

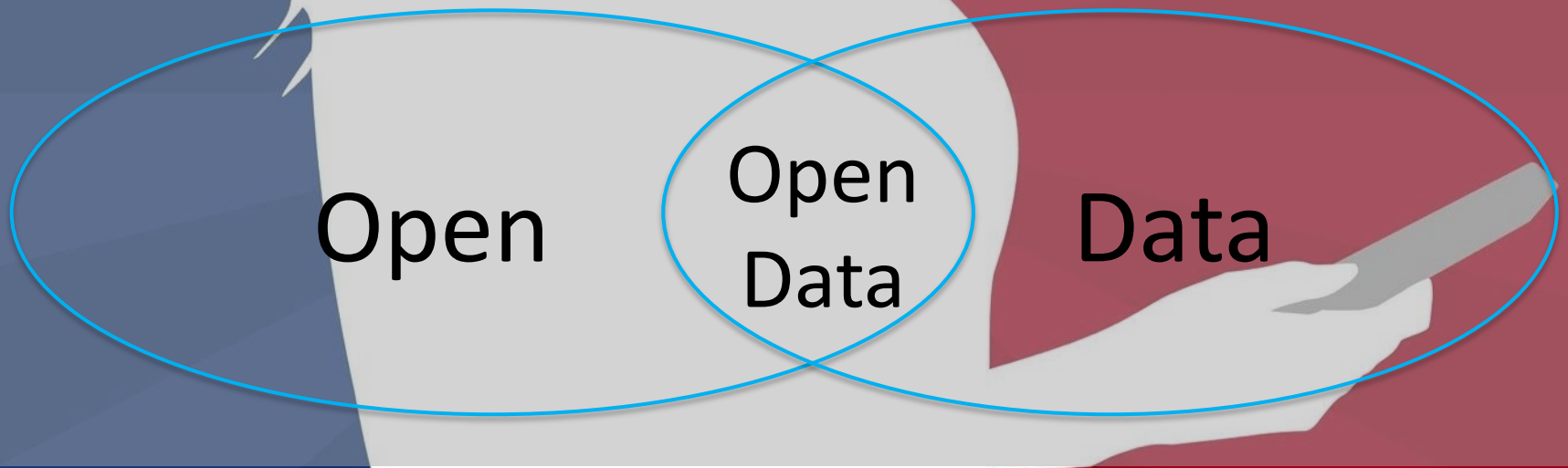
# From open data to open algorithms... and back again

