

Open Data, Myths and Reality

Data and Algorithms for Public Policy

Timothée Gidoin

SciencesPo, 2019-09-13

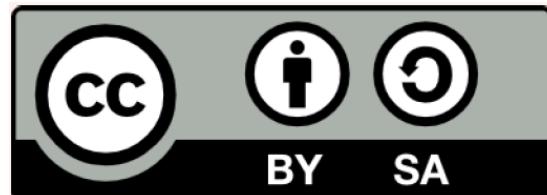
Before we start

Slides : https://gidoin.github.io/sciencespodata/lesson2_opendata.html

Sources : <https://github.com/Gidoin/sciencespodata/>

This production is freely reusable under the terms of the licence [Creative Commons 4.0 BY-SA](#).

The content of this presentation is partly inspired by other presentations made by Dataactivist team. I warmly thank them and notably [Joël Gombin](#) for his help.



Before we start



Simon Chignard is an expert in Open Data. He published in 2015 a best-seller book, [Datanomics](#), and advises Prime Minister services. [Twitter](#)



Jean-Marie John Mathews is a data Scientist and PhD student in algorithmics on the impacts of so-called "ethical by design" algorithms in artificial intelligence, he works with the [Medialab](#)



Timothée Gidoin cofounded [Datagora](#) an initiative that aims to democratize reliable sources. He works in parallel for [Dataactivist](#) as an Open Data consultant. [Twitter](#)

Before we start: reminder

Date	Session #		Teacher
05/09	1	Introduction to data policy	Simon Chignard
12/09	2	Open Data (data as a policy)	Timothée Gidoin
19/09	3	Public sector algorithms: « with great power comes great responsibility »	Simon Chignard
26/09	4	Workshop data manipulation #1: data preparation and analysis	Timothée Gidoin
03/10	5	Workshop data manipulation #2: datavisualization	Timothée Gidoin
10/10	6	Workshop Machine-learning #1 : introduction	Jean-Marie John Mathews
17/10	7	Use Case #1: Predisauvetage (Guest: Antoine Augusti)	Simon Chignard
24/10	8	Explainable algorithms: why ? how ?	Simon Chignard
07/11	9	Use case #2: Predictive policing (Guest: Bilel Benbouzid)	Simon Chignard
14/11	10	Workshop Machine-learning #2 : Fairness	Jean-Marie John Mathews
21/11	11	Workshop Machine-learning #3 : Explicability	Jean-Marie John Mathews
28/11	12	Oral presentations of coursework	Simon Chignard

Before we start: reminder

- **Midterm Exam :**
 - **25% of your total grade**
 - By group of **2 students**
 - Data manipulation, analysis visualisation exercise based on open data
 - To be submitted before **18/10 11:59 pm** (more details to come)
- **Final Exam :**
 - **75% of your total grade**
 - By group of **4 students**
 - 10-pages paper on the analysis of 3 uses cases (outside France) in one of the following topics: social benefits, police / justice, education, public sector human resources
 - 1 oral presentation (15 min) during the last session
 - 1 Medium blog post (1,5 pages) to present your findings
 - Evaluation of final exam: 50% quality of the analysis, 25% oral presentation, 25% quality of Medium blogpost

May the data be with you !



Data

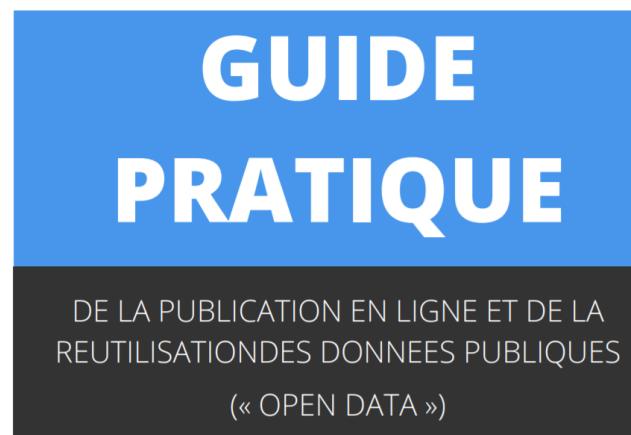
What are data ?



What are data ?

Une donnée correspond à la représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement

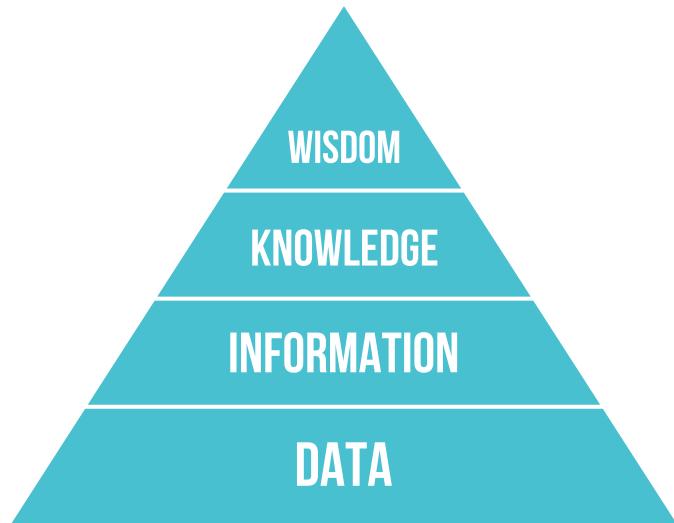
*Data correspond to the representation of information in a **conventional form** intended to facilitate its processing*



*Présentation du cadre juridique
de l'ouverture des données*

CNIL & CADA's Open Data Practical Guide, *in French*

Data-Information-Knowledge-Wisdom pyramid

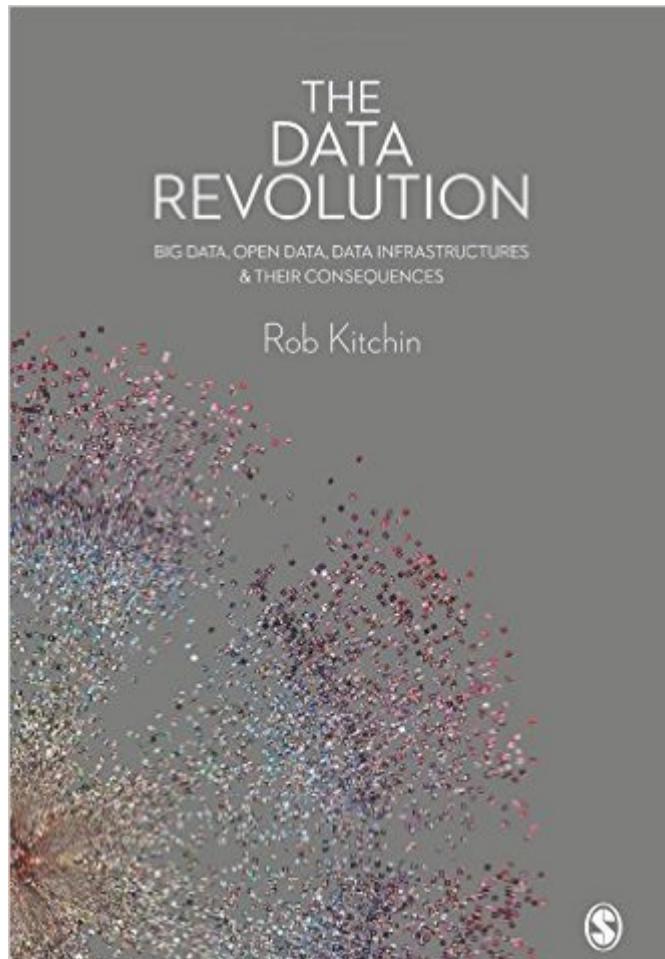


Attributed to [Russell Ackoff, 1989](#)

Data may be :

- Facts
- signals
- symbols

What are data ?



