



# DATA PUBLICA

## Ouverture des données (Open Data)

François Bancilhon  
Data Publica  
[twitter.com/fbancilhon](https://twitter.com/fbancilhon)  
[www.data-publica.com](http://www.data-publica.com)





ATA PUBLICA

# Plan

- Principes et périmètre
- Histoire et géographie
- Légal et financier
- Acteurs & marché
- Data Publica
- Applis et journalisme des données

# PRINCIPES ET PÉRIMÈTRE

# Ouverture des données, « open data »

- Mise à disposition des données publiques pour accès et réutilisation par les citoyens et les entreprises
- Type
  - Transport, cartographie, entreprises, statistiques, géographie, sociologie, écologie, environnement, économie, juridique, etc.
- Format
  - Chiffres et lettres (données structurées ou non)
  - xls, csv, xml, pdf, etc.

# Restrictions

- Secret d'état ou défense
- Protection de la vie privée
  - CNIL, anonymisation de fichiers
- Droit d'auteur
  - Données culturelles (photos, œuvre d'art, copyright, etc.)
- Exception culturelle
  - laquelle s'applique notamment aux données d'archives

# Ouverture des données :

## Motivations

- Produites avec l'argent des contribuables, elles appartiennent au contribuable
- Accessibles au citoyen pour des raisons de transparence (eGov)
- De grande valeur, l'état devrait les monétiser
- Carburant de la création d'applications innovantes et du développement d'un écosystème
- Peuvent servir à simplifier/améliorer la vie des citoyens



# Les 10 principes de la Sunlight Foundation

- Complètes
- Primaires
- Fraiches
- Accès physique et électronique
- Lisible par une machine
- Sans discrimination
- Standards ouverts
- Licence
- Permanentes
- Sans coût

# Caveat : le mieux est l'ennemi du bien

- Les raisons de ne pas ouvrir ses données sont nombreuses
- L'une est « le travail nécessaire à faire sur les données avant de les ouvrir »
- C'est un mauvais argument :
  - Ouvrez ce que vous avez
  - Si c'est bon pour vous, c'est bon pour moi

# Sérendipité

- L'ouverture des données permet dans de nombreux cas d'externaliser des travaux de l'éditeur public vers l'écosystème
- Qualité des données, intégration des données, notation des données, formatage des données
  - Exemple : les arrêts de bus britanniques
- Un nouvel argument pour faire de la donnée brute

# Les 1001 raisons de ne pas ouvrir ses données

- Elles contiennent des données personnelles
- C'est nouveau, je ne l'ai jamais fait
- Je suis légalement responsable
- Les problèmes légaux sont peu clairs
- On peut en faire mauvais usage
- On peut en faire mauvaise interprétation
- C'est cher
- C'est du travail
- Je ne suis pas prêt
- La DSI est déjà surchargée
- Mes données sont en mauvais état
- Je suis en train de réfléchir
- Je veux traiter le problème dans son ensemble

# HISTOIRE/GÉOGRAPHIE

# L'ouverture des données

2009



2010



# Ouverture des données : un mouvement mondial

- États Unis : initiative data.gov de l'administration Obama (2009)
- Royaume Uni : initiative data.gov.uk (2010)
- Union Européenne : directive européenne sur la réutilisation des données publiques (2003)
- Nombreux pays : France, Finlande, Australie, Nouvelle Zélande, Irlande du Nord, etc.

# US : data.gov

- 5 mars 2009, Vivek Kundra, DSI fédéral, met en place le programme data.gov
- 8 décembre 2009 publie la « directive de gouvernement ouvert »
  - Établit les principes du gouvernement ouvert et de l'initiative data.gov
    - Transparence, participation, collaboration
  - Fournit un agenda agressif
    - Chaque ministère à 45 jours pour publier au moins 3 jeux de données de « grand intérêt » qui n'était pas disponible auparavant
- 3 décembre 2009 publie le concept opérationnel de data.gov



# Royaume Uni : data.gov.uk

- Mis en place sous Gordon Brown
- Rattaché au Cabinet Office
- Mis sous le parrainage de Tim Berners-Lee
- Forte tendance « Web Sémantique »
- Repris par David Cameron
- De la transparence à l'efficacité
- Force chaque administration à publier toute dépense >25K£ (et les contrats associés)

