

Article sur Open Data,
inspiré par la lecture du livre de Simon Chignard ¹

Définition

fin 2011, le gouvernement français lançait une initiative sur les données ouvertes, à la suite des initiatives gouvernementales des Etats-Unis (Obama) et Britannique.

Le concept de données publiques ouvertes est formalisé en 2007 par Tim O'Reilly (qui a intronisé le terme Web 2.0) et Lawrence Lessig (professeur de droit à Stanford).

Les huit principes de l'Open Government Data sont exposés sur le site www.opengovdata.org en voici la traduction extraite de ¹

1. **Données complètes** : toutes les données publiques doivent être ouvertes, sauf exception : les données susceptibles de porter atteinte à la vie privée des individus (données nominatives : protection de la CNIL en France) ou à la sécurité du territoire (lois liées à la lutte anti-terroriste) ;
2. **Données brutes** : les données doivent être mises à disposition sous leur forme première avant tout travail d'agrégation ou de mise en forme ;
3. **Données à jour** : les données sont publiées aussi vite que possible pour préserver leur fraîcheur et leur valeur ;
4. **Données accessibles à tous** , pour tous les usages ;
5. **Données dont le traitement peut être automatisé** : les données sont structurées afin de permettre un traitement automatisé ;
6. **Accès non discriminant** : il n'est pas nécessaire de s'inscrire ou de s'enregistrer préalablement pour accéder aux jeux de données.
7. **Format non propriétaire** : les données sont mises à disposition dans des formats non-propriétaires. Ici , il ne faut pas raisonner forcément fichiers, il y a la variante de données mises à disposition sous forme d'API, par exemple horaires / retards dans le domaine du transport. Les puristes indiqueront que l'API sera souvent propriétaire.
8. **Licence libre** : les données ne doivent pas être soumises à des droits d'auteurs ou des droits de propriété intellectuelle. Des restrictions raisonnables à ce principe sont autorisées.

De plus il doit exister un contact pour répondre aux utilisateurs essayant d'utiliser la donnée, un contact pour répondre aux plaintes sur des violations aux principes ci-dessus.

Un juridiction ou une administration doit valider que les principes sont appliqués de manière appropriés².

Origines

Historiquement la lutte pour la transparence et l'ouverture des données remonte assez loin au siècle dernier. Par exemple le mathématicien Norbert Wiener après la seconde guerre mondiale se focalise sur ces questions de société, il milite pour l'ouverture de l'information en déclarant que le secret et la fermeture ont rendus possible Hiroshima et Auschwitz, l'ouverture et la transparence permettront à l'avenir d'éviter ces fléaux. Il refusera de travailler sur les travaux conduisant à l'arme nucléaire. Cette idée de la transparence comme outil démocratique est reprise aujourd'hui par de nombreux mouvements de la société civile.

1

1 Simon Chignard. Open Data. Comprendre l'ouverture des données publiques. Editions Fyp. 2012.

2 An administrative or judicial court must have the jurisdiction to review whether the agency has applied these principles appropriately.

Au royaume-Uni en 2009 Gordon Brown nomme deux conseillers pour mettre en place data.gov.uk
Tim Berners Lee (« inventeur du web ») et Nigel Shadbolt (chercheur en informatique)

Les courants de pensées et familles

Simon Chignard classe les courants de pensée de l'Open Data en trois familles :

1. La famille libérale d'origine anglo-saxonne. Le marché est au centre des préoccupations, la société civile fait pression sur le secteur public pour la promotion des acteurs économiques. Finalité marchande au mouvement. Dans cette tendance association Ifrap³.
2. La famille libérale-libertaire (analogue à un Richard Stallman dans le domaine de l'Open Source. Ouverture totale et gratuité revendiquées.
3. La famille participative. Mise en place d'une coproduction citoyens-acteurs publics. Associations Fing, liberTIC à Nantes.

Démarche

Il y a deux temps dans la démarche :

1. Rendre public les données
2. Engager de nouvelles relations avec les usagers autour de ces informations (famille participative)

Aspects juridiques

Loi **CADA** qui depuis 1978 donne aux citoyens l'accès aux documents administratifs, donc à des jeux de données. Avec des dérogations pour les opérateurs de l'état, le militaire et les données culturelles.

<http://www.cada.fr/>

CNIL . Pour protéger le citoyen sur le recueil de données personnelles. En conséquence les données de l'Open Data sont souvent (toujours?) anonymisées, et les effets de seuils pris en compte (pour ne pas identifier un individu qui serait le seul dans sa catégorie en recoupant plusieurs jeux de données).

