



# L'OpenData libère les données !

## Table des matières

1. Préambule : .....	2
2. Principes : .....	2
3. Historique : .....	2
4. Définition de « données publiques » : .....	2
5. Exemple de « données publiques » : .....	3
6. Les intérêts de l'OpenData .....	3
7. Quelques illustrations .....	4
8. Aspects Juridiques et Licences .....	5
Les Creative Commons des licences ouvertes sur mesure : .....	5
9. Accéder aux données .....	6
Les fichiers téléchargeables .....	6
Les interfaces de programmation (API) .....	6
10. Format des fichiers et des données .....	6
Les formats de fichiers .....	6
Les formats d'API .....	7
Format XML : .....	7
Format JSON : .....	7
11. Lexique de l'OpenData .....	9
12. Répertoire Français de l'OpenData .....	11

## Références :

- L'OPENDATA TOUT SAVOIR (ou presque) – LiberTIC
- Guide pratique de l'ouverture des données publiques territoriales – FING

## 1. Préambule :

Partout dans le monde, les Etats et les collectivités locales s'engagent dans le mouvement OpenData. Ce phénomène est fortement lié à l'avènement du web2.0 et nous laisse entrevoir notre futur proche à travers la notion de « smartcity » connue aussi sous l'appellation « ville2.0 ».



Figure 1 : logo OpenData

## 2. Principes :

Il s'agit de mettre à disposition des citoyens, des entreprises, des chercheurs, des associations, les masses de données numériques que les collectivités produisent : informations géographiques, statistiques, localisation des services, textes réglementaires, études, mesures, barèmes, informations transports, annuaires, etc. Et surtout d'en permettre la réutilisation par d'autres logiciels, applications et services.

## 3. Historique :

C'est le très connu organisme [W3C](#) en charge du développement du Web qui a publié les premières recommandations concernant le partage des données publiques gouvernementales en 2009 : [Publishing Open Government Data \(W3C Working Draft 8 September 2009\)](#)



Lors de sa présentation à la conférence TED en mars 2009, **Tim Berners-Lee**, le fondateur du W3C, en appelle au partage et à l'ouverture des données brutes. "Grâce au "www", on a pu mettre en ligne sur le web des milliers de documents, explique-t-il. Maintenant il faut y mettre des données". Quelles différences y a-t-il entre documents et données ? Les documents sont à lire, à analyser. Les données, si elles sont dans un format interopérable, sont directement utilisables, combinables... C'est le principe du "LinkedData" ("données liées") : les données du web de demain seront des objets relationnels.

## 4. Définition de « données publiques » :

Une donnée publique c'est une :

- Donnée collectée par les organismes publics
- Donnée non-nominative
- Donnée ne relevant pas de la vie privée
- Donnée ne relevant pas de la sécurité

En 2007, le groupe de travail « Open Government Data » définissait les 8 principes d'accessibilité des données : les données publiques sont considérées ouvertes si elles sont :

1. Complètes
2. Primaires
3. Opportunes
4. Accessibles
5. Exploitable
6. Non discriminatoires

7. Non propriétaires
8. Libre de droits

## 5. Exemple de « données publiques » :

<b>Des informations et des données</b>	
<ul style="list-style-type: none"> <li>• La description du territoire (cartes, cadastre...)</li> <li>• Des fonds documentaires (études, réglementation, statistiques...)</li> <li>• Les données de la décision publique (projets, enquêtes, délibérations, subventions...)</li> <li>• Le fonctionnement des réseaux urbains (eau, énergie, transports, logistique, télécoms...)</li> <li>• La localisation et les horaires d'ouverture des services et des commerces</li> </ul>	<ul style="list-style-type: none"> <li>• L'occupation des ressources et des capacités (voirie, bâtiments, espaces, parkings...)</li> <li>• Des mesures (environnement, trafic...)</li> <li>• Des événements (culture, sports...)</li> <li>• Des informations touristiques, culturelles, des données d'archives</li> <li>• Les flux urbains (circulation...)</li> <li>• Des données de surveillance...</li> </ul>
<b>Des applications et des services</b>	
<ul style="list-style-type: none"> <li>• Des systèmes d'information géographiques (SIG)</li> <li>• Des modèles (représentation, prévision...)</li> <li>• Des applications permettant de calculer des droits (sociaux, etc.)</li> </ul>	<ul style="list-style-type: none"> <li>• Des applications propres à chaque "métier" public</li> <li>• Des applications transversales : identification, localisation, paiement, sécurisation</li> <li>• Des services de paiement, de billetterie...</li> </ul>
<b>Des infrastructures et des espaces</b>	
<ul style="list-style-type: none"> <li>• Télécommunications</li> <li>• Capteurs urbains</li> <li>• Canaux d'information : écrans, hauts-parleurs, journaux, panneaux, sites web, écrans publics, autres canaux numériques...</li> <li>• Eclairage public</li> </ul>	<ul style="list-style-type: none"> <li>• Bâtiments</li> <li>• Espaces publics : voirie, places publiques, façades, espaces publics numériques, maisons des associations, lieux culturels, festifs et sportifs...</li> </ul>