*Data are commonly understood to be the raw material produced by **abstracting the world** into categories, measures and other representational forms – numbers, characters, symbols, images, sounds, electromagnetic waves, bits – that constitute the **building blocks** from which information and knowledge are created.*

Data or capta ?

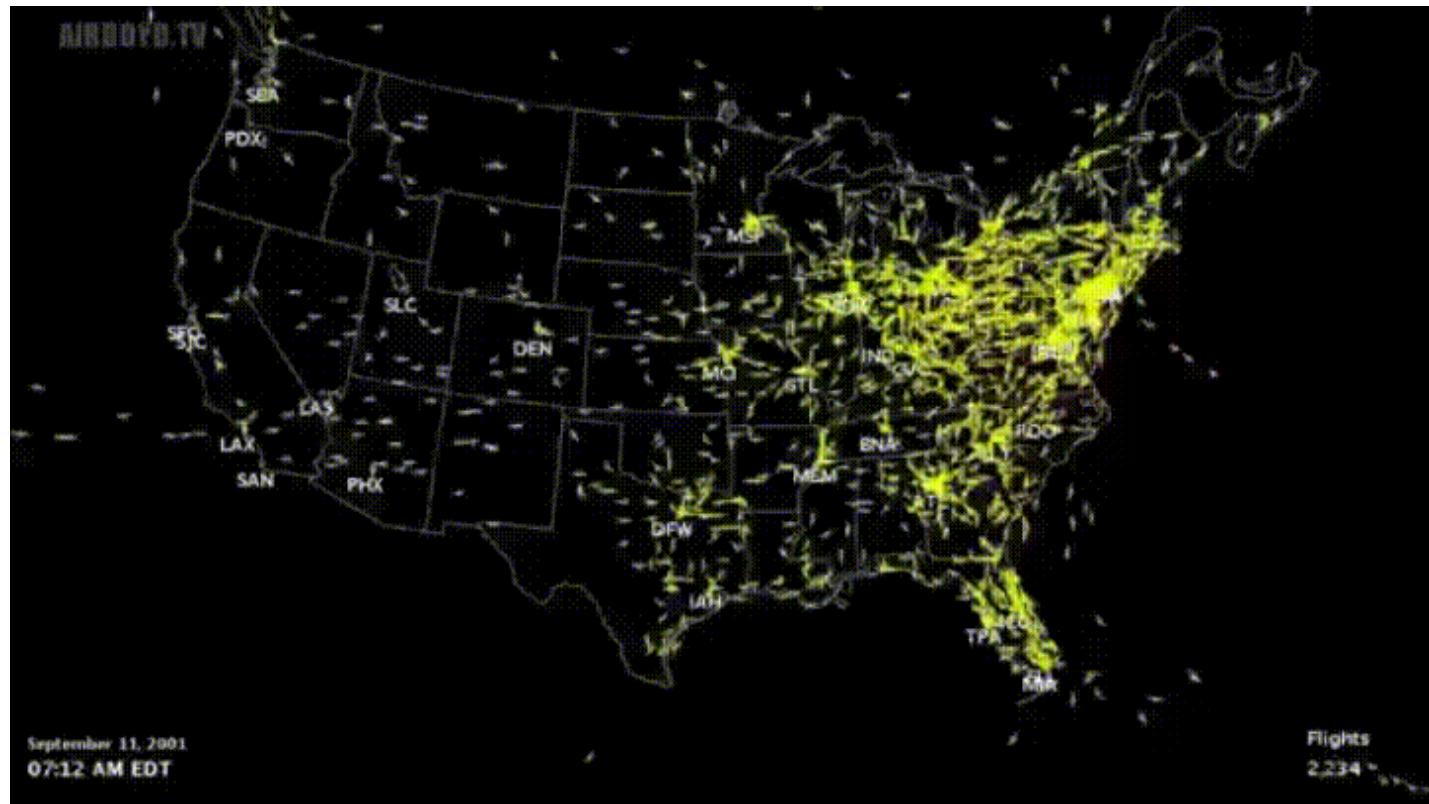
Technically, what we understand as data are actually **capta** (derived from the Latin capere, meaning 'to take'); those units of data that have been **selected and harvested** from the sum of all potential data.

Kitchin, 2014



Data are cool !

US Air traffic on September 11th 2001



Source

Data are cool !

"100 years of world cuisine"



NKB Dataviz

Data are cool !