# Europe

- Directive 2003/98/EC, 17 novembre 2003 sur la réutilisation des informations publiques
- Directive INSPIRE, 14 mars 2007
- EPSI Platform, information sur PSI (Public Sector Information)
- Projet LOD2 (Linked Open Data), projet FP7, Septembre 2010, 4 ans
- Share-PSI (ETSI & W3C)
- Digital Agenda (juin 2011)

# France

- 1978 : loi CADA
- 2003 : directive européenne
- 2005 : loi CADA révisée, complétée et mise en conformité avec la directive européenne
- 2007 : APIE
- 2010 : LIP
- 2011 : Etalab

# LÉGAL ET FINANCIER

# Licences

- Un domaine contrôlé par les anglo-saxons
- Problème de complexité
- Perplexité des collectivités locales
- Problème d'interopérabilité entre les licences

# Licences

- Conditions par défaut de l'APIE
- Licence d'information Publique (LIP)
- Creative Commons (CC), eg Wikipedia CC-by-sa, eg New Zealand, CC-by
- Future licence d'Etalab
- Open Database Licence (ODbL), eg OpenStreetMap
  - Tout droit de réutilisation, gratuit
  - Licence virale (des travaux dérivés doivent avoir la même licence, impossible de construire une base de données avec des licences incompatibles)
  - Notion de Base de données dérivée ou de base de données collective

# Gratuit ou payant ?

- APIE pousse au paiement
  - e.g. base de données de prix des carburants
- Le paiement est une spécificité française
- Les administrations ont du payant (cher) et du gratuit
  - e.g. IGN, DILA, INSEE, etc.
- Les industriels (établis) ne sont pas très fanas du gratuit

# Payant à quel prix ?

- Le « coût de mise à disposition »
- Une fonction du chiffre d'affaires généré par le ré-utilisateur
- Une fonction de la valeur
- Évolution
  - e.g. IGN et DILA



# La circulaire Fillon

- Circulaire du 27 mai 2011
- Le principe de base est la gratuité
- Le paiement doit être justifié, une liste des jeux de données payantes sera constituée, date limite juillet 2012.
- La liste sera agréée par le COEPIA
- Ce qui ne sera pas sur la liste sera gratuit

# ACTEURS

# Acteurs en France

- État
  - Éditeurs, APIE, Etalab
- Collectivités territoriales
  - Rennes, Paris, Nantes, Marseille, Montpellier, etc.
- Industriels
  - GFII (Groupement Français des Industries de l'Information)
- Presse
- Militants

# Éditeurs publics

- INSEE (Institut National de la Statistique et des Etudes Economiques), budget 418 M€
- DILA (Direction des Informations Légales et Administratives), budget 200M€
  - JO, BOAMP, BODAC, Legifrance, BALO, etc.
- IGN (Institut Géographique National), budget 220M€
- Ministère Écologie
- INRIA