## 6. Les intérêts de l'OpenData

- plus de transparence et donc plus de démocratie et de confiance
- plus d'information et donc plus de compréhension et de participation
- plus d'implications et donc plus d'échanges et d'amélioration
- Source d'Innovation par la transversalité et la superposition d'informations : les données publiques permettent à des professionnels ou des bénévoles de s'en emparer pour créer de nouvelles informations, de nouvelles connaissances et de nouveaux usages

## 7. Quelques illustrations

John Snow est pour ainsi dire un précurseur de l'OpenData. En effet, en 1854, il superposa sur une **carte** les **adresses des victimes** du choléra et constata qu'elles coïncidaient avec les **accès aux points d'eau**. Il découvrit l'origine des épidémies de choléra.



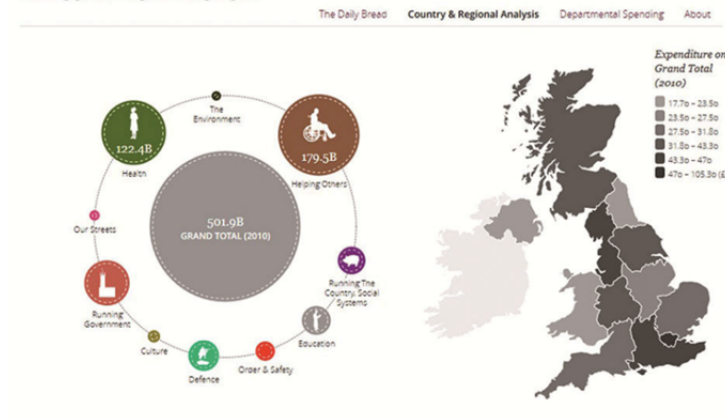
**Données + Données + Cartographie = Découverte**

Ce croisement pertinent de données est la base de toutes les applications basées sur l'OpenData.

Plus actuel, un exemple de démarche initiée par la **société civile** : « Where does my money go ? »

### WHERE DOES MY MONEY GO?

Showing you where your taxes get spent



Ce site développé par l'Open Knowledge Foundation permet de visualiser la répartition des budgets publics britanniques. Les données proviennent de la Country Regional Analysis (CRA).

Quelques exemples d'applications de services développées par des **start-up** autour des données publiques ouvertes :

- Statiophone est une application permettant de connaître en temps réel les places disponibles dans les parkings publics à Nantes.
- Home'n'go est un site qui permet de sélectionner des annonces immobilières d'agence partenaires en les croisant avec des données publiques selon plusieurs critères : géolocalisation, statistiques du prix au m2, taux d'imposition local, données démographiques, établissements scolaires ou de santé à proximité



## 8. Aspects Juridiques et Licences

Le cadre légal propre à l'exploitation des données publiques est défini par la loi du 17 juillet 1978. Elle reconnaît à toute personne un droit d'accès et de réutilisation des informations publiques. L'organisme en charge de la veille au respect de la loi sur l'accès aux documents administratifs est la CADA Commission d'Accès aux Documents Administratifs.

Afin que les données publiques soient accessibles au plus grand nombre, elles devraient être sous licence gratuite :

- C'est vrai dans beaucoup de pays (USA, UK, Nouvelle-Zélande, Australie, Espagne ...) : aux Etats-Unis toute création d'un organisme public tombe automatiquement dans le domaine public
- C'est beaucoup moins vrai en France ! 2 exemples parmi tant d'autres : pour accéder aux immatriculations ou au prix de l'essence : il faut payer une licence d'accès.

