nous avons mal détecté cette partie ou qu'il manque des informations, merci de nous l'indiquer en cliquant sur « Signaler un problème », nous vous proposerons un autre extrait de déclaration à saisir.

# Merci !

Contact : [samuel@dataactivist.coop](mailto:samuel@dataactivist.coop)