 [sue montgomery](#)
@MontgomerySue

[Suivre](#) ▾

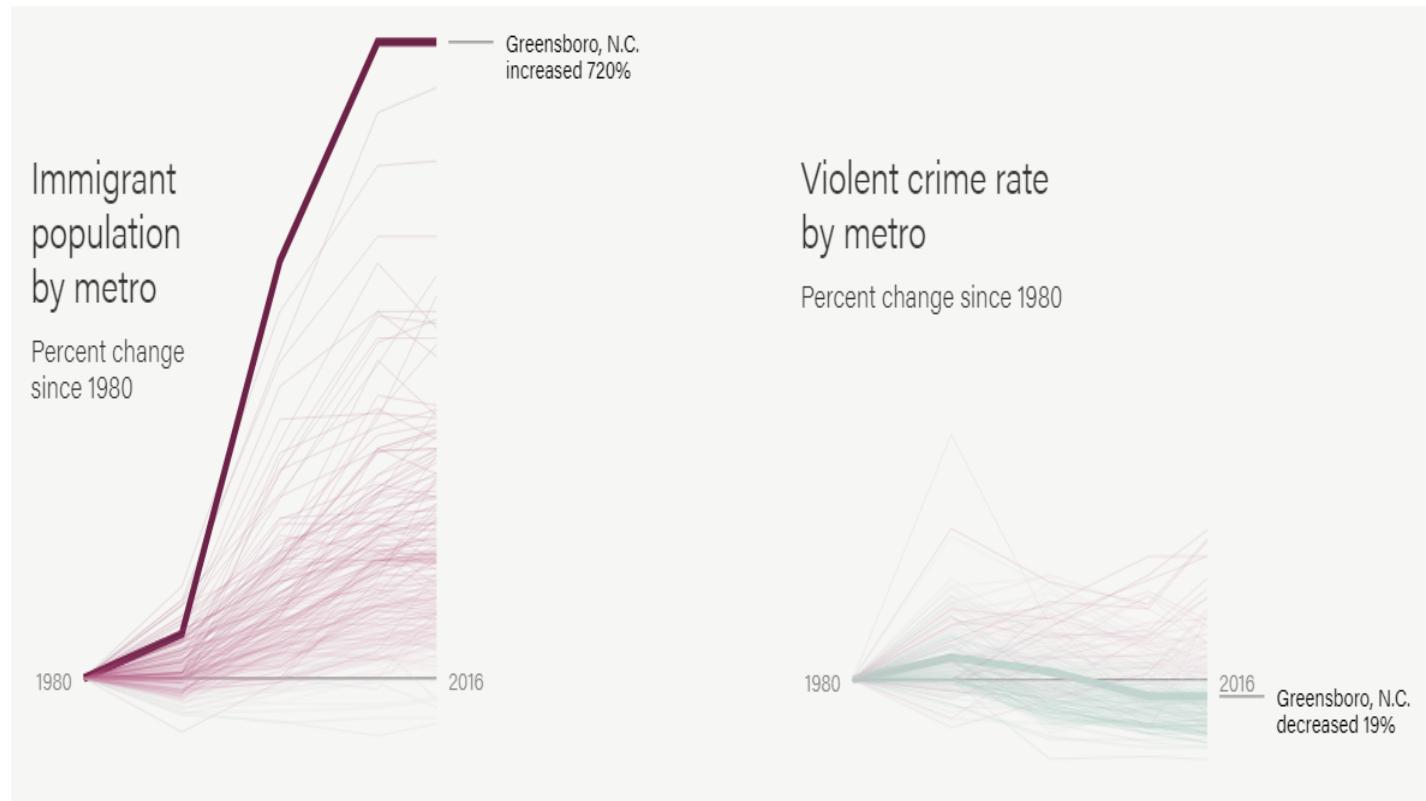
Je tricote dans le@conseil municipal parce que cela m'aide à me concentrer. J'ai lancé un projet dans lequel je tricote en rouge quand les hommes parlent et en vert quand les femmes parlent. Jour #1 #reclaimingmytime #femmes #itaintpretty #ecouter #pouvoir



17:24 - 13 mai 2019

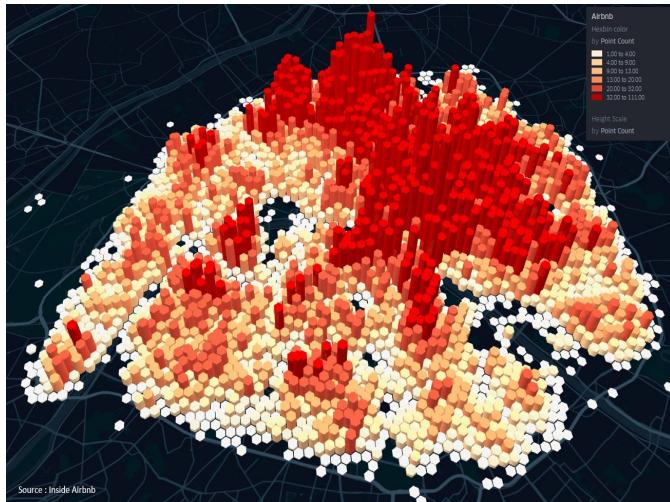
Source

Data are cool !



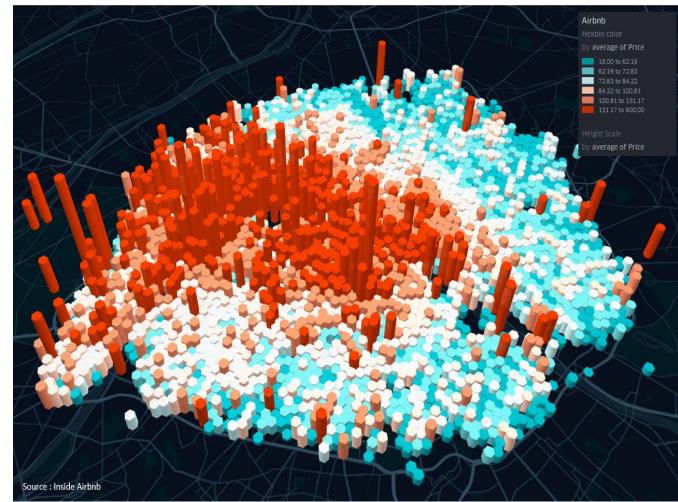
Source

Data are cool !



Number of Airbnb appartments available in Paris

Source



Average price of Airbnb appartments in Paris

Data are cool !



(<https://commons.wikimedia.org/wiki/Special:FilePath/Singapore%20Eng%202009%20ingredients.jpg?width=1000>)



(<https://commons.wikimedia.org/wiki/Special:FilePath/Bloody%20Mary%202009%20ingredients.jpg?width=1000>)

Wikidata Query Service

([https://query.wikidata.org/#SELECT%20%23Cocktail%20%23CocktailLabel%20\(SAMPLE\(%3Freipec](https://query.wikidata.org/#SELECT%20%23Cocktail%20%23CocktailLabel%20(SAMPLE(%3Freipec)

Open data

A talk to start..



Start the Ted talk at 09:00

Open data : definition

According to Wikipedia, open data is :

Open data is the idea that some data should be freely available to everyone to use and republish as they wish. One of the most important forms of open data is open government data (OGD), which is a form of open data created by ruling institutions.

Open data is both a ideological movement and a practical way of publishing data freely available and usable

According to French Government :

Open data deals with the effort that public organisms do in order to share the data that they own. This opening has to be done **for free, in public format and allow reuse of data**

French law considers that data produced/owned by public administrations/local authorities have to be made available to everyone. **This doesn't include private informations or data that may harm the national security**

Why should Gov open data ?



- **Innovation**
- **Modernisation**
- **Transparency**
- **Economic opportunities**

Open data : key milestones



DDHC, Art. 15.
La Société a le
droit de
demander
compte à tout
Agent public de
son
administration.

Loi CADA : droit
d'accès à
l'information
publique

Loi Lemaire :
ouverture des
données
devient
obligatoire

1789

1978

2016



1966

Freedom of
Information Act,
USA

2003

Directive
européenne

2007

Directive
européenne
INSPIRE

2013

Charte du G8
sur l'open data

7th December 2007 : Sebastopol meeting

 **What ?** : A meeting of the Open Government Group in Sebastopol (California), headquarters of O'Reilly editions

 **Why ?** : Influence the future president of the US to boost and implement Open Data

 **How ?** : By adopting a declaration that define the key principles of Open Government Data



1. Completeness

Datasets released by the government should be as complete as possible, reflecting the entirety of what is recorded about a particular subject.

