



Photo by Colleen Simon

# Open Data et ESR

## Opportunité de créer de nouveaux services

La collection numérique de l'Amue

SEPTEMBRE 2019

## Edito

Par Stéphane Athanase, directeur de l'Amue

La communauté Enseignement Supérieur et Recherche connaît le sujet de l'Open Science qui est un sujet important de politique publique pour notre recherche. Il pourrait occulter parfois le sujet plus global de l'Open Data qui mérite d'être appréhendé dans sa globalité, c'est-à-dire dans tous les contextes, de notre système de l'Enseignement supérieur et de la recherche français.

Cette logique d'ouverture et de partage provient d'un mouvement plus ancien, plus fondamental, qui rejoint la doctrine de gouvernance dite "open government" portant, entre autres, les valeurs de transparence de la vie publique, de co-production du service public, de transparence et de valorisation de l'action publique.

### Vecteur de transformation

L'Open Data, comme de nombreux vecteurs de transformation, repose sur un ensemble de mouvements : sociétaux, juridiques, numériques, culturels, organisationnels et citoyens.

Porté par l'Etat, repris au niveau des ministères et de nombreux opérateurs publics, ce mouvement aurait besoin, au sein de nos établissements d'enseignement supérieur, d'être amplifié pour répondre à cet enjeu de politique publique.

### Stratégie d'établissement

Il ne s'agit pas d'une charge supplémentaire mais bien d'une opportunité de promouvoir l'action menée par nos établissements, de proposer ou permettre de nouveaux services à valeurs ajoutées pour les usagers du service public autour de nos données. Interrogeons-nous sur notre patrimoine informationnel, valorisons notre action en ouvrant nos données, documents et algorithmes, imaginons de nouveaux services pour nos usagers, en un mot bâtissons une stratégie Open Data pour chacun de nos établissements. Stratégie à intégrer dans un ensemble plus large autour de la valorisation de l'information détenue par nos établissements.

### Sensibilisation

Ce numéro de la collection numérique a vocation à lancer une sensibilisation sur l'Open Data, à aider les gouvernances d'établissements à appréhender le sujet, découvrir de nouvelles expériences et se préparer à contribuer à ce mouvement sociétal. Une journée, organisée par l'Amue le 14 novembre, complétera ce dispositif qui alimentera votre réflexion sur l'Open Data, vous aidera à construire votre feuille de route et vous présentera les travaux de l'Amue en la matière. J'espère vous y retrouver nombreuses et nombreux.

#### DANS CE NUMERO



Photo by Amelie & Niklas Ohlroge on Unsplash

#### Etalab : le moteur de l'OpenData

Page #3

#### DANS CE NUMERO



Photo by Hans-Peter Gausler on Unsplash

#### Ouverture des données sur les parcours d'étude : expérimentation

Page #19



# La convergence de l'open data gouvernemental et de la science ouverte

Par Marin Dacos, conseiller pour la science ouverte du Directeur Général de la

Recherche et de l'Innovation, Ministère de l'enseignement supérieur, de la recherche et de l'innovation (MESRI)

## Plan National pour la science ouverte

Le plan national pour la science ouverte a été rendu public par Frédérique Vidal, Ministre de l'enseignement supérieur, de la recherche et de l'innovation, le 4 juillet 2018. Il met en place neuf mesures destinées à ouvrir les publications scientifiques et les données de la recherche. L'intérêt est d'abord politique : la science, essentiellement financée par la sphère publique, doit revenir au public, à tous les publics, et non être enfermée dans quelques campus ayant les moyens de s'offrir des abonnements aux prix élevés. L'intérêt est ensuite scientifique, car en faisant tomber les barrières à l'accès, l'accès universel aux publications améliore la qualité de la science. Elle présente enfin un avantage de société : une science plus ouverte est aussi plus accessible à la société.

Le plan national entame une convergence entre les différents mouvements d'ouverture qui touchent le secteur public depuis la déclaration de Budapest en 2002, qui invente le concept d'accès ouvert aux publications scientifiques, et la rencontre de Sebastopol en 2007, qui forge les principes de l'ouverture des données publiques. La France a pris en 2016 et 2018 deux initiatives qui ont pour effet de faire converger les deux mouvements.

## Ouvrir autant que possible, fermer autant que nécessaire

La loi pour une République numérique accorde un nouveau droit aux chercheurs et enseignants-chercheurs en tant qu'auteurs. Quand une recherche est financée à au moins 50% par l'Etat, l'auteur dispose de la faculté de déposer dans une archive ouverte le manuscrit auteur accepté, 6 mois (STM - sciences, techniques, médecine) ou 12 mois (SHS, sciences humaines et sociales) après publication. Si le contrat dit le contraire, il est réputé non écrit. Dans le domaine des données, la loi instaure une nouvelle obligation, qui a été trop peu identifiée par les établissements de l'ESRI (enseignement supérieur, recherche et innovation). L'article 6 de la loi fixe le principe d'ouverture par défaut pour toutes les données administratives, parmi lesquelles on compte les données de la recherche. Par défaut, le travail des agents publics est propriété du financeur,

donc en général de l'employeur. Mais les chercheurs et enseignants-chercheurs bénéficient d'une exception portant sur leurs articles et ouvrages publiés, qui n'appartiennent pas au financeur, mais à l'auteur. Cette exception n'est valable que pour les publications et pas pour la matière première de la recherche que sont les données. Celles-ci appartiennent au financeur, ce qui offre aux établissements un important levier pour mettre au point une véritable politique des données. Il est à noter que l'article 6 n'a pas été conçu principalement pour les données de la recherche, mais pour les administrations publiques en général. Il reste qu'elle s'applique aux établissements de l'ESRI au double titre des données de la recherche et des données mobilisées par les établissements pour assurer leur fonctionnement. Bien sûr, une demi-douzaine d'exceptions sont prévues par la loi, afin de ne pas ouvrir de façon inconsidérée.

### EXCEPTIONS A L'OBLIGATION D'OUVERTURE

Données personnelles sans consentement ; Données présentant des risques pour la sécurité publique ou la sécurité de l'établissement ; Secret de la défense nationale ; Secret professionnel ; Secret des affaires ; Secrets industriels et commerciaux ; Données présentant des risques pour la protection du potentiel scientifique et technique de la nation.

La doctrine du Plan national pour la science ouverte comme de la politique de science ouverte de l'Union européenne est même : "ouvrir autant que possible, fermer autant que nécessaire".

## Convergence des mouvements

Par ailleurs, l'open government partnership (OGP), structure internationale regroupant plus de 70 pays, dont la France, publie tous les deux ans des engagements nationaux au titre de la transparence de l'action publique. Issu du mouvement de l'open data, l'OGP est né en 2011. Dans [ses engagements de 2018](#), la France a décidé d'intégrer à ses "traditionnels" engagement d'ouverture des données un chapitre dédié à la science ouverte ("Engagement n°18 : Construire un écosystème de la science ouverte").