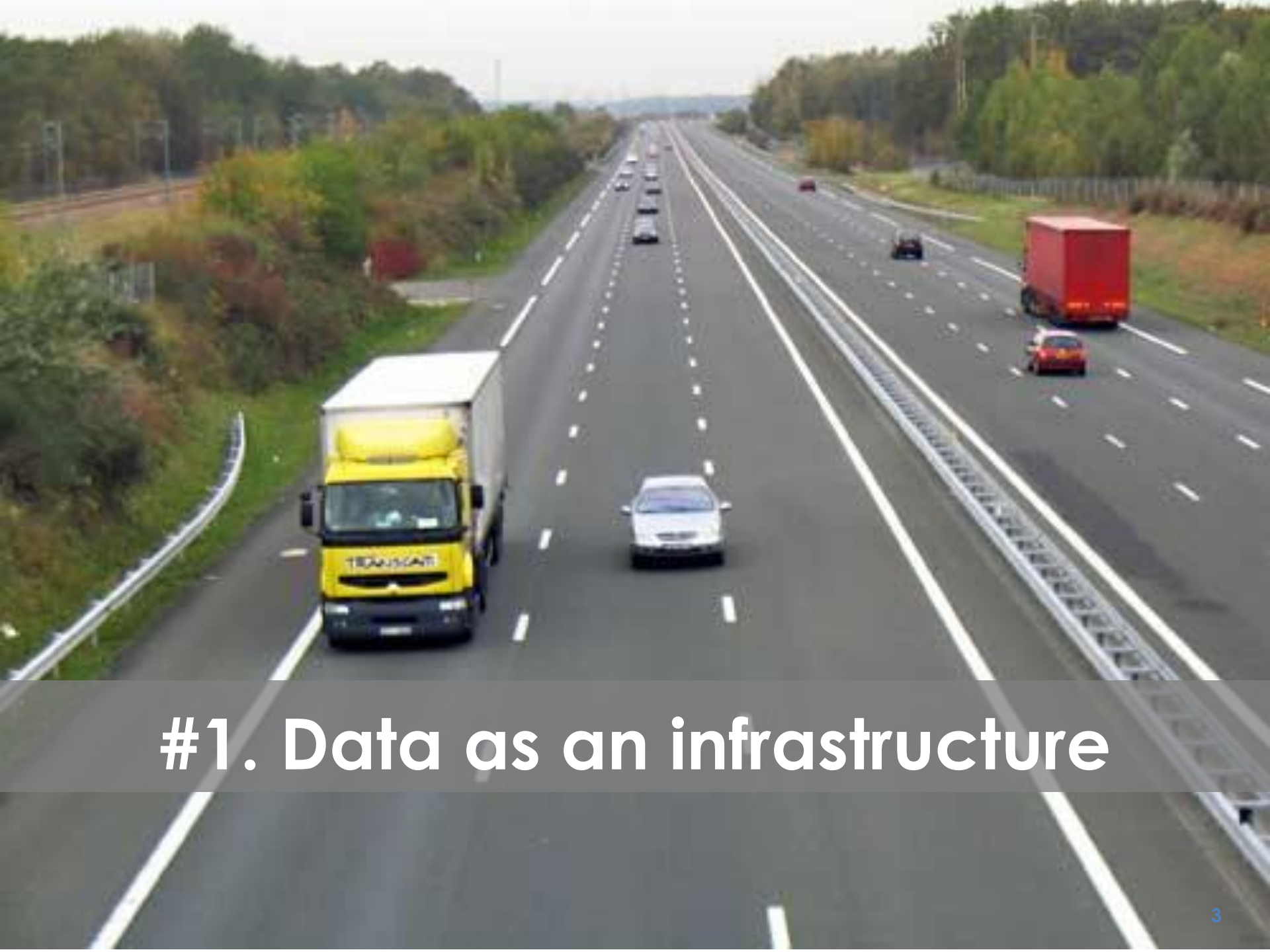
A French experience

*Internet Governance Forum*

November 14<sup>th</sup> 2018

**Etalab** : we contribute to the state digital transformation through open data and open government





# #1. Data as an infrastructure

# A new data governance for the state

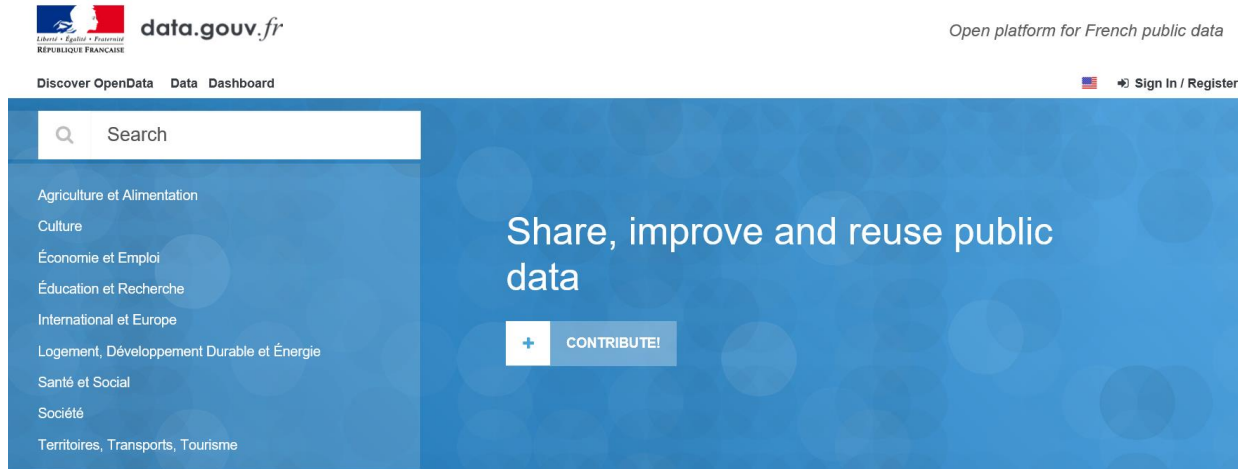


A national **chief data officer**

Producing guidelines and doctrine on data governance

**Building a community of departmental data officers**

# A national open data platform **Data.gouv.fr**



## BEST REUSE CASES



## LATEST REUSES



A **collaborative platform** open to all but especially public bodies

Which aims to **foster reuse** through :

- Curation and datavizs
- Tools to explore data
- Social features

# A strong legal framework for open data

## Law 78-753 of 17 July 1978 / French FIOA

Freedom of access to administrative documents

A few limits to such a freedom

Freedom to reuse

A regulatory authority, the CADA

## “Free of charge” Act

Free of charge reuse as a matter of principle

Very limited licence fees

## Digital Republic Act

**Open data by default** : any public organization of more than 50 employees **has to make available regularly updated databases and data with an economic, social, environmental interest** which can be communicated to everyone

**Open standards** : easily reusable and usable by an automated processing system

A **limited number of open licenses** (free of charge) to foster dissemination



# « Reference data » as a public service




## Recognizing the status of « reference data » :

- a **common reference to name** products, territories or persons ...
- with a **high economic and social impact**
- already **widely reused**

