

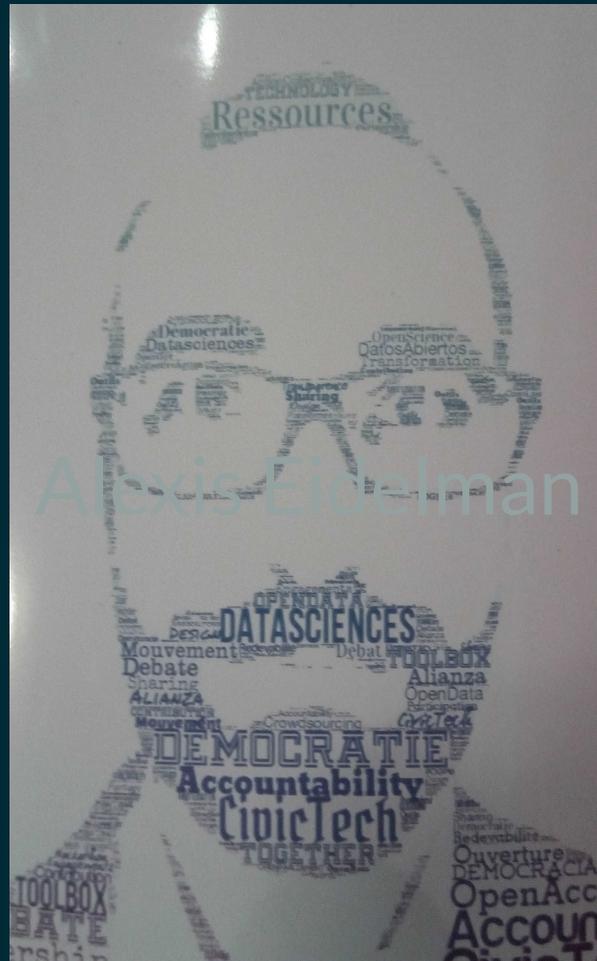
# OPEN DATA

## ENJEUX DE L'OUVERTURE DES DONNÉES

Télécharger les [slides du cours](#)

**DATA IS THE NEW OIL**

**DATA IS THE NEW SOIL**



Alexis Eidelman

# ALEXIS EIDELMAN

10 ans d'expertises dans la données

Un parcours dans les services statistiques publiques

Avec un passage à la direction du numérique

Aujourd'hui Lead Data Scientist au Ministère du Travail

# LE COURS

Une journée

évaluation : un dossier léger : Quelle opportunité  
d'ouvrir des données ?

# OPEN DATA

Aspect stratégiques, juridiques et techniques

1. Pourquoi faire de l'Open Data ?
2. Plusieurs points juridiques à avoir en tête
3. Comment ouvrir des données

# POURQUOI FAIRE DE L'OPEN DATA ?

1. Les motivations
2. Les raisons de ne pas ouvrir

# POURQUOI FAIRE DE L'OPEN DATA ?

1. Les raisons de ne pas ouvrir
2. Les motivations

# LES RAISONS DE NE PAS OUVRIR

- J'arme mes adversaires
- Je peux vendre plutôt que donner
- J'ai des choses à cacher
- J'ai peur que des gens fassent mieux que moi
- Mes données sont de mauvaise qualité
- Je ne vois pas l'intérêt
- Je ne suis pas sûr d'avoir le droit
- Je veux bien mais pourquoi le ferai-je ? (coût : serveur, communication, préparation des données)