Cette convergence des mouvements d'ouverture de la science et d'ouverture des données publiques est une clé de leur réussite. Nous devons, en effet, intégrer ce paradigme au cœur de nos dispositifs techniques, pour que l'ouverture soit une étape logique et intégrée dans le processus de travail de la donnée, pas un fardeau supplémentaire "hors circuit". Pour avancer dans cette direction, il faut mobiliser tous les acteurs de l'ESRI. Le mouvement touche l'ensemble du processus scientifique, en premier lieu la chercheuse et le chercheur, mais aussi tous les métiers d'accompagnement de la recherche, et tous les niveaux de l'organisation du monde de la recherche. Il semble acquis que la volonté collective existe, mais que la complexité technique, juridique et disciplinaire du sujet constitue une difficulté importante. C'est la raison pour laquelle a été mise en place une coordination nationale à travers le Comité pour la science ouverte (CoSO). À sa base, plus de 250 experts ayant manifesté leur intérêt, échangent sous forme numérique dans le cadre du Forum de la science ouverte. À son sommet, les présidents et présidentes des institutions qui représentent l'ensemble du secteur de l'enseignement supérieur, de la recherche et de l'innovation. Avec cette combinaison d'expertise de terrain, que nous espérons représentative de la diversité et de la complexité des situations, et de capacités décisionnelles au plus haut niveau, nous disposons d'un instrument puissant de coordination nationale et de contribution internationale.

Vous trouverez une bibliothèque de ressources incontournables et d'actualités de la science ouverte sur <https://www.ouvrirelascience.fr>. Vous y trouverez notamment un [guide juridique](#) sur l'ouverture des données de la recherche.

## Comité pour la science ouverte

Si vous souhaitez participer aux travaux du Comité pour la science ouverte, notamment en participant au forum, notez que l'appel à participation est toujours ouvert. Nous avons besoin de tous les métiers. Envoyez votre CV et lettre de motivation à [coso@recherche.gouv.fr](mailto:coso@recherche.gouv.fr)

### CONTACTS

Mail : [marin.dacos@recherche.gouv.fr](mailto:marin.dacos@recherche.gouv.fr)  
sur Twitter : [@marindacos](https://twitter.com/marindacos)



# Etalab : une mission au service de l'ouverture des données publiques

par Romain Tales, DINSIC

La mission Etalab, rattachée à la Direction Interministérielle du Numérique, du système d'Information et de Communication de l'Etat (DINSIC), a été créée en février 2011 afin de d'impulser la politique française en matière d'open data, c'est-à-dire l'ouverture des données publiques. Moins d'un an plus tard, le 5 décembre 2011, la plateforme nationale de l'open data est ouverte : il s'agit de data.gouv.fr. Cette logique d'ouverture avant-gardiste est finalement confirmée par la loi pour une République numérique promulguée en octobre 2016. Désormais, les données publiques doivent être diffusées par les administrations "par principe", et non plus seulement sur demande.

## L'open data par principe : une perte de temps ?

Les administrations au sens large (ministères, collectivités territoriales, établissements publics sous tutelle ou encore entreprises publiques) opposent souvent à l'obligation de l'ouverture des données publiques l'argument du manque de temps. Ouvrir les données viendrait concurrencer leur cœur de métier et leur ferait perdre un temps précieux. Or, la mise à disposition par principe des bases de données constitue en fait un gain de temps précieux : en diffusant pro activement un jeu de données sur la plateforme nationale, l'administration n'a plus besoin de répondre aux demandes individuelles ni de réaliser des extractions spécifiques qui la détourneraient de son activité principale.

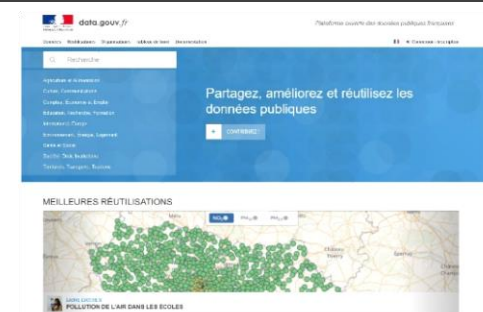
## Les arguments à l'ouverture : des bénéfices pour les administrations et les réutilisateurs des données

L'open data permet en premier lieu de donner de la visibilité aux travaux de l'administration en valorisant son travail. Outre le fait qu'il permet de faciliter le quotidien des administrations en leur faisant gagner du temps dans la mise en œuvre de leur mission de service public, il s'agit aussi d'un moyen de renforcer l'efficacité de l'action publique en général : la circulation des données crée des synergies entre les structures et améliore la mise en œuvre des politiques publiques à travers l'exploitation des données. Cette diffusion est une voie ouverte aux

réutilisateurs de données publiques pour les enrichir ou détecter des anomalies, voire les corriger. Cet ainsi que des données telle que la base des accidents corporels de la circulation a pu faire l'objet d'une montée en qualité à travers notamment les retours d'expérience des réutilisateurs.

Pour ces réutilisateurs, qui peuvent être aussi bien des administrations que des acteurs privés, la réutilisation des données diffusées en open data est le moyen de créer des usages innovants par le croisement de données, permettant de répondre à des besoins que les administrations ne présageaient pas forcément lors de la production initiale de la donnée.

## LA PLATE-FORME DE REFERENCE : DATA.GOUV.FR



Source : data.gouv.fr

## Quelques exemples des bénéfices de l'open data parmi d'autres...

### La création d'un Service Public de la Donnée

Introduit par la loi pour une République numérique, le service public de la donnée vise à mettre à disposition, en vue de faciliter leur réutilisation, les données de référence, c'est à dire celles qui présentent le plus fort impact économique et social.

Les données de référence satisfont aux trois conditions définies par la loi : elles constituent une référence commune pour nommer ou

identifier des produits, des services, des territoires ou des personnes ; elles sont réutilisées fréquemment par des personnes publiques ou privées autres que l'administration qui les détient ; leur réutilisation nécessite qu'elles soient mises à disposition avec un niveau élevé de qualité.

La liste des données de référence, fixée par décret en Conseil d'Etat, inclut à ce jour neuf jeux de données des domaines suivants :

- économique et associatif : base Sirene d'identification des établissements, répertoire national des associations, répertoire des emplois et des métiers,

- géographique : plan cadastral informatisé, base adresse nationale, code officiel géographique, ...

- administratif : référentiel de l'organisation administrative de l'Etat.

L'objectif principal du service public de la donnée est de faciliter la réutilisation des données présentant le plus fort impact pour l'économie et la société.