All raw information from a dataset should be released to the public, except to private information and information that may be sensitive for national safety

2. Primacy / Raw data

Datasets released by the government should be **primary source data**

3. Timely data

Datasets released by the government should be available to the public **as soon as possible**

4. Ease of Physical and Electronic Access

Datasets released by the government should be as accessible as possible, with accessibility defined as **the ease with which information can be obtained**, whether through physical or electronic means

5. Machine readability

Machines can handle certain kinds of inputs much better than others. Information shared in the widely-used PDF format, for example, is very difficult for machines to parse

Thus, information should be stored in widely-used file formats that easily lend themselves to machine processing.

6. Non-discriminatory access to data

“Non-discrimination” refers to who can access data and how they must do so

Non-discriminatory access to data means that any person can access the data at any time without having to identify him/herself or provide any justification for doing so.

7. Open standards

Open standards refer to who owns the format in which data is stored

Do you know a widespread proprietary format ?

Microsoft Excel is a fairly commonly-used spreadsheet program which costs money to use. Freely available alternative formats often exist by which stored data can be accessed without the need for a software license

8. Open Licence

Maximal openness includes clearly **labeling public information as a work of the government** and **available without restrictions on use as part of the public domain**

In France two type of licences : Licence Ouverte (CC-BY) ou ODBL (CC-BY-SA). what is the difference ?

LO (from Etalab) / ODBL : with both you can share, edit the database, create derived products and had a commercial usage

- **LO** : more "permissive" : you just have to mention the source and the date of update
- **ODBL** : you have to share and open your database at the same conditions

Open Data : to go further

Find the 8 principles of Open Government Data that were adopted in Sebastopol December 2007



Then in 2010 this list was slightly completed and updated (10 principles) by the Sunlight Foundation.

In 2013 the Sunlight Foundation wrote instructions and recommandations to implement concretely Open Data ("Open Data guidelines") based on those 10 principles

Open Government

What does OpenGov mean ?



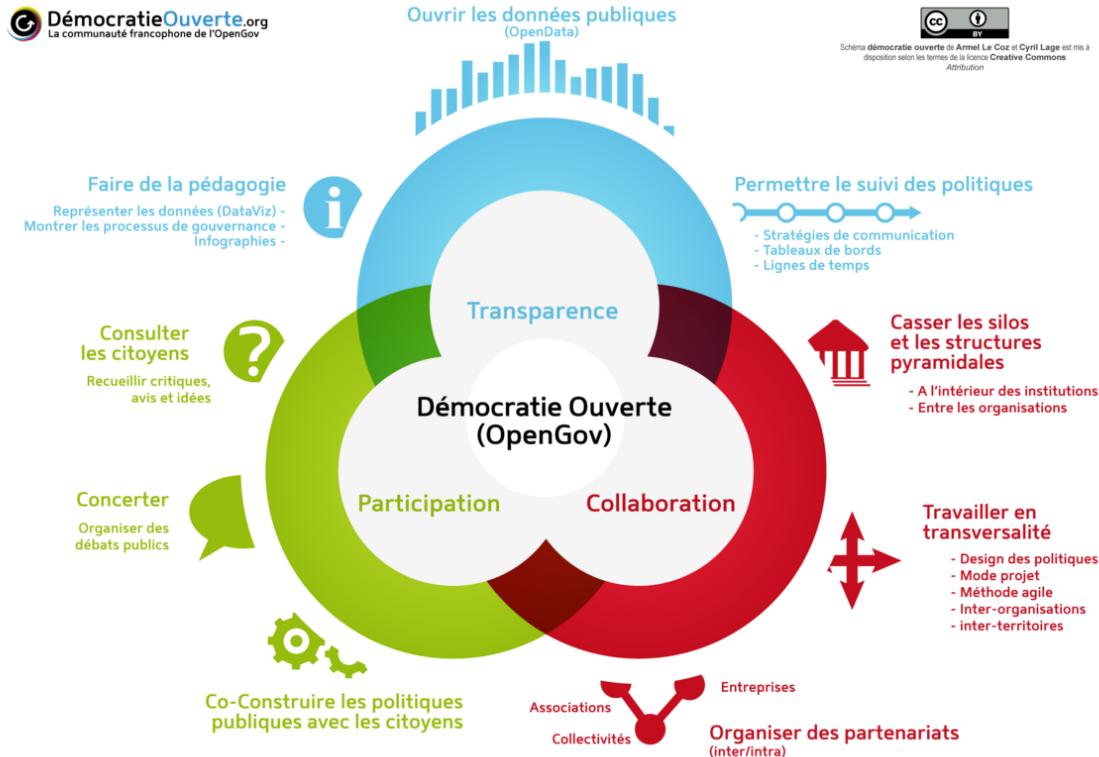
A definition of Opengov

The OECD defines open government as a culture of governance based on innovative and sustainable public policies and practices inspired by the principles of **transparency, accountability and participation** that **fosters democracy and inclusive growth**



OECD report

A definition of OpenGov



Source: Démocratie Ouverte

A definition of Opengov

Transparency

- *Ouvrir les données* : Open Data
- *Faire de la pédagogie* : being pedagogical
- *Permettre le suivi des politiques* : follow-up of public policies

Participation

- *Consulter les citoyens* : consult the citizens
- *Organiser des débats publics* : Organize public debate
- *Co-construire les politiques* : co-plan public policies with citizens

Collaboration

- *Casser les silos* : break silos between administrations
- *Travailler en transversalité* : work by being agile
- *Organiser des partenariats* : organize partnerships with civil society

A definition of Opengov

OpenGov aims at improving the efficiency and the responsibility of public governance. It affects both national and local authorities.

Three key pillars of Opengov:

1/ Transparency

2/ Accountability

3/ Public participation

The Opengov movement and ideas really increased in the last 10 years, notably through the support of a transnational actor...

The **Open Government Partnership** aka OGP

Open Government Partnership



In September 2011, a multilateral partnership has been created so as to promote the Open Gov principles and to translate them into concrete public policies : it's called **Open Government Partnership** (OGP)

Initially founded by 8 countries (Brazil, Indonesia, Mexico, Norway, Philippines, United Kingdom, the USA and... **South Africa** !), now OGP is composed of more than 70 countries, including France

Open Government Partnership

How does OGP work ?