Réutilisation de la donnée

au moins quatre classes de réutilisations possibles :

- la **consultation** : un internaute accède directement à une donnée en ligne, pour prendre une décision (quelle compagnie aérienne est la plus ponctuelle ?), recueillir une information (comment évolue les effectifs dans le collège de mon quartier ?), ...
- la **médiation** : les données sont retravaillées par un tiers pour les rendre accessibles au grand public, c'est par exemple le champ du datajournalisme ou de la visualisation de données,
- l'**application** : un développeur utilise les données ouvertes pour construire et proposer un service ou une application mobile,
- la **réutilisation spécialisée** : certaines données sont utilisées comme de la matière première qui rentre dans un process... ce sont des «intrants» pour des entreprises ou des organisations.

Ce **dernier type de réutilisation** mérite d'être détaillé, car il est souvent le plus méconnu.

L'un des fichiers techniques les plus populaires (en nombre de téléchargements) sur les plateformes open data concerne les **nomenclatures des voies et les adresses**. Pour le commun des mortels, ce fichier qui permet notamment d'associer une adresse postale (12, rue des Marronniers, Paris) à des

coordonnées géographiques, ne présente guère d'intérêt. Mais pour les entreprises qui font du e-commerce, de la livraison ou tout simplement qui veulent requalifier leur fichier clients en fonction de critères géographiques, **cette donnée a de la valeur.**

Guide du ré-utilisateur débutant « Open Data pour les nuls »: en 4 étapes :

1. La définition des objectifs et de la problématique
2. Le premier recensement des données disponibles et l'appréhension des éléments de contexte
3. Le traitement des données et le raffinement des objectifs initiaux
4. Le choix des modes de représentation des résultats

La suite (dans le contexte français)

après le changement de majorité présidentielle et parlementaire, le sujet n'est pas encore traité, au regard des 20 questions décrites dans ce billet :

<http://libertic.wordpress.com/2012/06/17/open-data-20-questions-qui-se-posedent-au-gouvernement-francais/>

- Faut-il définir juridiquement le terme « donnée ouverte » ?
- L'ouverture des données publiques doit-elle devenir une obligation légale ?
- Faut-il imposer l'usage de certaines licences ?
- L'Etat doit-il s'impliquer dans la standardisation des démarches ?
- Faut-il adapter le code des marchés public ?
- L'open data doit-il s'étendre au texte ?
- Les données culturelles doivent-elles être soumises au droit d'accès ?
- Faut-il revoir le droit d'auteur des agents publics ?
- Faut-il moderniser la CADA ?
- Faut-il réaffirmer l'obligation de la publication des répertoires ?
- Faut-il focaliser le développement de l'ouverture sur les données disponibles ?
- Faut-il définir une liste des données d'intérêt général soumises à publication obligatoire ?
- Faut-il étendre le devoir de publication aux organismes privés et civils ?
- Faut-il modifier le statut d'Etalab ?
- Faut-il upgrader data.gouv.fr ?
- Data.gouv.fr a-t-il vocation à accueillir les données de toutes les collectivités ? (vérifier la légende comme quoi etalab ne référence que les mairies UMP:-))
- Faut-il faire des données un outil de débat démocratique ?
- La notion de Gouvernance Ouverte doit-elle faire partie du vocabulaire politique ?
- La France doit-elle rejoindre l'Open Government Partnership ?
- Faut-il être cohérent dans ses positionnements ?

Conclusion

Le domaine de l'Open Data est en pleine croissance et présente aussi bien des aspects techniques innovants (mobilité, BigData, Web Sémantique) que des nouveaux usages à inventer. Cela nous évoque les périodes de croissances de l'internet/e-commerce dans les années 2000, avant l'éclatement de la bulle.

Il y a déjà une petite liste de startup françaises sur le sujet, d'où des emplois à la clé.

Références

Site associé au livre de Simon Chignard : <http://www.donneesouvertes.info>

Les événements auxquels j'ai assisté :

20 juin : Dataconnexion étape 2. Organiser par Etalab et sponsor technique Google.

5 juillet Datapero avec la SNCF chez Qnbe et aussi les gagnants de concours DataConnexion
<http://data.sncf.com/>

« Geo-bidouillage de donnée » 30 juin 2010 Loic Haic Slideshare.net

Etalab

<http://www.etalab.gouv.fr/>

Revue des initiatives en France
association LiberTIC (Nantes)

<http://libertic.wordpress.com/2012/01/02/carte-de-france-de-lopen-data-v4/>

Le Guide pratique de l'ouverture des données préparé par la Fing

<http://doc.openfing.org/RDPU/GuidePratiqueDonneesPubliquesv1beta.pdf>

(66 pages)

Les triplets RDF <http://www.w3.org/rdf> qui nous rapproche du monde du web sémantique cf DbPedia)

Le référentiel Opquast qui liste 72 bonnes pratiques de l'Open Data

<https://checklists.opquast.com/opendata>

Nantes : <http://actionsopendata.org/>

Rennes : <http://blog.asso-bug.org/>