La dynamique du service public de la donnée se traduit aujourd'hui par une forte adhésion des utilisateurs au dispositif. La progression du nombre de téléchargements l'illustre. Selon l'INSEE, le nombre d'utilisateurs de la mise à jour quotidienne de la base Sirene a été multiplié par 20 depuis son intégration dans le service public de la donnée.

Le cadastre était auparavant vendu (un seul acheteur en 2016) et diffusé gratuitement sur demande aux collectivités et administrations d'Etat. Sa mise en ligne accompagnée et outillée par Etalab a permis non seulement une meilleure diffusion des données (7 millions de téléchargement, tous types confondus depuis le début de l'année) mais a aussi permis de décharger les services des finances publics qui renvoient maintenant directement les demandes vers cadastre.data.gouv.fr.



Les utilisateurs professionnels des données se sont bien emparés du service public de la donnée (TomTom ou EDF par exemple). On voit aujourd'hui émerger des startups qui proposent des services de visualisation et d'utilisation des données de référence.

Les collectivités sont aussi des bénéficiaires directs des données de référence pour leurs usages propres (urbanisme, logement, vie associative). La plupart d'entre elles rediffusent d'ailleurs certaines bases sur leurs propres portails, dans une version localisée à la commune.

### Améliorer la prévision des accidents en mer...

Les Centres régionaux opérationnels de surveillance et de sauvetage (CROSS) ont publié les données relatives à leurs opérations en mer sur [data.gouv.fr](https://data.gouv.fr). Dans le cadre des défis "Entrepreneurs d'intérêt général" proposés par Etalab, et en partenariat avec les administrations publiques, des datascientists les ont mises à profit dans [Prédisauvetage](https://www.predisauvetage.fr). Cette carte interactive des opérations d'assistance et de sauvetage en mer propose de visualiser les interventions en leur appliquant plusieurs filtres, dont la date, la nature de l'opération, le lieu, la gravité etc.

Les CROSS ainsi que de nombreux acteurs du sauvetage en mer produisaient des données sur leurs interventions pour les rapports externes, mais n'en tiraient aucune information utile en interne pour l'exercice de leur fonction. Le traitement de ces données permet d'étendre les connaissances sur l'accidentologie en mer et d'améliorer la prévention. Les sauveteurs ont aussi une meilleure connaissance des risques sur certaines zones. Le projet a eu des impacts sur des secteurs inattendus : les assurances proposent par exemple des contrats personnalisés pour les activités nautiques. La Direction départementale des Territoires de la Mer a également amélioré la signalisation maritime par endroits, et des informations sont échangées avec les CHU. Les EIG ont entrepris un travail de normalisation des données afin de rendre le système Prédisauvetage compatible au niveau national.



Photo by Jonathan Gray

### Rétablir l'équilibre informationnel au sein de l'économie...

A l'occasion de la publication des données sur les transactions immobilières (Demandes de Valeurs foncières) par la DGFIP, Etalab a développé une application, [app.dvf.etalab.gouv.fr](https://app.dvf.etalab.gouv.fr), qui permet de visualiser les données sous la forme d'une carte, un moyen de démontrer l'importance d'exploiter les données ouvertes. "Il s'agissait de mettre à disposition du grand public, pour lequel manipuler des giga-octets de données n'est pas envisageable, une interface de visualisation simple et facile à prendre en main", déclare Marion Paclot, datascientist dans la mission Etalab. L'application, publiée sous sa version initiale peu après l'ouverture des données de DVF, a depuis connu plusieurs améliorations, notamment liées aux contributions des internautes. Le site reçoit environ 23000 visites par jour, et a été consulté à 2,5 millions de reprises depuis son ouverture.

#### POUR PLUS D'INFORMATIONS

Le site d'Etalab :  
<https://www.etalab.gouv.fr/>

La plate-forme nationale de l'Open Data :  
<https://www.data.gouv.fr>

Les données des opérations coordonnées par le CROSS Secmar :  
<https://www.data.gouv.fr/fr/datasets/operations-coordonnees-par-les-cross/>

Le site prédisauvetage / Carte SecMar  
<https://carte.snosan.fr/>



# Open Data : Tour d'horizon

par David Rongeat, DSPSI, Amue

## Un peu d'histoire internationale pour commencer

Open Government : l'Open Data est un sujet en lien avec l'"Open Gouvernement" qui porte les valeurs de transparence de la vie publique, concertation citoyenne, coproduction du service public pour améliorer la gouvernance des Etats, faire émerger des contre-pouvoirs et lutter contre la corruption.

Les grands principes de l'open data sont posés en 2007 lors de la rencontre dite de Sebastopol (Californie). En 2013, la [charte](#) sur l'open data du G8 crée les conditions d'une généralisation de l'open data par défaut. Parmi les pionniers signalons la ville de Chicago qui, dès 2005, ouvre les informations sur la criminalité avec un effet d'objectivation de celle-ci et un suivi de son évolution par les citoyens.

## Open Data : révolutionnaire ?

Mais historiquement, les premiers gènes de l'Open Data se trouvent du côté de la France avec l'article 15 de la déclaration universelle des Droits de l'homme et des citoyens (1789) "La société a le droit de demander compte à tout agent public de son administration". Bien plus tard, en 1978 (via la loi 78-753 du 17 juillet 1978) est créée la [CADA](#), Commission d'Accès aux Documents Administratifs, une autorité administrative indépendante qui a pour objectif de faciliter et contrôler l'accès des particuliers aux documents administratifs. Cette autorité peut être saisie par un citoyen auquel une administration n'a pas transmis des documents demandés.

A l'occasion du [discours](#) de Hourtin, le 25 août 1997, le 1er ministre Lionel Jospin proclame "Les données publiques essentielles doivent désormais être accessibles à tous gratuitement sur internet" puis "la diffusion internationale de nos documents publics doit à cet effet être favorisée". En 2011, l'État crée Etalab et ouvre la plateforme [data.gouv.fr](#) ; l'open data débute sa généralisation.

## Un cadre juridique

C'est la [loi numérique de 7 octobre 2016](#) qui pousse à la circulation des données et du savoir. Les données des opérateurs publics de plus de 50 agents ou commune de plus de 3500 habitants (et 50 agents) doivent, depuis la date d'application du 7 octobre 2018, être ouvertes par défaut. D'évidence, les réglementations

"Informatique et liberté" ou le RGPD (Règlement Général sur la Protection des Données) s'appliquent dans ce contexte d'ouverture et participent à la protection des citoyens : la lecture de ce [billet](#) permet de bien voir les interactions entre RGPD et Open Data.