To join OGP, a country has to be coopted by other civil society actors (for instance NGO such as Amnesty International) that testify the goodwill of the government

Once joined, the country has to write, by consulting civil society, a 2-years **national action plan** in which many **commitments** are exposed. Those public commitments have to relate with at least one of the 3 key OGP values : information transparency, public participation, accountability of public action

For instance :

- Developing a participatory budget representing X % of total budget of a local authority
- Implement an open data strategy that allows to open key datasets
- Consult citizens in the wake of a new project law

OGP and local gov

In 2016, OGP launched the “Subnational Government Pilot Program”

This decision recognized that many open government innovations and reforms are **happening at the local level** where governments can engage more directly with citizens and many crucial public services are delivered

The Pilot program consisted of 15 “pioneer” subnational governments who signed onto the Open Government Subnational Declaration and submitted their first Action Plans.

Then in 2018, OGP supported the launch of a global Community of Practice on Transparency and Local Open Government within the United Cities and Local Governments (UCLG)

This Community of Practice will support peer learning, networking, and wider awareness and capacity development on open governance and public integrity at the local level.

OGP and local gov : Paris

Commitments

Rows: 1-5 / 5

Page 1 of 1

Items: 5

Search Title, ID, Year, Country and Policy Areas

Title	ID	Year	Country	Policy Areas
A more inclusive participatory budget	PAR0001	2017	Paris, France	Marginalized Communities
To give the Parisians tools to be committed	PAR0002	2017	Paris, France	Capacity Building
Increasing community mobilization in new datasets request	PAR0003	2017	Paris, France	E-Government
Increase mobile and geolocalised crowdsourcing with DansMaRue V2	PAR0004	2017	Paris, France	E-Government
Kick-off of Paris city innovation lab	PAR0005	2017	Paris, France	Capacity Building

Paris 2016-2017 Action plan

OGP and local gov : Paris

Independent Reporting Mechanism (IRM): City of Paris Final Report 2017

Samuel Goëta, Dataactivist with the support of Timothée Gidoin

The Open Government Partnership (OGP) is a voluntary international initiative that aims to secure commitments from governments to their citizenry to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance. In 2016, OGP opened to subnational participants in their own right as part of a pilot program. The OGP Subnational Pilot Program consists of 15 subnational governments who submitted Action Plans and signed onto the Subnational Declaration at the Paris Global OGP Summit. This report summarizes the results of the development and implementation of Paris's pilot subnational action plan from January 2017 to December 2017.

The IRM reports for OGP pioneers will be published online primarily. As a result, this template is outlined in terms of the final site layout of the report.

Paris IRM final report

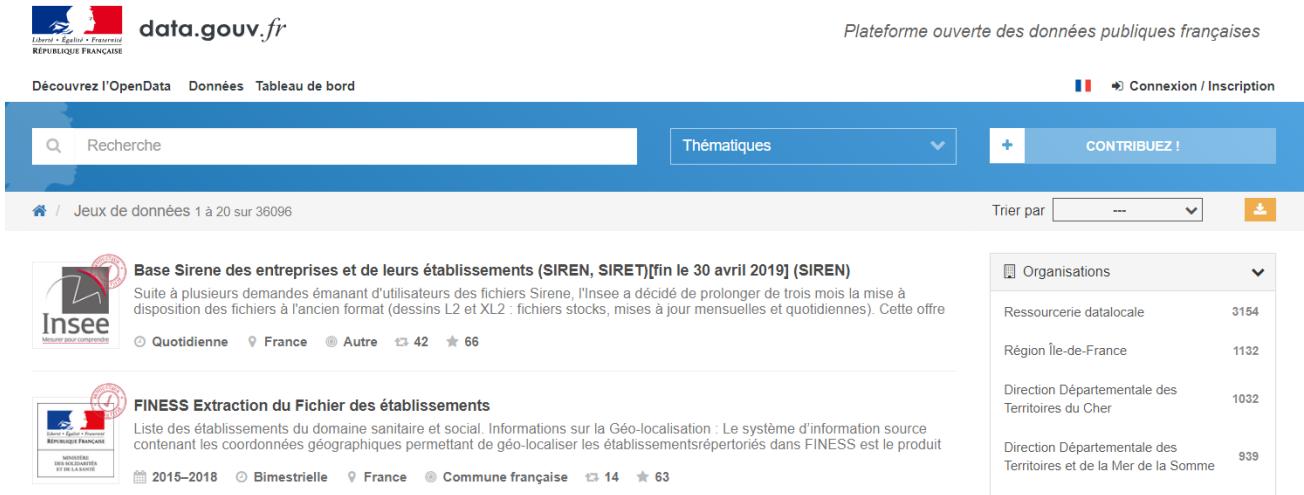
Coffee break : 5 minutes

Open Data platforms

**Have you already looked for
open data ?**

Data.gouv.fr ?

Data.gouv.fr is the national platform for French public data. It was designed and inaugurated in 2011, by the **Etalab** mission, and then refreshed in 2013.



The screenshot shows the homepage of data.gouv.fr. At the top, there is a logo of the French Republic (Liberté, Égalité, Fraternité) and the text "data.gouv.fr". Below the logo, there is a navigation bar with links to "Découvrez l'OpenData", "Données", and "Tableau de bord". On the right side of the navigation bar, there is a "Connexion / Inscription" link. The main search bar has a placeholder "Recherche". To the right of the search bar is a "Thématiques" dropdown menu. Below the search bar, there is a breadcrumb navigation showing "Jeux de données 1 à 20 sur 36096". To the right of the breadcrumb, there is a "Trier par" dropdown menu and a download icon. The main content area displays two datasets: "Base Sirene des entreprises et de leurs établissements (SIREN, SIRET)[fin le 30 avril 2019] (SIREN)" from Insee, and "FINESS Extraction du Fichier des établissements" from MINISTERE DES SOLS ET DE LA SANTE. Both datasets include details like frequency (Quotidienne, Bimestrielle), location (France, Autre), and ratings (42, 14 stars). To the right, there is a sidebar titled "Organisations" with a list of organizations and their data counts: Ressourcerie datalocale (3154), Région Île-de-France (1132), Direction Départementale des Territoires du Cher (1032), and Direction Départementale des Territoires et de la Mer de la Somme (939).

There are thousands of datasets (but not only) coming from different type of public data producers such as Ministries, independent agencies, statistical institutes, local authorities but also third parties producers such as OpenStreetMap or OpenFoodFacts

Beyond national OD platform

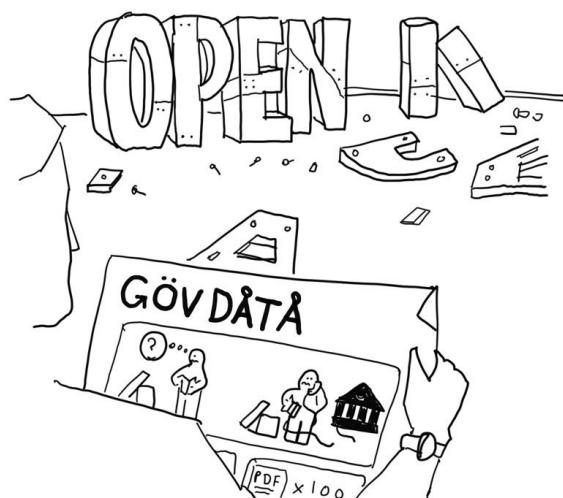
Data.gouv Portal is French **national** Open data platform but it's not the only website that publishes available public datasets. Do you know others ?

