

Étude de cas Open Data

par David Libeau
publié sous licence [Creative Commons Attribution \(CC-BY\)](#)

Table des matières

3. Étude de cas : l'Open Data, enjeux et fonctionnement.....	2
3.1 Avant-propos.....	2
3.2 Contexte.....	3
3.3 Importance du sujet.....	4
3.4 Introduction à l'Open Data.....	4
3.4.1 Définition.....	4
3.4.2 Contenu de l'étude de cas.....	5
3.4.3 C'est quoi "des données" ?.....	5
3.4.3.1 Données chaudes, données froides.....	5
3.4.3.2 Les producteurs de données.....	6
3.4.3.3 Données publiques, données privées.....	6
3.4.3.4 Données personnelles, données confidentielles.....	6
3.4.3.5 Le Big Data.....	6
3.4.4 Les enjeux de l'Open Data.....	6
3.4.4.1 Pourquoi ouvrir ses données ?.....	6
3.4.4.2 Pourquoi ne pas ouvrir ses données ?.....	7
3.4.5 Comment ouvrir ses données ?.....	8
3.4.5.1 Les formats.....	8
XML (Extensible Markup Language).....	9
JSON (JavaScript Object Notation).....	10
XSL.....	10
Que faut-il utiliser ?.....	10
3.4.5.2 L'API.....	11
3.4.5.3 La documentation.....	12
3.4.5.2 API ou téléchargement direct ?.....	12
3.4.5.2 Communication et communauté.....	12
3.5 Propositions pour l'Onisep.....	13
3.5.1 Ce qui a été fait.....	14
3.5.2 Ce qui peut être fait.....	15
3.6 Conclusion.....	16
3.7 Pour aller plus loin.....	16
Conclusion.....	17
Références bibliographiques.....	18
Glossaire.....	19

3. ÉTUDE DE CAS : L'OPEN DATA, ENJEUX ET FONCTIONNEMENT

3.1 Avant-propos

L'Open Data, l'ouverture des données publiques, a été rendu obligatoire pour les administrations publiques le 19 décembre dernier par un projet de loi voté par les parlementaires. Axelle LEMAIRE, Secrétaire d'Etat chargée du numérique, porte le projet de loi République Numérique depuis plusieurs mois et cette question de l'ouverture des données fait de plus en plus partie de l'actualité. Beaucoup de villes, d'organisations ou d'autres collectivités sont en avance sur la question, alors que pour d'autres, les mots "Open Data" sont inconnus. Mais que va apporter l'Open Data en France, pourquoi est-il plébiscité par de nombreux acteurs du web et comment une organisation peut-elle libérer gratuitement ses données ?

Dans cette étude de cas, une explication du concept d'ouverture des données sera faite grâce à des exemples concrets et à des explications synthétiques.

L'étude se portera sur des documents officiels, des analyses d'acteurs précurseurs de l'Open Data, mais aussi et surtout sur des problématiques concrètes et des exemples, plus ou moins vulgarisés, pour une compréhension facile de l'ouverture de données.

3.2 Contexte

Je suis étudiant préparant un DUT MMI et j'ai la chance de le faire en alternance pour l'Onisep, établissement public à caractère administratif.

Lors de mes périodes en entreprise, j'ai réalisé différents services web dont une cartographie des dates des journées portes ouvertes. Ce module web a été réalisé avec la base de données de l'Onisep. Lors de la phase de recettage (tests finaux avant mise en ligne), nous avons pu constater des disparités dans la qualité des données, surtout au niveau des coordonnées GPS que nous avons. J'ai alors proposé de vérifier ces données et de relever les incohérences ou erreurs en comparant les adresses postales aux adresses GPS. Cela a pu être fait principalement grâce au geocoder (transformateur d'adresse postale en coordonnées GPS) de la base de données d'adresse nationale (BAN), base de données ouverte, mais aussi, grâce aux données de Google Maps. L'utilisation des données ouvertes mises à disposition du gouvernement français et enrichies par les citoyens a soulevé

quelques questionnements en interne au niveau régional sur nos données et notre module cartographique. En plus, nous avons appris qu'un journal d'informations avait récupéré nos données de portes ouvertes grâce aux documents PDF publiés par l'Onisep pour réaliser une carte similaire pour positionner les journées portes ouvertes en Bretagne.