Enfin, pour conclure cette partie légale, un décret du [12 décembre 2018](#) précise les données et documents comportant des données à caractère personnel pouvant être rendus publics sans anonymisation. Dans notre contexte Enseignement Supérieur et Recherche, on retrouve par exemple les résultats aux examens et concours ou plus largement les organigrammes des établissements publics.

### POUR PLUS D'INFORMATIONS

- Loi numérique du 7/10/16 [ici](#)
- Decroix, J. (2017). Vie privée et open data : Harvard propose des règles du jeu à l'adresse [cnil.fr](#)
- Décret du 12 décembre 2018 relatif aux catégories de documents administratifs pouvant être rendus publics sans faire l'objet d'un processus d'anonymisation [Le décret](#)

## Quelques éléments clés

Au-delà de ce mouvement historique et de ce cadre légal, précisons ce qu'est l'Open Data avec [quelques notions](#) clés. Il s'agit bien de rendre accessible à tous (voir article sur données page 8) des données détenues par un opérateur public. Il faut préciser que derrière ce terme "Data" on retrouve également [les documents et code logiciel](#), pas uniquement des données issues des systèmes d'information.

Autre notion importante, il faut considérer qu'un établissement public a un [double positionnement](#) face à l'Open Data : en tant que producteur de données ouvertes mais également en tant que consommateur de données proposées par des tiers.

Les [consommateurs](#) des données ouvertes sont des citoyens, des ONG, des entreprises qui souhaitent créer des services autour de ces données, les médias et les opérateurs publics. Notons pour l'Enseignement Supérieur et Recherche que les étudiants et chercheurs sont parmi les plus grands consommateurs de données ouvertes. D'ailleurs, il est souvent décrit que l'Open Data trouve ses racines dans le monde universitaire (voir article sur Open Science Page 2).

Toutefois, il faut aussi prendre en compte certaines [réticences](#) liées à l'Open Data. Citons par exemple l'accès rendu possible à quelques-une de ses données par un concurrent; Ceci s'applique moins dans le contexte de l'Enseignement Supérieur et Recherche.

Enfin, mettre en place l'open data dans un établissement va au-delà du nécessaire respect de la réglementation. Cela consiste à s'interroger sur son patrimoine informationnel, à fixer une [stratégie](#) d'établissement sur le sujet et à identifier les [services](#) qui pourront être créés autour de ses données ouvertes.

## Stratégie d'établissement et apports de l'Open Data

Cette stratégie d'établissement peut puiser dans les différents apports identifiés pour l'Open Data :

Valoriser le travail et la production de l'établissement et le rendre attractif,

- Servir l'intérêt général,
- Améliorer l'efficacité de certains services,
- Permettre de questionner les politiques publiques,
- Faciliter le croisement de données en s'appuyant sur des référentiels ouverts,
- Pour son Système d'Information (SI) : Mieux connaître les données de son SI, améliorer la qualité des données.

En se plaçant du point de vue du citoyen et de la recherche, l'Open Data apporte de la transparence de l'action publique et éventuellement l'implication des usagers dans des dispositifs de production participative.

## Opportunité de créer de services

Enfin et c'est probablement le point essentiel, ouvrir les données de son établissement permet de créer de nouveaux services pour les usagers : quelques illustrations se trouvent dans l'article sur les usages Page 11.

### POUR ALLER PLUS LOIN

L'Open Data Canvas est une mine de ressources ouvertes  
<https://opendatacanvas.org>



# L'ouverture des données au Ministère de l'enseignement supérieur, de la recherche et de l'innovation

Par Emmanuel Weisenburger, département des outils d'aide à la décision, DGESIP/DGRI,

Sous-direction des systèmes d'information et des études statistiques, MESRI,

En septembre 2019, le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI) propose, sur son [portail dédié](#), intégré à son site institutionnel et synchronisé avec la [plate-forme interministérielle](#), une centaine de jeux de données ouverts de tous types : inscriptions d'étudiants et diplômes délivrés, ressources humaines, coordonnées des infrastructures de recherche et des principaux établissements ou organismes, accréditations de diplômes, lauréats de prix scientifiques, résultats des appels à projets de recherche, événements, agenda du ministre en charge de l'Enseignement supérieur, de la Recherche et de l'Innovation, ... Une interface spécifique à l'état de l'art permet d'explorer facilement l'ensemble des jeux de données (moteur de recherche, filtres), d'exporter ou d'intégrer directement et dynamiquement ses données dans des systèmes tiers (au moyen d'API).

## Initiative innovante

Cette offre est le résultat d'une initiative innovante, à l'échelle des Ministères, initiée en 2013. Cette démarche vise à aligner les principes de diffusion de l'information du Ministère sur le cadre légal émergent qui a progressivement instauré un principe d'ouverture par défaut des données publiques. Au-delà, par un partage large de l'information, le Ministère entend alimenter le débat public sur le système français d'enseignement supérieur, de recherche et d'innovation et sur les politiques publiques qui le soutiennent. Concomitamment, le Ministère perçoit le développement de l'ouverture des données comme une des voies de modernisation de son fonctionnement en favorisant une meilleure circulation interne de l'information.

Dans la mise en œuvre de cette démarche, le Ministère a adopté 3 principes structurants : privilégier la qualité à la quantité de jeux de données, encourager les réutilisations de l'information qu'il propose par la mise en place

de dispositifs d'accompagnement, développer des services et outils de datavisualisation sur les jeux de données clés.

Les jeux de données sont historisés, enrichis d'identifiants facilitant l'interopérabilité des données (introduction d'identifiant SIREN, UAI, Grid, Wikidata, Idref, RNSR,...). Des métadonnées précisent les conditions de production de cette information. Chaque jeu est conçu de manière à "porter" l'ensemble des informations nécessaires à son interprétation et à sa bonne réutilisation ; chaque variable est ainsi dédoublée : une première comprend un code, une seconde un libellé normalisé et compréhensible du public. Pour les jeux de données les plus complexes, un document détaille la structuration du jeu de données variable par variable et précise les conditions d'une utilisation rigoureuse de cette information.

## Visualisation des données

Le MESRI propose pour une grande partie de ses données un ensemble de services et d'outils de datavisualisation (par exemple scanR, moteur de la recherche et de l'innovation) référencés au sein de dataESR, portail recensant l'ensemble des ressources en données qu'il propose. Ces outils permettent aux réutilisateurs de disposer d'éléments de cadrage pour s'autocontrôler en confrontant leurs résultats à des agrégats de référence. Ils ont surtout pour vocation de rendre accessible à tous l'information ouverte qui, si elle était uniquement accessible via des fichiers de données brutes souvent complexes et massifs, seraient réservés aux seuls spécialistes de la donnée. Des outils de datavisualisation concernent par exemple les indicateurs de réussite en Licence, PACES et DUT à l'université, ou encore la mise à disposition de données APB, et bientôt Parcoursup, par formation de l'enseignement supérieur.