- There are also OD platforms led by **local authorities** at different levels. For instance, [Paris Data](#), [la Région Occitanie](#), [l'agglomération de Saint-Malo...](#)
- Open Data platforms led by **ministries**. Example : [data.education.gouv.fr](#)
- Open Data platforms led by **private organizations**. Example [DataNova](#) (La Poste)
- Datasets directly published in data provider website, without dedicated platform. Example : [Insee](#)

Challenge 1: data findability

Challenge 1 : data findability

Data findability is a major challenge. We have data portals and registries, but government agencies under one national government still publish data in different ways and different locations. (...) **Data findability is a prerequisite for open data to fulfill its potential and currently most data is very hard to find.**



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

Challenge 1 : data findability

According to you, how can we improve data findability ?

It is necessary to document/fill properly the metadata, that is to say the descriptive data associated to a dataset. For instance :

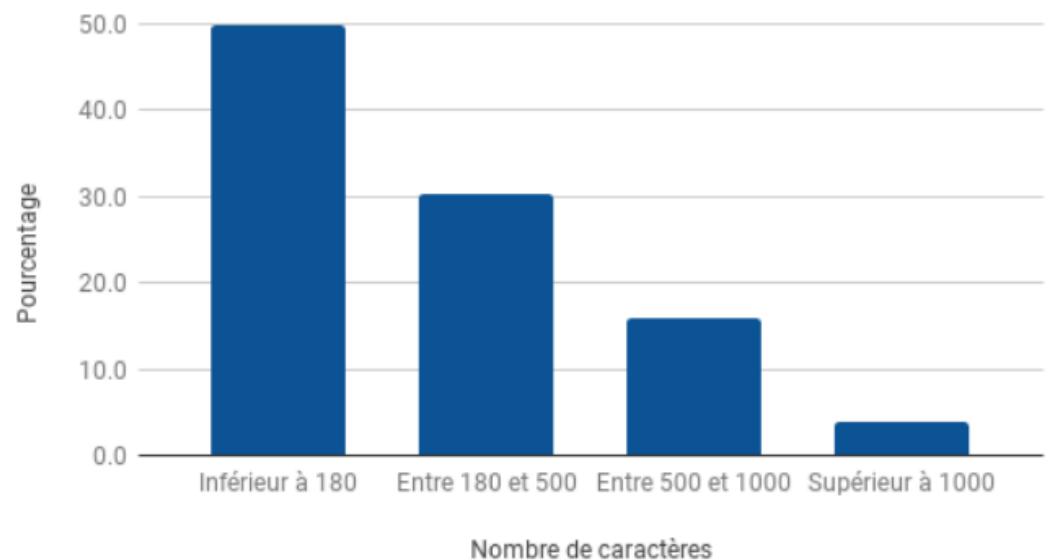
- Data producer
- Date of first publication
- Update frequency
- Date of last update
- Description of the dataset
- Variables explanation
- Topics / tags
- Time and space coverage (year, area, segmentation)

(This list is not exhaustive)

Challenge 1 : data findability

Summer 2017, Dataactivist realized a census of all the datasets opened by 15 major French cities (Paris, Lyon, Lille, Nantes...). More than 400 datasets were listed but identifying them was not easy at all...

Vue d'ensemble de la taille des descriptions des jeux de données ouverts recensés



Medium article

Challenge 1 : data findability

Thus,

Half of the descriptions of data opened by local authorities were written with less than 180 characters and only 4% of the datasets had a description above 1000 characters

Beyond very short descriptions, another challenge is to **detail the name of the variables**. Often they contain acronyms that make sense for public agents but can't be understood by common citizens.

In the example next slide, [l'enquête Etic du Ministère de l'éducation nationale](#), many columns have labels hardly readable such as "SiEquipementInf" or "Maint_PersEducHEcole" but the data producer made the effort to detail each of them in the metadata

Challenge 1 : data findability



data.education.gouv.fr

[Connexion](#)

[AIL](#) [DÉMARCHE](#) [DONNÉES](#) [API](#) [CARTOGRAPHIE](#) [LICENCE](#) [RÉUTILISATIONS](#)

Enquête ETIC 1er degré



[Informations](#)

[Tableau](#)

[Analyse](#)

[Export](#)

[API](#)

Departement	NbEleve	NbEnseignant	SiEquipementInfr	SiVoletNumProjetEcole	SiEntDisponible	EnseignantReferent	Maint_PersCom	Maint_PersEducHEcol
HAUTE-GARONNE	308	12	non	oui	non	oui	oui	non
HAUTE-GARONNE	78	3	non	non	non	non	oui	non
HAUTE-GARONNE	195	9	non	oui	oui	non	oui	non
HAUTE-GARONNE	40	2	non	oui	non	non	non	non
HAUTE-GARONNE	147	9	non	non	non	non	non	non
HAUTE-GARONNE	209	4	oui	oui	oui	non	oui	non
HAUTE-GARONNE	69	3	non	oui	non	non	non	non
HAUTE-GARONNE	232	11	oui	oui	non	non	oui	non
HAUTE-GARONNE	212	10	oui	oui	non	oui	oui	non
HAUTE-GARONNE	95	6	oui	oui	non	non	oui	non

Challenge 1 : data findability

EnseignantReferent

Y a-t-il un enseignant référent pour le numérique dans l'école ?

Nom (identifiant) **enseignantreferent**

Type texte

Exemple

non

Maint_PersCom

La maintenance et l'assistance technique des équipements numériques et du réseau pédagogique sont assurées par du personnel communal ou de la collectivité compétente ou mandaté par la collectivité

Nom (identifiant) **maintassistech_perscom**

Type texte

Exemple

non

Maint_PersEducHEcole

La maintenance et l'assistance technique des équipements numériques et du réseau pédagogique sont assurées par du personnel de l'éducation nationale hors école