Le concept d'ouverture des données m'a donc intéressé et j'ai choisi de traiter ce sujet pour informer, en premier lieu, mes collègues sur les principes de l'Open Data, et pour un peu plus creuser le sujet.

Du côté de la formation MMI, l'Open Data n'est pas étudié en tant que tel. L'intérêt de mon choix est donc double. Je pourrais découvrir et explorer un domaine différent et en vogue tout en utilisant les compétences de la formation surtout dans les domaines des réseaux, du développement web, du droit et de la communication. En effet, pour mettre en place un serveur et construire une API ou même pour utiliser un système d'Open Data déjà mis en place, des compétences techniques sont requises. De plus, le droit est fondamental pour étudier toutes les licences pour la publication des données.

3.3 Importance du sujet

Le sujet de l'Open Data me semble pertinent du fait de l'actualité (vote de la loi "République Numérique"), de ma poursuite professionnelle et de mon entreprise d'accueil. L'Onisep, établissement public à caractère administratif, n'a, à l'heure actuelle, aucune politique concernant l'ouverture de ses données. Il me semble important d'étudier la question de façon générale, technique et appliquée. De cette façon, mon étude de cas pourra lancer une réflexion autant pour mon entreprise au niveau local ou national que pour d'autres organisations comme la société de transport de Laval, pour la libération de ses données, ou encore l'amélioration du système existant de Rennes. De plus, cette étude me permettra de gagner en expertise sur ce nouveau domaine pointu et très en vogue.

3.4 Introduction à l'Open Data

3.4.1 Définition

L'ouverture des données publiques ("Open Data") est le principe selon lequel les données publiques ("Data") doivent être disponibles à tous, gratuitement ("Open"). Ces données proviennent des organismes publics, ils les recueillent, les

maintiennent et les utilisent. L'ouverture des données se base sur une démarche démocratique et citoyenne qui permet à tous, sans distinction, d'accéder à ces données, sans payer plus que leurs impôts. Ainsi, ils peuvent les réutiliser pour créer des applications, les consulter et les visualiser, ou même les enrichir en proposant des corrections ou des ajouts. Les entreprises vont pouvoir créer plus facilement des applications ou des services innovants grâce à ces données.

3.4.2 Contenu de l'étude de cas

La première partie de l'étude sera une présentation générale de l'Open Data. Il s'agira de définir les enjeux que l'ouverture des données implique (économiques, sociaux, politiques...) tout en ayant un regard critique sur ses conséquences et les risques. Une différenciation de l'Open Data et du Big Data sera faite et nous nous attarderons à définir tous les aspects qu'impliquent une ouverture des données, autant sur la documentation, que sur les licences, ou encore sur la "Dataviz", la visualisation ou mise en forme des données.

La deuxième partie se concentrera sur l'Onisep, l'organisation où je fais mon alternance, et de la possible ouverture de la base Ideo. Nous nous servirons des constatations précédentes pour proposer un jeu de données.

3.4.3 C'est quoi "des données" ?

L'emplacement précis d'un arrêt de bus et le prochain passage de la ligne B à cet arrêt, le prochain train pour Rennes au départ de Laval et le potentiel retard de ce train, les emplacements des poubelles ou conteneurs pour le verre aux alentours de la gare, les résultats de la dernière élection dans mon bureau de vote, les prix des carburants sur le territoire français, la pollution de l'air, ou encore, les dates des journées portes ouvertes des IUT de Bretagne... tout cela, c'est des données. Ce sont des informations connues qui ont été informatisées pour pouvoir les stocker et les transmettre.

3.4.3.1 Données chaudes, données froides

On distingue deux types de données, les données chaudes et les données froides. Les données froides sont des informations qui ne changent pas ou peu fréquemment. Par exemple, un arrêt de bus, physiquement, c'est statique, ça ne bouge pas, donc cette donnée est une donnée froide. En revanche, le temps avant le prochain passage d'un bus est une donnée mise à jour très fréquemment, toutes les

secondes. De même, pour les données météorologiques comme la température, ses mises à jour sont fréquentes, ce sont donc des données chaudes.