C'est ce qui explique certainement le succès de la mise à disposition des données du MESRI. La fréquentation de la plateforme open data ne cesse d'augmenter. Depuis son ouverture, plus de 650 000 fichiers de données ont été téléchargés, dont 175 000 sur les seuls 8 premiers mois de l'année 2019. Mais l'ouverture des données est un processus continu et le MESRI poursuit ses efforts pour élargir son offre, améliorer toujours la qualité des données proposées et automatiser les mises à jour. Il suit également avec intérêt les travaux réalisés en établissement pour s'adapter au cadre légal de la Loi République numérique et développer des offres locales de données. Dans ce contexte, il est d'autant plus crucial que ministère et établissements s'attachent à organiser ensemble les conditions d'une production soutenable de données ouvertes et à développer l'interopérabilité des données exposées.

### LIENS UTILES

<https://data.enseignementsup-recherche.gouv.fr/> : plate-forme ministérielle d'exposition de données ouvertes

<https://data.esr.gouv.fr/> : portail fédérant l'ensemble des ressources (jeux de données, datavisualisation, tableaux de bord, publications, API ...) en données sur l'enseignement supérieur, la recherche et l'innovation



<https://scanr.enseignementsup-recherche.gouv.fr/> : scanR, moteur de la recherche et de l'innovation qui permet d'analyser le paysage français de la recherche et de l'innovation

<https://data.esr.gouv.fr/dataercycle/> : Les parcours et la réussite des bacheliers et des bacheliers en Licence, PACES et en DUT



# Utilisation de données ouvertes du Cnous au service de nos étudiants et personnels

Retour d'expérience par Thierry Cheminade et Philippe Berger, La Rochelle Université

**Historiquement** : En 2014, le CNOUS met à disposition des données en provenance de différents acteurs, principalement des CROUS mais aussi de ses partenaires et de particuliers. Ces données, stockées sur un serveur parking, sont accessibles à tous et disponibles nationalement. Elles concernent les emplois et les logements étudiants ainsi que les descriptifs des restaurants

Enfin, La Rochelle Université se charge de développer deux applications web (de type portlet) basées sur Twitter Bootstrap :

- Esup-crousannonces : pour l'affichage des annonces de logements et des emplois en provenance des centrales Lokaviz et Jobaviz.
- Esup-dining : pour l'affichage des restaurants universitaires et des menus

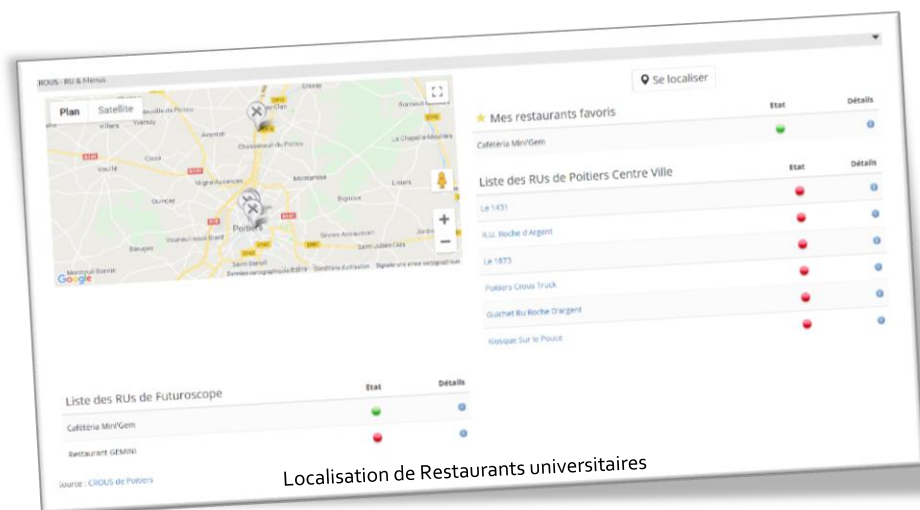
Une fois la portlet déployée et configurée, le service esup-dining pourra :

- Géolocaliser les restaurants universitaires (RUs) les plus proches
- Lister les RUs disponibles dans la zone géographique
- Afficher les informations génériques pour chacun des restaurants : ses coordonnées, son accessibilité, ses horaires d'ouvertures
- Afficher les menus servis

L'utilisateur peut également personnaliser ce service :

- En marquant un restaurant en tant que favori
- En indiquant ses préférences nutritionnelles et/ou les aliments auxquels il est allergique (\*)

(\*) A noter que cette fonctionnalité est disponible dans la portlet mais l'information peut être absente du flux : certaines instances ne souhaitant pas prendre la responsabilité de classifier les aliments allergènes.



universitaires et les menus qui y sont proposés.

Le CNOUS propose déjà un accès à cette masse d'information au travers des centrales Jobaviz et Lokaviz, deux portails qui permettent de consulter les emplois et les logements étudiants. Malheureusement ces services souffrent d'un manque de visibilité et sont méconnus de la population étudiante.

Pour y remédier le CROUS de Poitiers et La Rochelle Université décident de rendre ces informations accessibles au plus près de l'utilisateur en les intégrant dans l'Environnement Numérique de Travail (ENT) à travers de nouveaux services spécifiques.

Au préalable, un travail commun consiste à établir et à structurer le flux des données qui doit être utilisé par ces futurs services : au format XML dans un premier temps, puis en JSON dans un second temps et ce, afin d'optimiser la taille des fichiers et obtenir de meilleures performances.

## Focus sur le service "esup-dining"

Peu de prérequis sont nécessaires pour son installation : simplement disposer d'un ENT EsupPortail / uPortal V4 ou supérieur et d'un SGBD (PostgreSQL ou MySQL).

Quelques réglages sont nécessaires avant sa mise en route : indiquer le rôle utilisateur qui pourra accéder au service en mode administrateur (back office), définir l'URL du flux JSON distribué par le CROUS et correspondant à la zone géographique de votre établissement.