En France, c'est l'Agence du Patrimoine Immatériel de l'Etat, l'APIE, qui est chargée notamment de créer le cadre de réutilisation des données publiques. Les modèles de licence développés :

- Licence pour une livraison unique d'informations publiques
- Licence avec livraison successive des informations

Les modèles développés par l'APIE laissent une part importante aux licences payantes.

### Les Creative Commons des licences ouvertes sur mesure :

L'auteur décline les options pour définir sa licence et peut l'indiquer sous forme de visuel tel que:



Désignation complète du contrat	Terme abrégé désignant la licence	Symboles désignant la licence				Type de licence
Paternité	CC-BY					<a href="#">Licence libre non copyleft</a>
Paternité Partage des conditions initiales à l'identique	CC-BY-SA					<a href="#">Licence libre copyleft</a>
Paternité Pas de modification	CC-BY-ND					<a href="#">Licence de libre diffusion</a>
Paternité Pas d'utilisation commerciale	CC-BY-NC					<a href="#">Licence de libre diffusion</a>
Paternité Pas d'utilisation commerciale Partage des conditions initiales à l'identique	CC-BY-NC-SA					<a href="#">Licence de libre diffusion</a>
Paternité Pas d'utilisation commerciale Pas de modification	CC-BY-NC-ND					<a href="#">Licence de libre diffusion</a>

**Paternité [BY] (*Attribution*)** : l'œuvre peut être librement utilisée, à la condition de l'attribuer à l'auteur en citant son nom.

**Pas d'utilisation commerciale [NC] (*Noncommercial*)** : le titulaire de droits peut autoriser tous les types d'utilisation ou au contraire restreindre aux utilisations non commerciales (les utilisations commerciales restant soumises à son autorisation).

**Pas de modification [ND] (*NoDerivs*)** : le titulaire de droits peut continuer à réserver la faculté de réaliser des œuvres de type dérivées ou au contraire autoriser à l'avance les modifications, traductions.

**Partage des conditions initiales à l'identique [SA] (*ShareAlike*)** : le titulaire des droits peut autoriser à l'avance les modifications ; peut se superposer l'obligation (SA) pour les œuvres dites dérivées d'être proposées au public avec les mêmes libertés (sous les mêmes options Creative Commons) que l'œuvre originale.

## 9. Accéder aux données

La diffusion des données est une étape clé du processus de partage des données publiques. Face à ces problématiques, le producteur de données doit respecter un certains nombres de règles.

- Les données doivent être documentées (les données des données = métadonnées)
- Les données doivent respecter une sémantique (exemple de la date : quel calendrier sert de repère).
- Les données doivent respecter une syntaxe.

Le fournisseur de données, lui, doit s'assurer de l'interopérabilité de son accès. Pour cela, Il existe deux façons de donner accès à des données sur le web : le téléchargement d'un fichier, ou l'accès direct à la source de la donnée :

### Les fichiers téléchargeables

L'accès direct à un fichier contenant les données est souvent la méthode la plus simple car les outils informatiques proposent généralement une fonction d'export des données sous forme de fichier. Elle est cependant moins bien adaptée lorsque les données changent souvent et qu'il y en a beaucoup : cela oblige les réutilisateurs à recharger fréquemment l'ensemble du fichier et peut être consommateur en ressources système et réseau.

### Les interfaces de programmation (API)

Une "interface de programmation" (API en anglais) va permettre à une application "demandeuse" (celle du réutilisateur, le client) de demander à l'application "fournisseuse" (celle du détenteur des données "sources", le serveur) les informations dont elle a besoin. Les grands acteurs du numérique en général et du web particulier, comme Google, Amazon, Facebook, etc., ont établi leur succès grâce à leurs APIs : leurs données sont en quelque sorte "réutilisables", sans que leurs bases de données ne quitte jamais leurs ordinateurs.