3.4.3.2 Les producteurs de données

Beaucoup d'administrations publiques produisent des données, mais des entreprises privées sont aussi des producteurs. Par exemple, Keolis est l'entreprise en charge du réseau de bus à Laval et à Rennes. Ceux-ci possèdent toutes les données sur le réseau comme les horaires, les emplacements des arrêts, etc. A Rennes, ces données sont en Open Data.

3.4.3.3 Données publiques, données privées

Les données ouvertes, en Open Data, ne sont pas obligatoirement des données publiques. On l'a vu, si ces données sont produites par des administrations publiques, ou même gratuitement par des citoyens, ces données sont publiques. En revanche, il existe un très grand nombre de données privées. Nous pouvons le voir avec Keolis qui a ouvert ses données à Rennes mais pas à Laval. L'entreprise Keolis est privée et c'est sa politique qui détermine s'ils ouvrent leurs données ou non.

Donc, les données sont privées ou publiques selon leur producteur.

3.4.3.4 Données personnelles, données confidentielles

Attention, l'ouverture de données ne doit pas se faire sur des données personnelles ou confidentielles ! Un hôpital se doit, pour des raisons évidentes de confidentialité, de ne pas ouvrir les dossiers de ses patients.

3.4.3.5 Le Big Data

Vous avez sûrement déjà entendu parler du Big Data. Aussi nommé mégadonnées, ça représente une base de données tellement grande qu'il est difficile de la traiter pour des systèmes informatiques. L'Open Data enrichit l'écosystème global des données en mettant accessible des données supplémentaires. Cependant, le Big Data est bien un autre sujet.

3.4.4 Les enjeux de l'Open Data

3.4.4.1 Pourquoi ouvrir ses données ?

La transparence, c'est la raison démocratique prônée lors de l'ouverture des données publiques. Chaque citoyen paie ses impôts et est en droit de pouvoir accéder aux documents administratifs des administrations publiques (qu'il finance). La CADA (Commission d'Accès aux Documents Administratifs) est chargée de veiller au respect de la liberté d'accès aux documents administratifs et aux archives publiques. N'importe quel citoyen peut faire un recours à la CADA s'il n'arrive pas à obtenir un document.

Ouvrir ses données, c'est aussi et surtout reprendre le pouvoir sur ses données. Lorsqu'on commence à se poser des questions sur l'ouverture des données dans une administration, on se rencontre souvent du manque de fiabilité des données (données incomplètes...), du manque de documentation (explication de la donnée), etc. Alors, le discours suivant s'installe : "Je ne suis pas sûr que mes données soient suffisamment fiables pour pouvoir être publiées.". Ce discours est légitime et est souvent source d'arrêt pour la prise d'initiative. Sauf que le problème n'est pas la publication de données non fiables mais l'utilisation dans l'administration de données non fiables. Comment puis-je accepter que les données présentes dans mon système d'informations ne soient pas fiables ? L'Open Data peut donc aider l'administration à remettre en cause sa base de données pour l'améliorer.

Mais l'amélioration peut aussi venir des utilisateurs via un crowdfixing ! En effet, ouvrir ses données n'est pas seulement les mettre en ligne et créer une page web pour les télécharger. L'ouverture des données doit s'accompagner par la construction d'une communauté de développeurs, de citoyens ou d'acteurs. Ceux-ci vont réutiliser les données en créant de nouveaux services, valorisons leurs applications, leurs sites web. De plus, en utilisant les données, ils vont sûrement tomber sur des bogues et découvrir des incohérences ou problèmes sur les données. L'intérêt pour l'administration est qu'en créant un esprit communautaire, les utilisateurs vont pouvoir nous signaler ces problèmes et nous allons les résoudre ! Sans l'Open Data, il aurait fallu passer un temps considérable pour vérifier toutes les données alors que les utilisateurs de l'Open Data, en créant de nouveaux services, s'en rendent compte par eux-mêmes. La création d'une communauté est primordiale, cela peut se faire tout simplement en communiquant sur les réseaux sociaux mais

des initiatives comme un concours de création d'application est très apprécié par les développeurs et permet de faire connaître ses données.