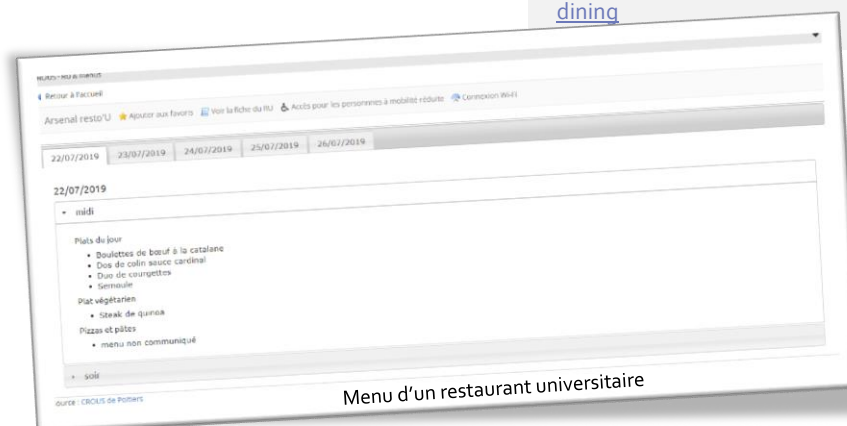
### POUR PLUS D'INFORMATIONS

Esup-crousannonces et esup-dining sont estampillés "esup" car La Rochelle Université a mis ces services à la disposition du Consortium ESUP-Portail et sont donc proposés dans leur catalogue.

Vous pouvez retrouver les sources sur le github d'ESUP-portail :

<https://github.com/EsupPortail/esup-crousannonces>

<https://github.com/EsupPortail/esup-dining>

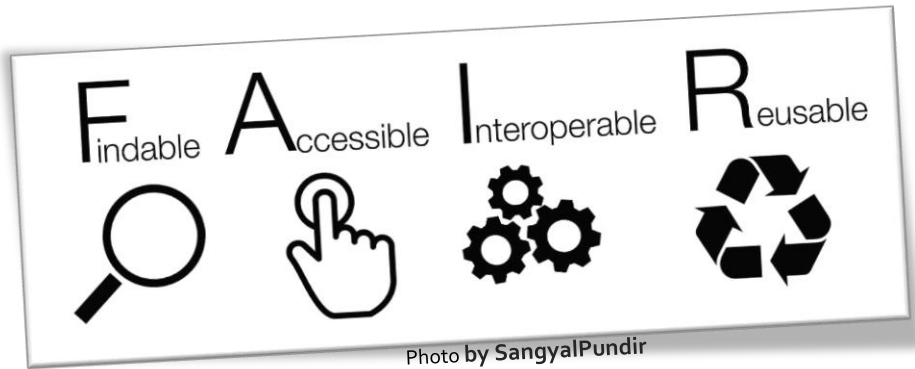


# Les données au cœur de l'Open Data Par Amue-DSPSI

L'acronyme FAIR (Findable Accessible Interoperable Reusable) DATA résume à lui seul quelques-unes des principales caractéristiques que devraient présenter les données ouvertes. Nous allons les détailler ici.

les jeux de données ne doivent pas utiliser des formats de fichiers propriétaire (tels que xls par ex) mais bien des **formats ouverts** (CSV par exemples).

de données **interopérable**. Ces référentiels, eux même publiés en open data, sont par exemple ceux publiés par l'Insee, [la base adresse nationale](#), les UAI (Unité Administrative Immatriculée) qui remplacent le code RNE, ...



## Principales caractéristiques

Les données ouvertes doivent être avant tout des données **primaires** (au sens n'ayant subi aucune transformation), non agrégées tant qu'elles respectent les exigences réglementaires (RGPD, anonymisation par ex). La temporalité est également un point important : ouvertes au plus tôt (notion de **fraîcheur**), il est important de respecter la **temporalité** annoncée pour son jeu de données pour faciliter le traitement par les consommateurs. Ces deux notions rejoignent le thème du cycle de vie de la donnée. Enfin le caractère complet (**complétude**) et **qualité** viennent compléter les caractéristiques des données ouvertes.

L'éligibilité d'un jeu de données à l'ouverture se mesure à l'aune de ces caractéristiques.

## Ouvertes donc accessibles

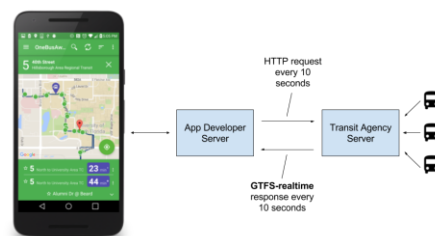
En utilisant la plate-forme data.gouv.fr (voir article page 8), le producteur de données est incité à **documenter** son jeu de données, le rendant ainsi beaucoup plus compréhensible par un tiers. Cette documentation est pour le producteur l'occasion de décrire son jeu de données (schéma, fréquence de publication,...) et de maîtriser la stratégie de ré-utilisation de ses données par un ou des consommateurs.

Cette plate-forme permet aussi de rendre les jeux de données **accessibles** et utilisables par le plus grand nombre d'utilisateurs potentiels et les rend plus facilement exploitables automatiquement par des machines. Enfin

## Standard et ouverture

Les données ouvertes doivent s'appuyer sur des **licences ouvertes** permettant leurs utilisations par des tiers. Pour bien choisir sa licence suivre les indications de <https://www.data.gouv.fr/fr/licences>, pour sa part l'Amue a choisi, pour ses données ouvertes, la licence d'Etat [Licence Ouverte 2.0](#). La réutilisation sera également facilitée si les jeux de données s'appuient sur des **standards**. Citons par exemple le standard GTFS pour publier des jeux de données de transport (voir [ici](#)) ou pour les données géographiques se référer à la page spécifique de [geo.data.gouv.fr](http://geo.data.gouv.fr)

### FLUX TEMPS REEL POUR LES TRANSPORTS



*Flux d'information en temps réel typique d'une agence de transport vers une application mobile au format GTFS*

## Référentiels

Pour simplifier l'utilisation des jeux de données et, surtout, le croisement entre des jeux de données il est important que les données constitutives s'appuient sur des **référentiels**. Cela rend naturellement le jeu

## Dans une cible SI Décisionnel

Comment traiter des aspects de standardisation, référentiels, qualité de données sans évoquer le cas d'un projet de SI décisionnel dans un contexte Open Data : il serait à la fois consommateur de données ouvertes par d'autres (référentiels et indicateurs externes par ex) mais également un point central d'ouverture des données pour peu qu'elles demeurent non agrégées. Ces données bénéficieront alors des apports en cohérence transverse métier d'un projet Décisionnel (qualité, identification des sources, horodatage des données, standardisation par ex). En fonction de leur stratégie, des établissements pourraient même ouvrir leurs indicateurs en open data.

### POUR ALLER PLUS LOIN

Le modèle de déploiement en 5 étoiles de Tim Bunders Lee > à voir sur <https://5stardata.info/fr/>



### POUR ALLER PLUS LOIN

De L'Open Data au Big Data par la société coopérative Dataactivist

<https://dataactivist.coop/dataliteracy>





# L'ouverture pour affirmer sa souveraineté sur les données publiques

Par Jérémie Valentin, chef de projet Open Data, Montpellier Méditerranée

Métropole

## Une première phase lancée en 2011

Initiée en 2011-2012, la politique d'ouverture des données de la Métropole de Montpellier était avant tout axée sur la création potentielle de services innovants. Pour se faire l'objectif reposait sur l'accumulation le plus rapidement possible d'un nombre important de données quel que soit la qualité et la valeur de ces dernières. Cette première phase ignorait le rôle de l'open data comme facteur de transparence de l'action publique et comme outil de synergie interne valorisant le travail des agents de la collectivité. Huit années après, la quantité et surtout la qualité permettent aujourd'hui d'affirmer qu'une politique d'ouverture des données peut, entre autre, aboutir à la création de services innovants pour la collectivité et son territoire.