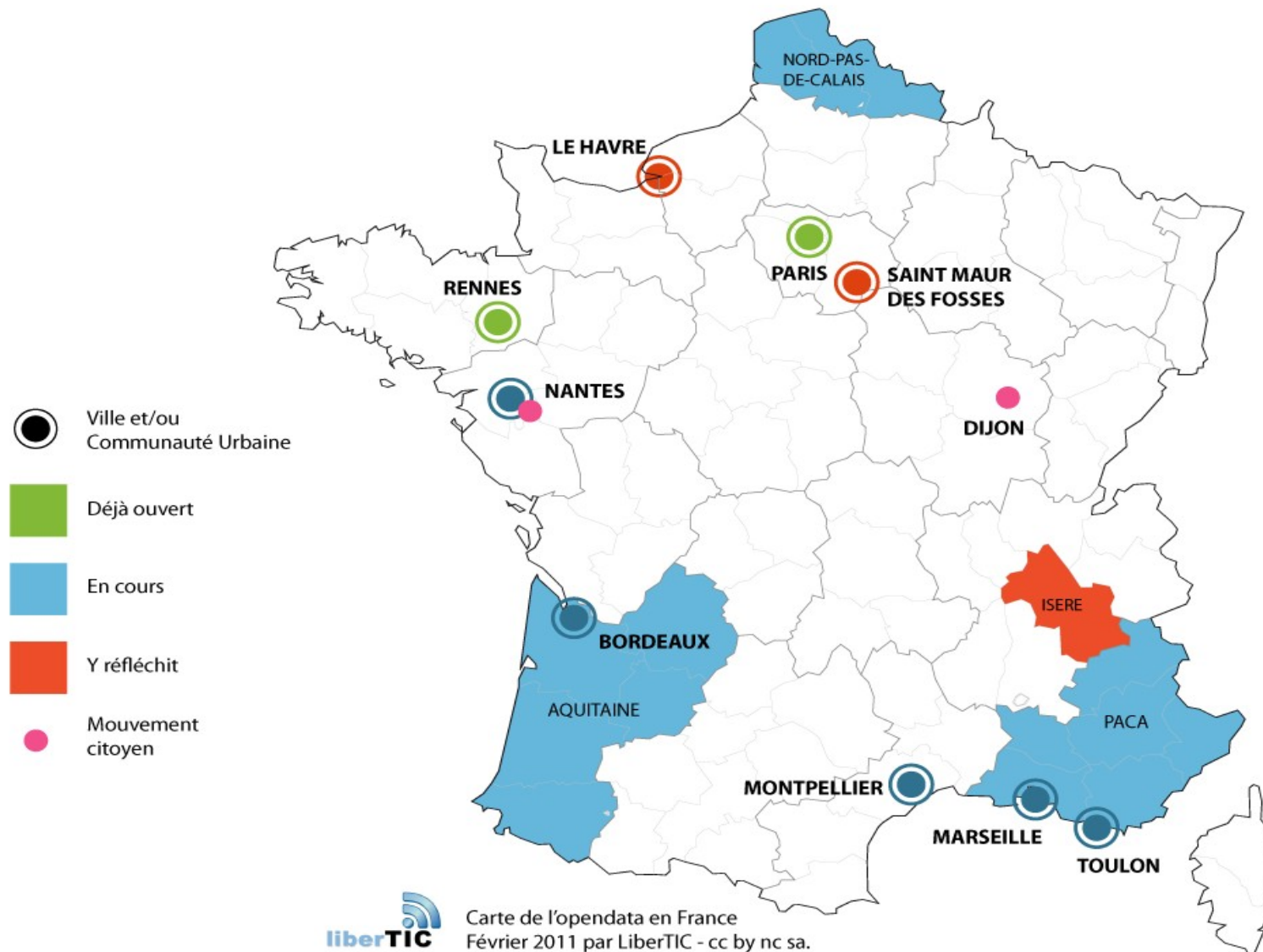
# APIE

- Agence pour le Patrimoine Immatériel de l'Etat
- Patrimoine immatériel
  - Brevets, Marques, Œuvres d'Art, Données, etc.
- « Optimiser l'impact de la gestion du patrimoine immatériel sur l'économie ;
- Tirer parti d'une meilleure valorisation des actifs pour moderniser les services publics, soutenir la conduite des politiques publiques au profit des usagers et contribuer au désendettement ;
- Prémunir l'État et les usagers contre d'éventuels risques de détournement. »

# Etalab

- Recommandation du rapport Riester
- Mission créée auprès du premier ministre en février 2011
- En charge de mettre en place data.gouv.fr
- Appel d'offre émis en mai 2011, sous administration de la DILA
- Objectif de mettre en place un annuaire des données publiques avant la fin 2011

# Collectivités territoriales



# Militants

- FING
  - Fondation Internet Nouvelle Génération
- LiberTIC
  - Nantes
- Fondapol
- Regards Citoyens
  - [NosDéputés.fr](http://NosDéputés.fr), [NosDonnees.fr](http://NosDonnees.fr)



# Acteurs internationaux

Organization	Type	Location	# DS	Business model	Started	Investors
Aggdata	Company	US	600	sells marketing data	2006	
DataMarket	Company	Iceland	1 300	subscription to a platform with tools and data	2010	
Infochimps	Company	US	5 000	resells data (50% commission)	2010	
DataMarketPlace	Company	US	30	market place for data buyers and sellers	2009	Y combinator
Factual	Company	US		subscription to a platform with tools and data	2007	Andreessen-Horowitz etc.
Pachube	Company	US		Subscription connects apps to sensors around the world	2008	
Strikelron	Company	US		web services to provide access to data and implement apps	2003	Vcs
Swivel	Company	US		platform to share and work on data (visualization)	2006	
Timetric	Company	UK	12	tools and data for statistics	2007	
CKAN	Foundation	World	1 369	crowdsourcing community		
Data.gov	Gov site	US	275 000	official US Federal Govt site		
Data.gov.uk	Gov site	UK	3 900	official UK govt site		
Amazon WS-PDS	Project	US	34	work place		
Google Public Data Explorer	Project	US	22	provides good graphical visualization tools		
ManyEyes	Project	US	4 400	IBM visualization r&d project		
Microsoft Datamarket	Project	US	10	data integration tools from the Azure platform		