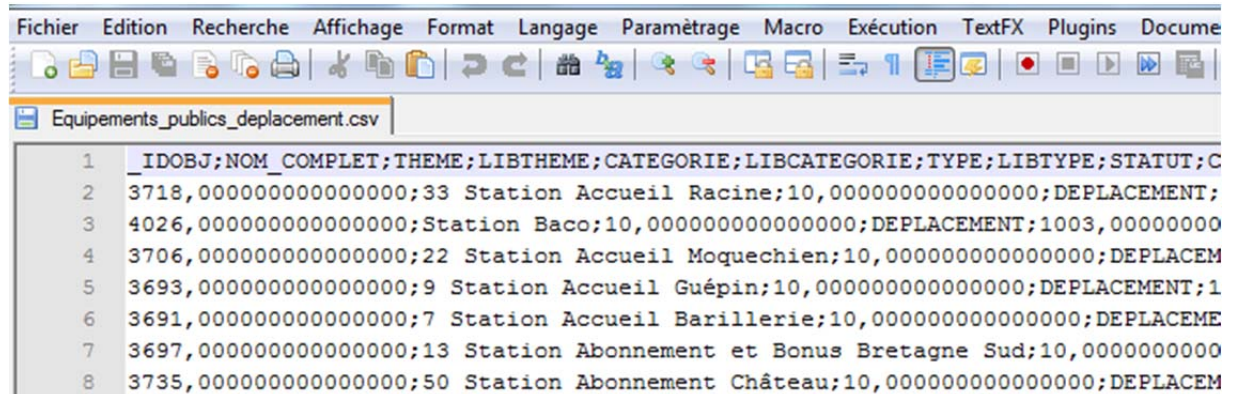
## 10. Format des fichiers et des données

### Les formats de fichiers

Les formats ouverts garantissent une meilleure réutilisabilité des données.



- Le format CSV (Comma Separated Values) en français, « valeurs séparées par des virgules ». Il s'agit d'un format de fichiers ouvert qui permet de stocker les données d'un tableau. Chaque ligne du fichier correspond à une ligne du tableau ; les colonnes sont en général séparées par des virgules. On peut très bien remplacer les virgules par des tabulations ou tout autre caractère. Ce format est utilisé pour échanger de manière interopérable des données de tableurs, bases de données, annuaires, *etc.* entre logiciels différents et/ou plateformes différentes.



- Certains formats de fichiers propriétaires peuvent cependant être acceptables. Par exemple, le format .xls (tableur Excel de Microsoft), peut éventuellement convenir.

## Les formats d'API

Dans l'ensemble des API web disponibles à travers le monde, les formats ouverts XML et JSON sont les formats de référence.

### Format XML :

XML (entendez *eXtensible Markup Language* et traduisez *Langage à balises étendu*, ou *Langage à balises extensible*) en tant que telle doit être vu comme un outil permettant de définir un langage (on dit alors qu'il s'agit d'un métalangage), permettant de créer des documents structurés à l'aide de balises (markup). Une balise est une chaîne de caractère du type : <balise>

Ainsi, un **document XML**, c'est-à-dire le fichier créé en suivant les spécifications de la *norme XML* pourra par exemple ressembler à cette arborescence :

```
<annuaire>
  <personne class = "etudiant">
    <nom>Pillou</nom>
    <prenom>Jean-Francois</prenom>
    <telephone>555-123456</telephone>
    <email>webmaster@commentcamarche.net</email>
  </personne>
</annuaire>
```

### Format JSON :

JSON (JavaScript Object Notation – Notation Objet issue de JavaScript) est un format léger d'échange de données. Il est facile à lire ou à écrire pour des humains. Il est aisément analysable ou générable

par des machines. JSON est un format texte complètement indépendant de tout langage, mais les conventions qu'il utilise seront familières à tout programmeur habitué aux langages descendant du C. Ces propriétés font de JSON un langage d'échange de données idéal.

Un exemple simple, définition d'un menu (soit un objet composé de membres qui sont un attribut et un tableau lequel contient d'autres objets, les lignes du menu).

```
{
  "menu": "Fichier",
  "commandes": [
    {
      "title": "Nouveau",
      "action": "CreateDoc"
    },
    {
      "title": "Ouvrir",
      "action": "OpenDoc"
    },
    {
      "title": "Fermer",
      "action": "CloseDoc"
    }
  ]
}
```