3.4.4.2 Pourquoi ne pas ouvrir ses données ?

Pour les administrations publiques, la loi République Numérique les y oblige. Mais on peut émettre des réserves pour les entreprises privées ou même pour certaines administrations.

En effet, pour reprendre Keolis, à Laval ils n'ont pas ouvert leurs données. La concurrence fait souvent peur dans ces cas-là. L'entreprise a déjà son application officielle, d'autres entreprises concurrentes pourraient venir sur le marché grâce à l'Open Data. Cet argumentaire n'est pas forcément très réfléchi car le produit que vend Keolis est bien le transport d'individus grâce à son réseau de bus. En ouvrant leurs données, des applications, souvent bien mieux que l'application officielle, seront créées. Cependant, ces applications ne vont prendre aucun marché puisqu'elles vont aider les utilisateurs à prendre le bus. Le bénéfice est donc autant pour la collectivité que pour Keolis qui a donc, selon moi, tout intérêt à ouvrir leurs données.

Dans le cas d'une administration, dans la majeure partie des cas, la concurrence est inexistante. Cependant, pour l'Onisep, par exemple, l'office a de nombreux concurrents : L'étudiant, Studyrama, etc.

3.4.5 Comment ouvrir ses données ?

Pour ouvrir ses données, certains principes ou règles ont été posés pour que l'Open Data soit de qualité.

3.4.5.1 Les formats

Le format des données est un des éléments primordiaux dans l'Open Data. Pour le comprendre, il faut penser aux développeurs-citoyens qui vont s'approprier la donnée.

Si on prend des pages de l'ancien relevé de naissance d'une ville, numérisée en format image, par exemple, ce n'est que très peu exploitable pour un informaticien qui cherche à faire des statistiques automatiques, par exemple.

Nous allons revenir sur les différents formats utilisés à l'heure actuelle.

CSV (Comma-Separated Values)

Le format CSV est le format le plus basique qu'il existe. Comme son nom l'indique, les données sont séparées entre elles par des virgules.

Le principal avantage est le faible poids pour le transport des données.

Exemple :

```
id_arret;nom_arret;coord_X;coord_Y
1;Technopolis;1.2345;45.6789
2;Courbertin;1.2745;45.6389
```

XML (Extensible Markup Language)

Le langage XML est un format très pratique car il permet une compréhension facile, à la fois des humains et des ordinateurs. Il est composé de balises comme le langage HTML.

L'avantage premier est bien la lisibilité naturelle des données par les humains sans alourdir les algorithmes des ordinateurs. La possibilité de hiérarchiser l'information sur plusieurs niveaux est souvent aussi recherchée et ce format le permet.

Exemple :

```
<?xml version="1.0" encoding="UTF-8"?>
<OpenDataArretsDeBus>
  <arret>
    <id>1</id>
    <nom>Technopolis</nom>
    <emplacement>
      <coord_X>1.2345</coord_X>
      <coord_Y>45.6789</coord_Y>
    </emplacement>
  </arret>
  <arret>
    <id>2</id>
    <nom>Courbertin</nom>
    <emplacement>
      <coord_X>1.2745</coord_X>
      <coord_Y>45.6389</coord_Y>
```

```
</emplacement>
</arret>
</OpenDataArretsDeBus>
```

JSON (*JavaScript Object Notation*)

Comme son nom l'indique, il est souvent utilisé dans le langage JavaScript (langage de programmation web).

Exemple :

```
{
  "arret" : {
    "id": "1",
    "nom": "Technopolis",
    "emplacement": [
      {
        "coord_X": "1.2345",
        "coord_Y": "45.6789"
      }
    ]
  }
}
```

XSL

Le simple tableau Excel.

Il a l'avantage d'être simple d'utilisation mais nécessite l'achat d'une licence Office (ce qui est non démocratique).

Exemple :

	A	B	C	D
1	id	nom	coord_X	coord_Y
2	1	Technopolis	1,2345	45,6789
3	2	Coubertin	1,2745	45,6389
4				

Que faut-il utiliser ?

L'Open Data est la libération des données pour tous. Pas uniquement ceux qui ont la suite payante Office ou Adobe sur leur ordinateur. Utiliser des formats ouverts est synonyme de neutralité et d'accessibilité. C'est gage de qualité pour l'Open Data.