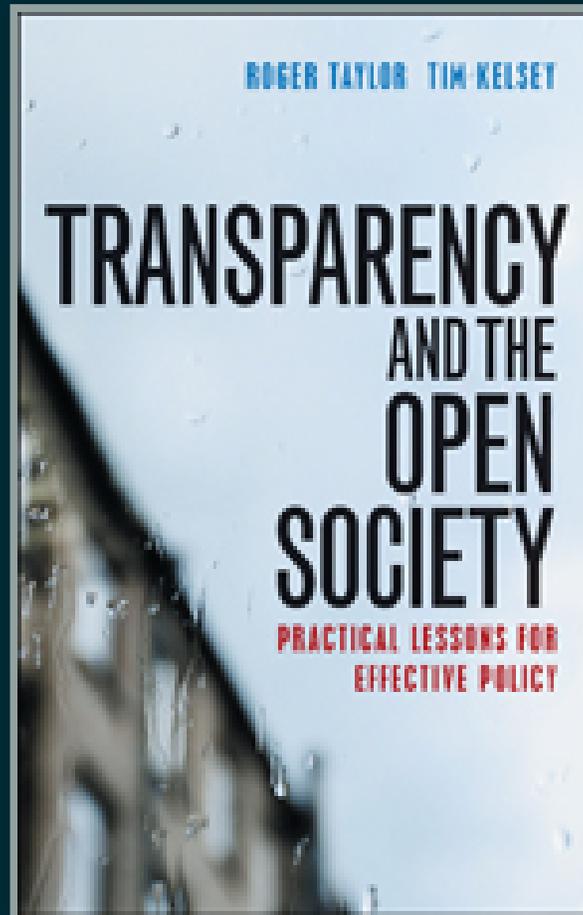
- J'arme mes adversaires
- J'ai peur que les données favorisent un comportement que je veux éviter
  - Ne pas dire au voleur, où le gendarme le cherche
  - Les données rendent plus facile l'utilisation d'un service qui m'est coûteux
  - En montrant tout, je montre mes failles et facilite les attaques

# LES MOTIVATIONS

- Comme l'Open Source ?
- pour une entreprise
- pour une administration

- Transparence : je montre ce que je fais
- Communication : je me fais connaître
- Stratégique 1 : devenir le centre d'un réseau
- Stratégique 2 : capter l'innovation
- Economique : des données pour réguler
- Ethique, politique : je n'ai rien à cacher

# TRANSPARENCY



# LA JUSTICE

un cas particulier



# LA JUSTICE

un cas particulier



Et sinon : quelles données sont ouvertes ?

# LES DONNÉES PRIVÉES

- Des données crowd-sourcées : OpenStreetMap, Wikipédia, Wikidata
- Des données privées :
  - Saemes (parking)
  - SNCF
  - TeamSky :)
  - Enel, Enedis
- Des données de recherche : PubMed, autres
- Les données des challenges

# LES DONNÉES PUBLIQUES

- Base des accidents de la route
- le fichier des équidés
- la base transparence santé
- la base nationale d'adresse
- les données des services d'inspections des restaurants

# LES DONNÉES PUBLIQUES

Des données pour réguler

- name and shame (inégalité salariale, liste noire compagnie aérienne)
- la qualité des réseaux téléphoniques
- la liste de station de recharge
- les remboursements de frais des agents publics
- la transparence pour lutter contre la corruption

# LES DONNÉES PUBLIQUES

Le service public de la donnée

- <https://www.data.gouv.fr/fr/reference>

# RÉALISER UNE OUVERTURE DE DONNÉES ?

Une entreprise ou une administration démarre en une ouverture de données.

Est-ce qu'on peut imaginer des réactions négatives ?

# Les causes possibles de mécontentement

## Les causes possibles de mécontentement

- Du Open Washing (des données payantes)

## Les causes possibles de mécontentement

- Du Open Washing (des données payantes)
- Une ouverture durable

## Les causes possibles de mécontentement

- Du Open Washing (des données payantes)
- Une ouverture durable
- Une limitation des usages

## Les causes possibles de mécontentement

- Du Open Washing (des données payantes)
- Une ouverture durable
- Une limitation des usages
- Un caractère personnel ! l'exemple d'AOL

- Les données !

- Les données !
  - Des données agrégées, peu intéressantes

- Les données !
  - Des données agrégées, peu intéressantes
  - Des données obsolètes

- Les données !
  - Des données agrégées, peu intéressantes
  - Des données obsolètes
  - Des données incompréhensibles

- Les données !
  - Des données agrégées, peu intéressantes
  - Des données obsolètes
  - Des données incompréhensibles
  - Des données partielles

- Mais surtout une ouverture de données est ratée  
quand ...

**L'OPEN DATA EST UNE TRANSFORMATION**

- Mais surtout une ouverture de données est ratée  
quand ...  
les données ne sont pas utilisées !

**L'OPEN DATA EST UNE TRANSFORMATION**

- Mais surtout une ouverture de données est ratée  
quand ...

les données ne sont pas utilisées !