L'équivalent XML:

```
<?xml version="1.0" ?>
<root>
  <menu>Fichier</menu>
  <commands>
    <item>
      <title>Nouveau</value>
      <action>CreateDoc</action>
    </item>
    <item>
      <title>Ouvrir</value>
      <action>OpenDoc</action>
    </item>
    <item>
      <title>Fermer</value>
      <action>CloseDoc</action>
    </item>
  </commands>
</root>
```



## 11. Lexique de l'OpenData

### **API (Application Programming Interface)**

Désigne une interface de programmation fournie par un programme informatique pour permettre à d'autres programmeurs d'en utiliser certaines fonctions ou données. A travers le développement d'API, il s'agit pour un programmeur de développer une fois pour toute une application contenant tous les outils de base nécessaires, et de la mettre à disposition de la communauté des développeurs pour leur faciliter le travail.

Par exemple, les API de Google permettent à d'autres développeurs d'utiliser simplement, au sein de leur propre application ou service, des fonctions telles que la recherche ou les cartes de Google maps. C'est par l'intermédiaire d'API (en l'occurrence, des requêtes plus ou moins préformantes) que des sites extérieurs peuvent obtenir des données actualisées disponibles sur des sites dont ils ne sont pas les éditeurs, par exemple pour interroger une grande base de données ou récupérer des informations temps réel sur les transports.

Les API peuvent être mises à disposition de manière ouverte (accessible à tous) ou réservée à quelques acteurs, gratuite ou payante.

### **Crowdsourcing (enrichissement, approvisionnement par la "foule")**

Désigne le fait d'utiliser l'intelligence, la créativité, le savoir-faire d'un grand nombre de personnes (en l'occurrence des internautes) pour enrichir, alimenter un service, une application, un site web, la conception d'un produit, etc. L'idée sous-jacente est que la participation d'un grand groupe de personnes peut produire des connaissances, des contenus, une "œuvre collective", dont la valeur dépasse de loin celle que pourrait produire n'importe quel participant individuel.

### **Data journalism (Journalisme de données)**

Désigne une nouvelle façon de faire de l'investigation journalistique en se basant sur l'analyse de données et (souvent) la représentation visuelle. Le journaliste utilise des bases de données comme sources et en déduit des connaissances, des corrélations ou des intuitions qui ne seraient pas accessibles par les méthodes traditionnelles de l'enquêtes journalistique. Même si l'article reste la composante de base, l'illustration des idées par l'image graphique, un schéma, une carte, etc., prend une place plus importante

### **Linked Data (les données liées)**

Désigne une approche du web poussée par les tenants du "web sémantique" qui encourage à décrire toutes les données d'une manière lisible par des ordinateurs, et à les lier entre elles en décrivant leurs relations, ou encore, en facilitant leur mise en relation : ainsi, deux informations produites par des acteurs différents, mais correctement décrites comme faisant référence à un même territoire, pourront être "liées" même si leurs producteurs n'y auraient jamais pensé. Les Linked Data sont liées les unes avec les autres à l'aide de certains "formats" (propres au web sémantique : RDF, OWL, SIOC...) lisibles par un ordinateur. Celui-ci devient alors capable de les associer de manière automatique. Tim Berners-Lee, "père du web" et champion du Linked Data, a contribué à la mise en place du portail gouvernemental britannique des données publiques, data.gov.uk.

**Interopérabilité**

Désigne la capacité que possède un produit ou un système, dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs et ce sans restriction d'accès ou de mise en œuvre.

**Métadonnée**

Désigne un ensemble structuré d'informations servant à définir ou décrire une ressource (donnée, document...) quel que soit son support (papier ou électronique). S'agissant d'un livre par exemple, le titre, le nom d'auteur, l'éditeur, l'année de publication en constituent des métadonnées.

Les métadonnées sont donc "des données sur les données", qui permettent à un individu ou un ordinateur d'en comprendre le sens et l'organisation. Elles sont à la base des techniques du web sémantique.

**Mashup (combiner, mélanger ensemble)**

Designe une pratique consistant à combiner différentes sources (le plus souvent externes) d'informations ou de données afin d'alimenter son propre site web. De plus en plus de sites web (Google, Yahoo!, Amazon, eBay...) proposent gratuitement des API, afin d'encourager la communauté des développeurs à créer des mashup utilisant leur contenu.