Les cinq étoiles de Tim Berners-Lee

Le site 5stardata.info classe selon cinq catégories l'accessibilité des données.

- * Données accessibles sur le web (sans condition de formats)
- ** Données accessibles structurées (exemple : Excel au lieu de l'image d'un tableau)
- *** Format non propriétaires (exemple : CSV au lieu d'Excel)
- **** Usage d'URL pour identifier les données
- ***** Données liées sémantiquement (lier une donnée à des données d'autres producteurs pour mieux les contextualiser)

3.4.5.2 L'API

L'API (Application Programming Interface) est l'interface des développeurs. Il s'agit d'un site web où les développeurs trouveront plein d'informations et de documentation pour utiliser le système d'information auquel il est lié.

Pour l'Open Data, c'est une plateforme web sur laquelle on accédera aux données par un système d'URL.

Par exemple, l'adresse <http://data.bus-de-laval.fr/arret?id=1&format=xml> renverrait en format XML l'élément "arret" d'id 1. Il me donnerait l'emplacement de l'arrêt de bus. L'intérêt d'utiliser une API est la facilité de mise à jour des données et aussi la publication de données chaudes. On est obligé d'utiliser une API pour les données chaudes car leur mise à jour peut être fréquente. La création d'une API et sa mise en place est assez technique pour l'entreprise ou l'administration. C'est lourd, au niveau des infrastructures informatiques et au niveau des compétences techniques à mobiliser, et donc difficile à entreprendre. Il faut de bonnes connaissances informatiques pour réaliser un projet de ce type et le résultat peut être parfois d'une qualité moyenne. Une API de qualité doit être très bien documentée (c.f. paragraphe suivant) et facile d'utilisation, intuitive, pour les développeurs.

Des erreurs peuvent, en effet, rendre l'API très difficile à utiliser pour les algorithmes des développeurs. Par exemple, l'Open Data de Keolis Rennes (précurseur, lancé en 2009) était très bien, surtout pour l'époque, mais la mise en place maladroite des versions rend son utilisation compliquée. L'API dispose de plusieurs versions, dans la théorie, c'est très important mais ici, dans la pratique, les

fonctions de chaque version ne sont utilisables que dans ses versions. Concrètement pour le développeur, celui-ci doit obligatoirement connaître le numéro de version associé à la fonction utilisée, ce n'est que peu pratique lorsque l'on construit une application où l'on automatise beaucoup le fonctionnement.

Aussi, pour des raisons statistiques ou de contrôles, certaines organisations mettent en place un système d'identification. Pour faire leurs requêtes, chaque développeur doit renseigner son identifiant. Très pratique, pour l'organisation, c'est, par principe, contraire à l'Open Data, qui doit être accessible à tous, sans inscription d'utilisateur.

Cependant, le plus important dans une API est sa documentation.

3.4.5.3 La documentation

Extrêmement importante, la documentation va donner au développeur les moyens d'utiliser l'API. Par exemple, on va lui indiquer que l'adresse <http://data.bus-de-laval.fr/arret> permet de récupérer un arrêt ; que le développeur doit obligatoirement indiquer le format de réponse, et au moins un élément "id", "nom", ou "coord_X" et "coord_Y" ; que s'il indique les coordonnées X et Y, l'arrêt le plus proche, dans un rayon d'un kilomètre maximum, sera renvoyé ; ou même le format de réponse pour les erreurs. Bref, tout ce qu'il a besoin de savoir pour travailler !

3.4.5.2 API ou téléchargement direct ?

La mise en place d'une API peut être souvent lourde, au niveau technique et demander beaucoup de temps pour la rédaction de la documentation. Bien souvent le téléchargement direct de fichiers CSV ou XML est d'abord envisagé mais si nous avons des données chaudes à publier, l'API est inévitable. Cependant, même si nous avons une API, rien ne nous empêche, et c'est même conseillé, de proposer les données en téléchargement en plus de l'API.

La réponse est donc, les deux, si on peut !

3.4.5.2 Communication et communauté

Lors de l'ouverture d'une base de données, il ne faut évidemment pas négliger la communication et l'aspect communautaire déjà évoqué précédemment.