## Requiring a high degree of quality and reliability both in content and dissemination

9 data sets, including the **land, company, NGO registries**



Logo	Dataset Name	Description	Engagement Metrics
	<b>Base Adresse Nationale (BAN)</b>	La Base Adresse Nationale est une base de données qui a pour but de référencer l'intégralité des adresses du territoire	2 26 34 France Point d'intérêt
	<b>Plan Cadastral Informatisé (PCI)</b>	Le plan cadastral est un assemblage d'environ 600 000 feuilles ou planches représentant chacune une section ou une	3 2 54 France
	<b>Référentiel à grande échelle (RGE)</b>	L'État a confié à l'IGN le développement du référentiel à grande échelle (RGE). Pour ce faire, il fait appel à ses moyens propres	12 2 8 France
	<b>Base Sirene des entreprises et de leurs établissements (SIREN, SIRET)[fin le 31 janvier 2019]</b>	ATTENTION ! Ce jeu de données est remplacé par un nouveau jeu de données et prendra fin le 31 janvier 2019. La base	522 44 63 France
	<b>Registre parcellaire graphique (RPG) : contours des parcelles et îlots cultureux et leur groupe de cultures majoritaire</b>	Le registre parcellaire graphique est une base de données géographiques servant de référence à l'instruction des aides de la	16 2 7 2013–2017 France
	<b>Répertoire National des Associations (RNA)</b>	Le Répertoire National des Associations (RNA) contient l'ensemble des associations relevant de la loi 1901, à savoir toutes les	27 5 21 1901–2017 France Point

# Source codes as open data

## From a court ruling

A Paris Administrative Court ruled on 10 March 2016 that the software simulating the calculation of personal income tax was an administrative document subject to freedom of information / open data.

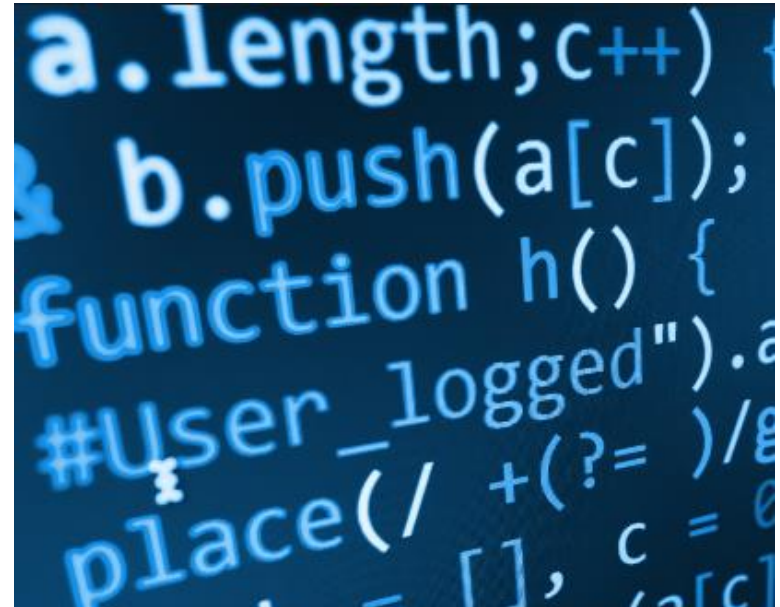
## To law

Digital Republic Act included clearly **source codes** in administrative documents

## And a state sponsored policy

#codeimpot

Open source contributing policy for public organizations





## #2. Opening public algorithms



# A new legal framework for public algorithms

## A new problem : code is law

Fears about automated decision making (ie. university affectations, tax calculation etc.)



## A new legal framework

### Digital Republic Act

Transparency rules for algorithms used to make individual decisions:

- **mention** in each decision
- **right to explanation**: the person can get detailed information on how the algorithm was implemented for his specific situation and with which data
- **online publication of the rules of the main public algorithms**

### French implementation of GDPR

**Automated decision making allowed** for administrative decisions when:

- **right to human intervention**
- transparency
- control of algorithm / **rules-based algorithms**

# A tough implementation

## **Getting better understanding of the situation**

*What is automated decision making? Which public administration use algorithmic treatment? Who is developing the algorithm? Can we access it?*

## **Making sure the existing practices are compliant with the law**

*With the new transparency framework*

*With the relevant legislation: does the code correspond to the legal text? Do we have to be more specific in the text?*

## **Developing best practices**

*Ex post:*

*Is transparency enough? Should we share the source code or more explanations ? How to best explain? Is it worth it without data?*

*Is auditing the solution? How?*

*Ex ante:*

*Where should we position the developing team?*

*How to develop algorithms that are accountable « by design »?*

# Making accountable AI through open data

## Preparing the next public algorithms generation

Deep learning benefits

Efficient private sector algorithms

→ rule-based algorithm will be challenged rapidly

## Open training datasets

Non discriminatory

Diverse

For basic AI functions: text, image

And more critical public services

## Open testing frameworks

A few projects

## Open algorithms





# Thank you !

[www.etalab.gouv.fr](http://www.etalab.gouv.fr)  
[www.data.gouv.fr](http://www.data.gouv.fr)

Maud Choquet : [maud.choquet@modernisation.gouv.fr](mailto:maud.choquet@modernisation.gouv.fr)

Amélie Banzet : [amelie.banzet@modernisation.gouv.fr](mailto:amelie.banzet@modernisation.gouv.fr)