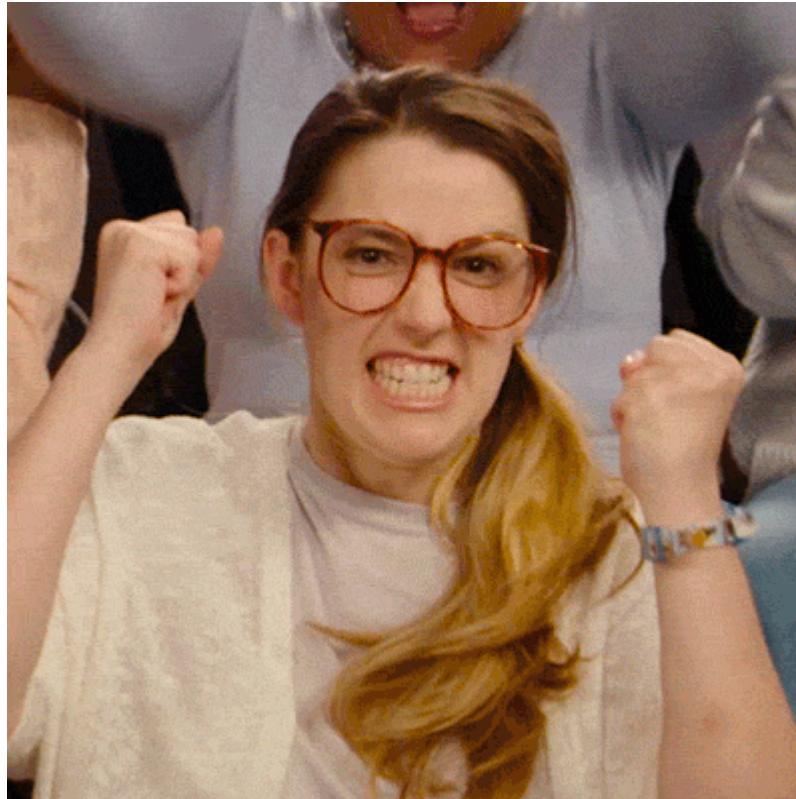
Nom (identifiant) **maintassistech_perseduchecole**

Type texte

Exemple

non

Challenge 1 : data findability



Challenge 1 : data findability



[En savoir plus](#) sur l'inclusion de vos ensembles de données dans l'outil de recherche d'ensembles de données.

Read [Simon Chignard's article](#) on Google Dataset Search

Challenge 2 : data quality

Challenge 2 : data quality

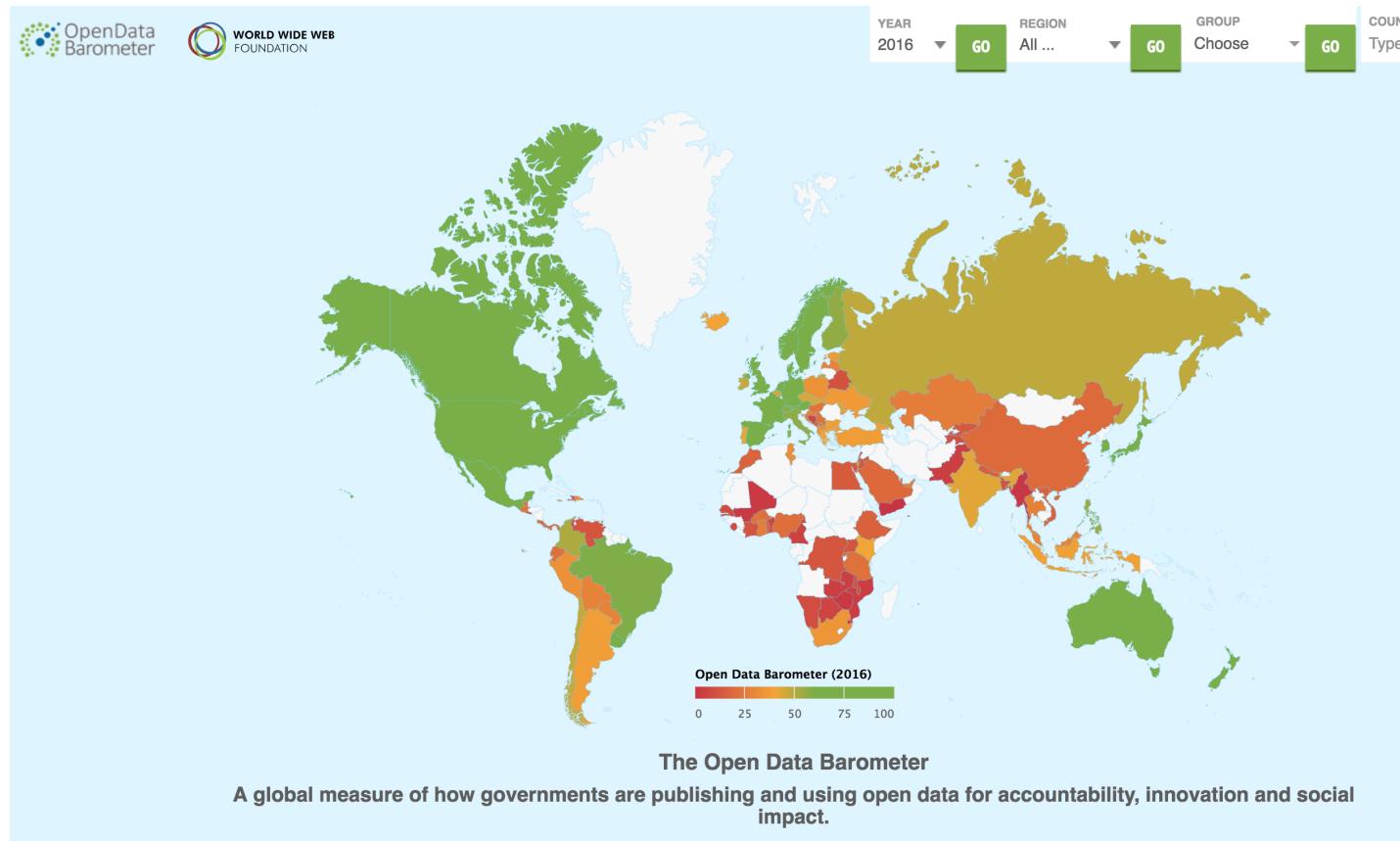
Government data is usually incomplete, out of date, of low quality, and fragmented. In most cases, open data catalogues or portals are manually fed as the result of informal data management approaches. **Procedures, timelines, and responsibilities are frequently unclear among government institutions tasked with this work.**

OpenDataBarometer ?

It's a global measure of how governments are publishing and using open data for accountability, innovation and social impact. The Leaders Edition looks at the 30 governments that have adopted the Open Data Charter and those that, as G20 members, have committed to G20 Anti-Corruption Open Data Principles.

<http://opendatabarometer.org/4thedition/report/>

Challenge 2 : data quality



Challenge 2 : data quality

Country	Score  <small>OUT OF 100</small>	Score Change  <small>SINCE FIRST EDITION</small>	Score Trend  <small>OVER PAST EDITIONS</small>	Readiness  <small>OUT OF 100</small>	Implementation  <small>OUT OF 100</small>	Emerging Impact  <small>OUT OF 100</small>
 Canada <small>See details</small>	76	18 	      	86	87	55
 United Kingdom <small>See details</small>	76	-4 	    	83	89	57
 Australia <small>See details</small>	75	17 	  	79	84	62
 France <small>See details</small>	72	17 	                                                                    <img alt="Up arrow" data-bbox="335 3945 345 3			

Challenge 2 : data quality

Sometimes data are well too aggregated...