La communication paraît naturelle pour une entreprise qui propose ce service innovant ou même pour une collectivité qui doit satisfaire ses électeurs mais pour certaines administrations, cela peut manquer. Certaines ouvertures de données ont

tellement peu de communication que leur existence même n'est pas connue des citoyens (c.f. partie suivante sur l'Onisep), un comble !

3.4.5.3 Licences

Juridiquement, les licences sont là pour protéger les propriétaires et les utilisateurs. Elles apportent la sécurité juridique nécessaire aux producteurs et aux réutilisateurs des données en fixant des règles d'utilisation. Il existe des multitudes de licences pour une multitude d'utilisation. Cependant, inventer une licence pour sa propre utilisation est fortement proscrite.

Pour l'Open Data, on notera différentes licences qui sont souvent utilisées (liste non exhaustive) :

Licence ouverte / Open licence, a été créé par Etalab pour l'Open Data : <https://www.etalab.gouv.fr/licence-ouverte-open-licence>. Libre et gratuite, elle autorise la reproduction, la redistribution, l'adaptation et l'exploitation commerciale des données mais oblige la mention de la paternité.

Creative Commons, est une licence américaine donnant une solution alternative légale aux personnes souhaitant libérer leurs œuvres des droits de propriété intellectuelle standard de leur pays.

ODT, Open Document, est un format ouvert et normalisé pour la bureautique et le successeur du format initial d'OpenOffice.org

3.4.5.3 Dataviz

La visualisation de données, "Dataviz", est l'outil essentiel pour une démocratisation des données publiées. Grâce à la visualisation, l'Open Data n'est pas seulement intéressant pour les développeurs. Il permet l'accès aux données pour tous grâce à une mise en forme accessible (graphiques, tableaux, cartes...).

3.5 Propositions pour l'Onisep

Suite à la loi "République Numérique" toutes les administrations publiques vont être obligées d'ouvrir leurs données. L'Onisep a déjà ouvert une partie de ses données. En effet, le Comité interministériel pour la modernisation de l'action

publique (CIMAP) de décembre 2013
 (https://www.etalab.gouv.fr/lesdecisionssurlopendataducimapdu18decembre2013), a
 acté et a sommé l'Onisep d'ouvrir ses données.

3.5.1 Ce qui a été fait

L'Onisep a choisit la plateforme Data Gouv (data.gouv.fr) pour publier quelques données.

Office national d'information sur les enseignements et les professions

L'Onisep (Office national d'information sur les enseignements et les professions) est un établissement public qui dépend du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. Éditeur public, l'Onisep produit et diffuse toute l'information sur les formations et les métiers. Il propose aussi des nouveaux services aux élèves, aux parents et aux équipes éducatives. Notre mission :

- Informer sur les formations, les métiers, les secteurs professionnels ; donner des chiffres de l'insertion professionnelle.
- Accompagner les jeunes et leurs familles pour les aider à préparer leur avenir.
- Conseiller les parents, les élèves et les équipes éducatives qui aident les jeunes à choisir leur orientation.

JEUX DE DONNÉES MEMBRES ABONNÉS

<p>Organismes d'information et d'orientation du Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la ...</p> <p>1 0</p>	<p>Liste des métiers ONISEP</p> <p>1 0</p>	<p>Formations initiales en France</p> <p>0 0</p>
<p>Etablissements d'enseignement secondaire</p> <p>2 1</p>	<p>Etablissements d'enseignement supérieur</p> <p>5 4</p>	

TÉLÉCHARGER EN CSV

Figure 6. La page de l'Onisep sur DataGouv

Les cinq jeux de données, tous au format CSV et XML sous la licence d'Etalab "Licence Ouverte / Open Licence" :

Liste des organismes d'information et d'orientation du Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche (CIO, SCUIO)

Il contient le nom de l'organisme, le sigle et les coordonnées géographiques (adresse postale, commune, département, académie, région, longitude X, latitude Y) et le lien d'accès direct sur le site Onisep.

Liste des métiers répertoriés par l'Onisep. Il contient le libellé du métier et le lien d'accès direct sur le site Onisep.