## Un portail pour les 31 communes

Aujourd'hui la Métropole de Montpellier a dépassé, mais pas abandonné, cette quête obsessionnelle de la création de services grâce à l'évolution du cadre réglementaire en 2016 (Loi pour une République Numérique) et la Métropole se retrouve aujourd'hui dans la proposition d'un service public de base, utile aux entreprises, chercheurs et citoyens. Ce changement de paradigme a entraîné la Métropole de Montpellier dans une nouvelle phase du projet, en proposant le portail open data comme une offre de service gratuite pour les 31 communes de notre territoire. L'idée est de regrouper et d'homogénéiser les données des 31 communes sur un portail unique. Ce travail d'acculturation à la donnée s'est accompagné d'une prise de conscience sur la qualité de la donnée et sa fraîcheur. C'est ainsi que la Métropole a multiplié les initiatives pour ouvrir des données à forte valeur ajoutée tout comme des données dont la mise à jour se fait en temps réel. Il s'agit là des conditions minimales pour voir la création de service avec des modèles économiques pérennes.

## Vers une cité intelligente

Parallèlement il est devenu clair que l'ouverture des données ne représentait pas à elle seule les enjeux de la donnée pour une collectivité, mais qu'elle était une brique de la politique de cité intelligente que la Métropole de Montpellier expérimentait depuis 2010. En effet une plateforme open data reste un dispositif technique et il est nécessaire d'organiser son administration afin de répondre aux problématiques de la smart city qu'elles soient internes ou externes.

Aujourd'hui dans sa feuille de route Cité intelligente, la Métropole de Montpellier se "tourne vers la gestion de la donnée urbaine, pour s'assurer de sa qualité, de sa pérennité et de sa diffusion". La Métropole se pose en tiers de confiance sur le respect de la vie privée, de l'indépendance vis-à-vis d'acteurs tiers et de la diffusion des données. Afin d'appliquer cette indépendance et sa souveraineté sur les données, la Métropole applique un accès non privilégié et non discriminant à ses données, via l'unique canal de diffusion de l'open data.

L'OPEN DATA A MONTPELLIER  
MEDITERRANEE METROPOLE

 **OPENDATA**  
Montpellier Méditerranée Métropole

  
Montpellier  
Méditerranée  
métropole

Le portail Open Data de Montpellier  
Méditerranée Métropole :

<http://data.montpellier3m.fr/>



Source : data.montpellier3m.fr



# La réutilisation des données publiques

Par Jean-Marie Bourgogne, délégué

général de Open Data France

Engagées dans l'open data depuis près de 10 ans, plus de 450 collectivités territoriales ont déjà ouvert leurs données publiques. Des milliers de jeux de données sont ainsi la disposition de citoyens, de chercheurs ou d'entreprises qui les exploitent plus facilement, peuvent créer de nouveaux usages ou enrichir des services existants.

Les collectivités ont encore du mal à communiquer sur les usages qui sont fait de ces données en "open data". Quelques sites présentent ici ou là des cas d'usages exemplaires mais il ne s'agit bien souvent que d'une petite partie visible des usages de l'open data. Une étude plus approfondie nous montre que de très nombreux services utilisent régulièrement de multiples sources de données dont une grande partie sont disponibles en open data. Et les réutilisateurs ne sont pas exclusivement des starts-up mais une infinité d'acteurs publics, associatifs ou privés incluant des entreprises anciennes ou très puissantes. Dans le cadre de son observatoire open data des territoires, OpenDataFrance a publié une synthèse de cas de réutilisation des données publiques à travers des dossiers thématiques (mobilité, environnement, emploi, etc.) et des entretiens, montrant ainsi toute l'utilité des données ouvertes et le dynamisme des écosystèmes liés à leur utilisation. (<https://frama.link/reutilisationsOD>)

Quels sont les principaux enseignements de ce travail d'enquête ?

## 1. L'iceberg des réutilisations

Les usages de l'open data ressemblent à un iceberg. Il y a d'un côté une partie visible, évidente et en particulier les réutilisations des collectivités elles-mêmes. Mais de l'autre, notamment quand on dépasse les seules plateformes locales, des usages moins connus. Certains usages (données géographiques en particulier) préexistaient à l'open data et seul le mode d'accès a changé. D'autres sont mal cernés du fait du manque de traçabilité des téléchargements et API. Aussi si l'on s'en tient aux réutilisations déclarées - telles que signalées sur les portails des collectivités - elles peuvent paraître faibles, si l'on considère ces réutilisations "cachées" elles sont beaucoup plus significatives.

## 2. Quelques secteurs tirent les usages

Tous les jours des millions de personnes sollicitent des applications de calcul

d'itinéraire, de GoogleMaps à City Mapper en passant par toutes celles créées par des territoires ou des start-ups. Ces applications reposent très largement sur de l'open data. Dans le domaine de l'urbanisme également, architectes, urbanistes, aménageurs sont d'énormes consommateurs de données territorialisées que l'open data a contribué à rendre plus aisément exploitables. La force d'entraînement de ces domaines est considérable avec un effet tâche d'huile sur des sujets connexes. On citera le stationnement, le vélo ou l'auto-partage pour la mobilité. Mais c'est aussi vrai pour toutes les données associées à la parcelle (réglementation, adressage, desserte réseau...) ou encore les événements (bâtiments publics, fréquentation, handicap...).

## 3. Un phénomène d'intermédiation

Apidae, Navitia, OpenDataSoft, Ozwillo, OpenAgenda... et toutes les plates-formes de données thématiques sont très appréciées des réutilisateurs à qui elles simplifient la recherche et l'exploitation de données. Leur force ? réunir sur une seule plate-forme des données ouvertes (et parfois non ouvertes) et les proposer sous forme d'API ou de widgets prêts à l'emploi. Ces plateformes ont cependant pour conséquence de faire écran : les collectivités connaissent mal les réutilisations des données exploitées via des API. Et dans certains cas, les réutilisateurs eux-mêmes ne savent pas nécessairement que le service (payant pour certaines plateformes) est produit à partir de données ouvertes.