**Web 2.0**

Il n'existe pas de définition stable. L'expression a été proposée en 2004 par l'éditeur Tim O'Reilly en 2004 pour décrire une certaine évolution du web fondée sur quatre piliers :

- Le recentrage de l'information et des services autour de la personne ;
- Une facilitation considérable de l'intervention des utilisateurs dans la production et l'échange de textes, images et vidéos ;
- Des interfaces standards permettant de partager et d'agencer de diverses manières informations et services ;
- Et une expérience plus "sociale" du web par laquelle l'innovation et la valeur proviennent, dans une large mesure, de l'interaction entre les utilisateurs.

Les services et plateformes emblématiques du web2.0 sont Wikipedia, eBay et Craigslist, Youtube, Facebook et Twitter.

**Web<sup>2</sup> ("web squared", le web puissance 2) :**

L'expression est proposée par Tim O'Reilly et John Battelle, pour décrire le développement non plus arithmétique mais exponentiel des possibilités du web. Le web "1.0" reliait des documents, le web "2.0" reliait des gens, le web au carré relie des informations individuelles, des données. Pour certains, cette expression un peu ad hoc désigne une phase intermédiaire avant le "web sémantique", pour d'autres, les deux expressions désignent un peu la même chose.

**Web sémantique ou web des données :**

Désigne un ensemble de technologies visant à rendre l'ensemble des ressources du web (documents et "grains" élémentaires d'informations) accessible,, intelligibles et utilisables par des programmes et agents logiciels, grâce à un système de métadonnées. Les machines pourront ainsi traiter, relier et combiner automatiquement un certain nombre de données.

## 12. Répertoire Français de l'OpenData

### **APIE – Agence du patrimoine immatériel de l'Etat**

L'agence, créée en septembre 2007, est un service à compétence nationale, rattaché conjointement au directeur général du Trésor et de la politique économique et au directeur général des Finances publiques. Ses missions sont :

- De sensibiliser les acteurs publics aux enjeux attachés aux actifs immatériels de l'État ;
- de faciliter la prise en compte de ces enjeux et l'adoption de nouveaux modes de gestion au sein de l'administration ;
- d'accompagner les administrations et les établissements publics pour la définition de nouvelles stratégies et la conduite de projets concrets de valorisation de leur patrimoine immatériel.

Depuis sa création, l'APIE mène des actions concrètes pour aider les opérateurs économiques à mieux accéder aux informations publiques réutilisables et pour sensibiliser et accompagner les administrations dans la mise à disposition de leurs informations.

L'objectif est de favoriser l'émergence d'un cadre pratique, juridique et financier homogène de nature à faciliter la réutilisation.

### **Cada - Commission d'accès aux documents administratifs**

La CADA est une autorité administrative indépendante dont le rôle est consultatif. Elle est chargée de veiller au respect de la liberté d'accès aux documents administratifs et aux archives publiques. Elle émet des avis lorsqu'elle est saisie par une personne qui rencontre des difficultés pour obtenir la communication d'un document administratif ou pour consulter des documents d'archives publiques. La saisine de la CADA pour avis est un préalable obligatoire à l'exercice d'un recours contentieux.

### **Fing**

La Fing est une association dont la vocation est de produire et partager des idées neuves pour anticiper les transformations numériques. Elle aide les grandes entreprises et les start-ups, les territoires et les décideurs politiques, les chercheurs, les créateurs, les innovateurs sociaux... à anticiper les opportunités et les risques associés aux technologies, à leurs usages et au système d'innovation qui les accompagne. L'ouverture des données publiques est l'un de ces projets.

### **Regards citoyens**

Le collectif RegardsCitoyens.org est une association constituée de citoyens de tous âges et régions mus par un désir commun de proposer un accès simplifié au fonctionnement de nos institutions démocratiques. Regards Citoyens a notamment créé le portail "Nos Données", qui vise à apporter plus de visibilité aux données publiques librement accessibles à chacun

### **LiberTIC**

LiberTIC est une association nantaise créée en décembre 2009 suite au premier Open Data Camp de Paris. Elle a pour objectif de promouvoir l'ouverture des données publiques, l'e-démocratie, le gouvernement 2.0 et d'accompagner le territoire ligérien dans le développement et l'utilisation d'outils innovants. Parallèlement, Libertic participe au mouvement de rapprochement de l'Economie sociale et solidaire et des TIC afin de créer des synergies au niveau local.