Liste des formations initiales de France. Il contient le code CNIS/NSF de la nomenclature des spécialités des formations, le type de formation, le libellé principal

et le libellé complémentaire, la durée, le ministère responsable, le niveau de formation et le lien d'accès sur le site Onisep.

Liste des établissements d'enseignement supérieur. Il contient le code UAI (unité administrative immatriculée), le type d'établissement, le nom, le sigle, le statut (public, privé...), la tutelle (ministères, chambres consulaires...), université de rattachement, coordonnées géographiques (adresse postale, commune, département, académie, région, longitude X, latitude Y) et le lien d'accès direct sur le site Onisep.

Liste des établissements d'enseignement secondaire. Il contient la même chose que la liste des établissements d'enseignement supérieur.

3.5.2 Ce qui peut être fait

L'ouverture des données de l'Onisep est encourageante. On retrouve beaucoup d'informations pour les établissements et pour les organismes d'information (on relèvera les coordonnées GPS par exemple). En revanche, pour la liste des métiers, on sera plus dubitatif. Ce jeu de données est un peu pauvre, nous n'avons que le libellé (le nom) et le lien vers l'Onisep alors que sur le site, plein d'informations sont délivrées (niveau d'étude, salaire, centre d'intérêt...). C'est un premier pas, mais les applications ne sont que faible (c.f. ma réutilisation "Métiers au hasard").

Ce qui manque dans l'Open Data est des données sémantiques, replacées dans leur contexte, liées entre elles. Par exemple, nous pourrions lier les établissements avec les formations, en indiquant dans les établissements quelles formations celui-ci dispensent. De même pour les métiers, on pourrait rajouter un jeu de données "Secteurs d'activité" et lié chaque métier avec son ou ses secteur-s d'activité-s.

En plus de cela, un point important sur lequel l'Onisep n'a pas ouvert ses données est les dates des journées portes ouvertes (c.f. la carte des JPO dans mes missions). Ces données mises à jour annuellement sont fortement demandées par les utilisateurs de l'Onisep.

Pour finir, on terminera sur la forme de cet Open Data. Aucune communication n'a été faite et par conséquent, il n'y a que peu de réutilisations publiées sur Data Gov.

3.6 Conclusion

L'Open Data est novateur et rejoint les désirs de démocratisation de l'information. Les données pour tous est un enjeu important mais les avancées dans le domaine ne se font que lentement. Certaines lois aident l'Open Data, avant de se faire rattraper par les lobbys ou les pressions d'entreprises qui voient dans l'OpenData une concurrence démocratique gênante. Je préconiserais plus d'éducation et de communication sur l'OpenData, pour bien comprendre les enjeux et les bénéfices.

3.7 Pour aller plus loin

Découvrez des réutilisations plutôt humoristiques de l'Open Data sur mon profil Data Gouv : <https://www.data.gouv.fr/fr/users/david-libeau-2/reuses/>

RÉFÉRENCES BIBLIOGRAPHIQUES / SOURCES

KOBER Vincent, *Open data Ouverture, exploitation, valorisation des données publiques*, territorial éditions, 2014

CHIGNARD Simon, *Open data Comprendre l'ouverture des données publiques*, FYP éditions, 2012

ADBS, *Big data - Open data Quelles valeurs ? Quels enjeux ?*, De Boek, 2015

Site data gouv : <https://www.data.gouv.fr/fr/> - 13 novembre 2015

Site République numérique <https://www.republique-numerique.fr/> - 17 décembre 2015

Site Portail de l'Economie et des Finances <http://www.economie.gouv.fr/> – 1 avril 2016

Site Wikipédia <https://fr.wikipedia.org/> - 20 mars 2016

Site Open Data Keolis Rennes <https://data.keolis-rennes.com/> - 24 janvier 2016

Site Open Data de Rennes métropole <http://www.data.rennes-metropole.fr/> - 7 février 2016

Site d'Etalab <https://www.etalab.gouv.fr/> - 16 janvier 2016

Site de Creative Commons <http://creativecommons.fr/> - 21 septembre 2015

Site de la CADA <http://www.cada.fr/> - 27 février 2016

Site de Tim Berners-Lee <http://5stardata.info/en/> - 5 mars 2016