# MARCHÉ

# Marché des données électroniques en France

- 1,6 Milliard d'euros annuel
- Environ 60% vient des données publiques
- 170 acteurs
- 9 verticaux
- 2 horizontaux (veille et traitement de contenu)

# Les 9 verticaux

Vertical	Exemple	Taille (M€)
Financier	Reuters	300
Presse	Press Index	250
Juridique	Francis Lefebvre	240
Solvabilité	Altarès	160
STM	Meteo France	160
Image	Sipa	60
Economique	Société.com	55
Marketing	Acxiom	55
Brevets	Reuters	25

# Les trois étages



# Data Publica

**CATALOGUE  
DES DONNEES  
ET MOTEUR  
DE RECHERCHE**

**SUPERMARCHE  
DES DONNEES**

**JEU DE  
DONNEES  
SUR MESURE**

**L'ACTU DES  
DONNEES**

# Data Publica

**CHERCHER**

**CATALOGUE  
DES DONNEES  
ET MOTEUR  
DE RECHERCHE**

**ACHETER**

**SUPERMARCHE  
DES DONNEES**

**COMMANDER**

**JEU DE  
DONNEES  
SUR MESURE**

**S'INFORMER**

**L'ACTU DES  
DONNEES**



## Données

Ici, vous pouvez voir les jeux de données disponibles, et vous pouvez faire des recherches parmi ces jeux de données.

- Agriculture & Pêche (161)
- Arts & Culture (10)
- Banque & Assurance (39)
- Climat & Météorologie (21)
- Commerce & Services (59)
- Démocratie & Elections (53)
- Economie & Finances (623)
- Education & Formation (117)
- Environnement & Energie (275)
- Géographie & Espace (21)
- Immobilier & Construction (96)
- Industrie & Production (62)
- Informatique & Communications (10)
- Population & Démographie (199)
- Presse & Médias (9)
- Santé & Nutrition (116)
- Sciences & Technologies (33)
- Sociétés & Conditions de vie (132)
- Sports & Loisirs (13)
- Sécurité & Justice (39)
- Tourisme & Voyages (20)
- Transport & Logistique (164)
- Travail & Salaires (126)
- Autres (13)

**Rechercher** Recherche avancée

Flux RSS



### Accès membre

Vous possédez déjà un compte Data Publica, connectez vous ici.

[Connexion](#)

[J'ai oublié mon mot de passe !](#)

### Inscription

Nouveau sur Data Publica ?  
Inscrivez vous maintenant, c'est facile et gratuit.

[Inscription](#)

[Blog](#)





# LES APPLICATIONS BASÉES SUR LES DONNÉES OUVERTES

# Applications mobiles

- Transport (collectif et individuel)
  - Tourisme
  - Culture et événements
  - Commerce
- 
- Au concours Mobilife 2011, sur les 12 finalistes, 7 avaient besoin de données publiques et des difficultés pour les trouver

# La vue à plus long terme

- Observer
  - Les bouchons de décembre 2010 sur la A86
  - La hausse de mortalité pendant la canicule de 2003
- Mesurer
  - La taille des Françaises
- Prévoir
  - Les ventes par point de vente
  - Le nombre de médecins
  - L'influence de la météo sur les ventes

# JOURNALISME DES DONNÉES

# Journalisme des données (Data Journalism)

- Publier des articles à partir de données publiques
  - Identification des données, croisement, extraction d'information, visualisation (graphiques)
- Un sujet très tendance
- L'avenir de la presse ?
- Tim Berners-Lee : « The Press will have to become Data savvy »

# Journalisme des données (Data Journalism)

- La presse comme intermédiaire entre les éditeurs de données et le grand public
- France
  - OWNI, Dataveyes, Le Monde Magazine, Streetpress, DataNews,
- Monde
  - Freakonomics
  - The Guardian (Simon Rogers)

# Conclusion

- Un mouvement général, mondial et puissant
- De nombreux acteurs concernés :
  - entreprises, organismes publics,
  - citoyens, presse, chercheurs, etc.
- Plusieurs dimensions
  - Technique, sociale & politique, économique
- Nous ne sommes qu'au début



# DATA PUBLICA

## Ouverture des données (Open Data)

François Bancilhon  
Data Publica  
[twitter.com/fbancilhon](https://twitter.com/fbancilhon)  
[www.data-publica.com](http://www.data-publica.com)