**\* ANIMATIONS DE COMMUNAUTÉ**

**L'OPEN DATA EST UNE TRANSFORMATION**

- Mais surtout une ouverture de données est ratée  
quand ...  
les données ne sont pas utilisées !

**\* ANIMATIONS DE COMMUNAUTÉ**

**L'OPEN DATA EST UNE TRANSFORMATION**

et doit donc être appréhendée comme telle

# ÉLÉMENTS JURIDIQUES

- Publication et réutilisation
- Les licences
- Ce que la loi ne permet pas
- Ce que la loi permet
- Le caractère personnel, l'anonymisation et la pseudonymisation

# PUBLICATION N'EST PAS OUVERTURE

La notion de réutilisation est nouvelle.

# POURQUOI LA PUBLICATION SEULE NE PERMET PAS LA RÉUTILISATION ?

- Diffusion éparpillée (panneaux des mairies, sites des entreprises)
- Des formats variés rendant impossible le traitement automatique

En cas de données personnelles, la publication n'entraîne pas la possibilité de réutilisation.

## L'exemple de la base transparence santé.

- un lieu de publication centralisé (un site)
- un format de publication standardisé
- un format du fichier ouvert (json, csv, xml)
- des référentiels ...
- un vecteur de diffusion
- un droit de réutilisation

# LES LICENCES

Comme dans le logiciel libre, plusieurs licences peuvent être envisagées :

- La licence ouverte
- La licence ODbL (share-alike)
- Les licences spécifiques
- Les licences incompatibles avec l'Open Data

# CE QUI NE PERMET PAS

La loi Informatique et Libertés de 1978 : la définition du caractère personnel

- Des informations sensibles mais pas d'autre hiérarchie des informations
- Pas de hiérarchie des « réidentifieurs »
- Doit résister à toutes les méthodes de réidentification d'aujourd'hui et de demain
- Et d'autres éléments à connaître

# CE QUI NE PERMET PAS

La loi Informatique et Libertés de 1978 : la définition du caractère personnel

- Des informations sensibles mais pas d'autre hiérarchie des informations
- Pas de hiérarchie des « réidentifieurs »
- Doit résister à toutes les méthodes de réidentification d'aujourd'hui et de demain
- Et d'autres éléments à connaître

Renforcé par le RGPD

# CE QUI NE PERMET PAS (SUITE)

Les secrets :

- médical
- fiscal
- statistique
- etc.

# LES LOIS QUI PERMETTENT

- La loi Informatique et Libertés
- La loi d'accès aux documents administratifs. En cas de caractère personnel, trois options :
  - le consentement
  - l'anonymisation
  - la disposition législative ou réglementaire
- Des lois sectorielles
- La loi pour une république numérique : l'ouverture par défaut

# La spécificité des décisions de justice

# ELÉMENTS TECHNIQUES DE L'OPEN DATA

Tim Berners-Lee : **What we want is raw data !**

Tim Berners-Lee : **What we want is raw data !**

Mais ...

# QUEL FORMAT DE DONNÉES

- le format : csv, json, utf8
- la structure du fichier (**forme normale**)

# QUEL FORMAT DE DONNÉES

- le format : csv, json, utf8
- la structure du fichier (**forme normale**)

Rendre les données brutes

# La qualité de l'ouverture



Nettoyer les bad data ?

Mais les données ne suffisent pas, il faut des :

- explications : la donnée est une obtenue
- de la documentation
- une diffusion de qualité
- une communication avec les réutilisateurs

# DIFFUSION

téléchargement, data.gouv.fr, autres, API

# ANONYMISATION

# ANONYMISATION

Retirer le caractère personnel, anonymiser, c'est nécessairement supprimer une partie de l'information contenue dans les données.

Retirer l'identification directe ne suffit pas !  
Anonymisation

- k-anonymisation
- l-diversité

Profiter de l'anonymisation pour animer un écosystème

# PLAN RAPPORT

1. Description du sujet, du contexte

De quelle données parle-t-on ?

2. Motivations : les + et les - ?

3. Aspects juridiques

4. Aspects techniques : coûts

Quelles transformations des données ?

Quel moyen de diffusion ?

5. Votre recommandation

avec quelle licence ?