The image consists of two side-by-side screenshots. The left screenshot shows a Twitter profile for 'Samuel Goëta' (@samgoeta). The bio reads: 'Rechercher des données : une déception en 4 actes'. Below the bio, there are four tabs: 'Tweets', 'Images', 'Maps', and 'Vidéos'. The 'Tweets' tab is selected, showing a single tweet from 'Jules Grandin' (@JulesGrandin) dated 29 nov. 2017. The tweet content is: 'je voyageur sur RER - Data.gouv.fr gouv.fr/fr/datasets/comptage-voyageur ;> contient le comptage voyageur effectué'. The right screenshot shows a list of users on a platform. At the top, it says 'Angeline Jolie 21 1 < favoris' and 'Bertrand Cantat 1 243 favoris'. Below that is a list of users with their names, follower counts, and profile icons: 'David Silva 7 959 favoris', 'Affaire Maëlys 1 416 favoris', and 'Hit Sale'. At the bottom of the list, there are links: '2017 Twitter - À propos', 'Confidentialité - Conditions', 'Politique de confidentialité - Cookies', and 'Informations sur la propriété'.

Challenge 2 : data quality

Or hardly exploitable...

Source

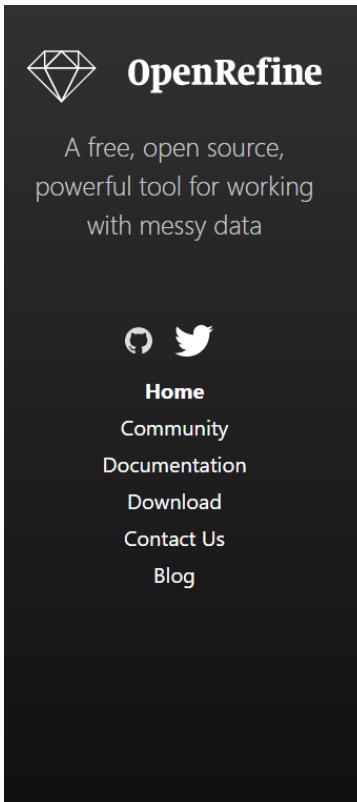
Challenge 2 : data quality

Or hardly exploitable...²

A	B	C	D	E	F	G
13						
14	Nombre de places assises.....	2 306 places				
15	(dont médiathèques centrales 1 665 places)					
16						
17	Entrées.....	982 793				
18						
19	Inscrits.....	56 821				
20						
21	Prêts.....	1 955 381				
22						
23	Ensemble des collections.....	plus d'1 000 000 de documents				
24	dont :					
25	disques et textes enregistrés	97 792				
26	vidéocassettes et DVD	42 839				
27	partitions	3 547				
28	documents adaptés aux personnes handicapées.....	10 075				
29	les collections patrimoniales :	321 398				
30	pages numérisés.....	226 300				
31						
32	Services informatiques et numérisations					
33	Accès à la vidéo à la demande,					
34	Accès au téléchargement de musique et de livres					

Source

Challenge 2 : data quality



Google News Initiative

Introduction to OpenRefine

1. Explore Data

OpenRefine can help you explore large data sets with ease. You can find out more about this functionality by watching the video below.

The image shows a screenshot of the Google Refine 2.0 software interface. The title bar says "Google Refine 2.0 - Intro...". The main area is a data grid with columns: "Contract ID", "Contractor Name", "Type of Contract", "Date of Award", "Start Date", "End Date", and "Total value o". The data grid contains 10 rows of contract information. A video player icon is overlaid on the grid, indicating a video tutorial is available.

[Download Open Refine \(+ tutorials\)](#)

Challenge 2 : data quality



Mode d'emploi

Cet outil vous permet de créer un fichier CSV en vous assurant qu'il est conforme à un schéma, c'est-à-dire que ses données sont complètes, valides et structurées.

1. Sélectionnez le schéma qui vous intéresse dans la liste déroulante, les schémas disponibles ici étant ceux référencés sur schema.data.gouv.fr.
2. Remplissez le formulaire : vous allez ainsi créer la première ligne de votre fichier CSV.
3. L'outil vous prévient d'éventuelles erreurs de validation, le cas échéant vous pouvez les corriger.
4. Une fois votre formulaire valide, les valeurs apparaissent sous la forme d'une ligne dans un tableau récapitulatif.
5. Vous pouvez alors choisir d'ajouter une ou plusieurs lignes (répétez les étapes 2 à 4) ou télécharger le fichier CSV correspondant au tableau récapitulatif.

Choisissez un schéma à utiliser :

csv-gg est un [logiciel libre](#) développé par [Etalab](#).

Attention il s'agit d'un projet expérimental. En cas de question ou de problème, vous pouvez [ouvrir un ticket ici](#) ou nous [envoyer un email](#).

Have a look to CSV GG (an Etalab initiative)

Tidy data



Tidy data Paradigm (Hadley Wickham)

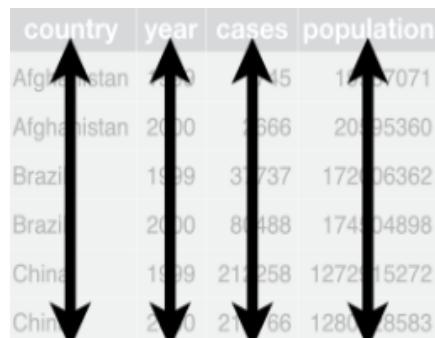
“All happy families are alike, but every unhappy family is unhappy in its own way” – Leon Tolstoi

“Tidy datasets are all alike, but every messy dataset is messy in its own way.” – Hadley Wickham

Tidy data

Tidy data principles ("données ordonnées")

- Each variable in the data set is placed in its own column
- Each observation is placed in its own row
- Each value is placed in its own cell



country	year	cases	population
Afghanistan	1990	145	187071
Afghanistan	2000	2666	2095360
Brazil	1999	31737	17206362
Brazil	2000	80488	17404898
China	1999	21258	127215272
China	2000	21666	128042583

variables



country	year	cases	population
Afghanistan	1990	145	187071
Afghanistan	2000	2666	2095360
Brazil	1999	31737	17206362
Brazil	2000	80488	17404898
China	1999	21258	127215272
China	2000	21666	128042583

observations



country	year	cases	population
Afghanistan	1990	145	187071
Afghanistan	2000	2666	2095360
Brazil	1999	31737	17206362
Brazil	2000	80488	17404898
China	1999	21258	127215272
China	2000	21666	128042583

values

See you in two weeks !

So happyyyyy together !



Thank you !

Contact : timothee.gidoin@sciencespo.fr