## 4. Les "monsieur Jourdain" de l'open-data

De plus en plus de services sont créés par des développeurs à partir de "briques", qu'il s'agisse de logiciels ou de données. Si ces informaticiens connaissent les API, ils ne sont pas nécessairement sensibilisés à l'open data. Le cas d'OpenStreetMap, boosté par l'augmentation des tarifs de GoogleMaps, est exemplaire : on est passé d'une contribution militante à une contribution massive (y compris par des institutions) sans que les contributeurs sachent nécessairement qu'il s'agit là d'une des plus grandes bases open data au monde. Produire de l'open data (et l'utiliser) sans le savoir est peut-être un signe de maturité d'un secteur qui a quitté le militantisme.

## 5. Crowdsourcing et open data : un cercle vertueux

Au-delà d'OSM, de plus en plus de projets open data sont associés à du crowdsourcing, autrement dit à une mobilisation des habitants ou des acteurs locaux pour enrichir et compléter des "communs" de données. Cette mobilisation, facilitée et renforcée par la généralisation des smartphones, est particulièrement forte dans les domaines du handicap, de la mobilité ou encore du tourisme. La mobilisation des habitants ou des acteurs locaux concourt à la création d'un cercle vertueux car plus une donnée est utilisée, plus elle est corrigée ou enrichie, plus elle sera fiable et donc (ré)utilisée.



L'association [OpenDataFrance](https://opendatafrance.org/) a été créée en octobre 2013. Elle a pour but de regrouper et soutenir les collectivités engagées activement dans une démarche d'ouverture des données publiques et de favoriser toutes les démarches entreprises par ces collectivités pour promotion l'open data.

Animateur et fédérateur, elle agit en mode projet :

- apporter à ses membres des informations, de conseils ou autres soutiens nécessaires à l'ouverture des données publiques ;
  - favoriser les productions communes par des groupes de travail ou ateliers sur des thèmes définis en commun comme devant apporter des éléments de réponse aux questions qui se posent sur l'open data et sa mise en œuvre ;
- Elle a ainsi produit un référentiel pédagogique et méthodologique librement accessible en ligne, elle a mis en place un ambitieux programme d'accompagnement local des collectivités (OpenDataLocale), elle a enfin mis en œuvre un ensemble d'outillages visant à la simplification de la production, à l'amélioration de la qualité et à la standardisation des données publiques (OpenDataFactory).
- représenter ses membres auprès de toute autorité publique et privée dans le but d'assurer le développement harmonieux de l'open data.



# Concrètement, qu'apporterait l'open data aux établissements ? Une compilation d'usages

Par Amue-DSPSI



Photo par justgrimes via visualhunt.com

Impossible d'être exhaustif dans le domaine de l'Open Data, tant les cas d'usages sont variés. Mais rien de tel que de présenter des cas concrets, réalisés ou prévus, principalement dans le domaine de l'Enseignement Supérieur et Recherche pour aider à appréhender le sujet de l'Open Data.

## En consommant des données ouvertes pour améliorer la qualité de ses données

Consommer des référentiels ouverts, comme les données ouvertes par l'Insee sur les informations des entreprises permet de qualifier, corriger, expurger les données liées aux entreprises dans son Système d'information (fournisseurs, CRM, référentiel pour les stages,...). Dans le même état d'esprit de mise en qualité de ses données, exploiter la Base Adresse Nationale est un moyen de contrôler la qualité de ses données d'adresse : une seule 'adresse' pour cela : <https://adresse.data.gouv.fr/>. Logique à prendre en compte au plus tôt dans ses projets de SI pour garantir des données de qualité dès la saisie.

Dans ce même objectif de qualité de ses données, leur ouverture peut inciter des consommateurs à demander des corrections ou des compléments sur les données publiées. Pour peu qu'un processus de prise en compte de ces retours soit mis en œuvre, l'ouverture peut devenir un vecteur de

correction de ses propres données. Citons comme exemple la diffusion de géolocalisation de stations de transports en commun que les usagers pouvaient proposer de mettre à jour, correction participative par des usagers.

## Pour de meilleurs services aux usagers

Informers les opérateurs de services aux usagers des variations de flux de population (arrivées massives pour un colloque international, horaires exceptionnels pour les examens, fermeture temporaire de campus ou de site, séminaires...) par campus peut les aider à adapter leurs services. L'Open Data est un vecteur possible de cette transmission d'informations. Les opérateurs de transport en commun dans les villes moyennes qui desservent des campus peuvent ainsi optimiser leur service pour une meilleure qualité de service aux usagers (étudiants, personnels et invités) en adaptant les horaires, fréquence à la réalité des présences. Ces mêmes informations normalisées peuvent aussi aider les services de restauration en adaptant leur offre au besoin.

## Un gain pour l'employabilité et l'attractivité d'un territoire.

Pour attirer les entreprises qui recrutent des cadres, un établissement pourrait publier, en anticipation, le nombre et qualification de ses prochaines cohortes de diplômés. Dans le même esprit, cette idée proposée par un établissement : ouvrir les données, pour son territoire, qui présentent les laboratoires, les sujets de recherche, nombre de chercheurs, grands équipements, ... pour attirer des talents et des entreprises dans une zone géographique.

Publier, pour un établissement, les données d'insertion professionnelle de ses anciens étudiants est un vecteur de valorisation de ses formations qui peut être repris pour montrer les débouchés d'un diplôme. En poussant plus

### POUR ALLER PLUS LOIN

Données ouvertes par Pôle Emploi  
<http://www.pole-emploi.org/accueil/actualites/infographies/open-data-des-donnees-pour-eclair.html?type=article>



avant cette logique, des recruteurs pourraient utiliser ces données (quels emplois pourvus dans le temps par quels diplômés, de manière anonyme) pour identifier les diplômés qui ont réellement abondés certains types d'emplois et ainsi cibler par emploi à pourvoir les diplômés requis.

## Information de l'institution

Il est aisé de publier, en Open Data, des documents institutionnels de l'établissement (Compte rendu de CA par exemple), tout comme des données sociales et RH issues du bilan social, permettant par exemple de mettre en valeur une parité optimale. Dans une logique de transparence et une volonté stratégique singulière, un établissement pourrait publier les subventions qu'il octroie aux différentes associations.

Les informations des événements organisés par l'établissement ou des associations universitaires peuvent être utiles aux acteurs territoriaux (transport, valorisation d'une agglomération, service culturels,...) ou pour des applications à destination des citoyens,... : des dispositifs existent déjà, c'est un avantage de l'open data, et cette diffusion pourrait se faire en s'appuyant sur le projet [OpenEventDataBase](#).

L'université de Paris Nanterre publie son offre de formation sur le site [data.gouv.fr](http://data.gouv.fr). Bien que ces informations soient généralement accessibles sur les sites institutionnels, la mise en open data permet de rendre ces données exploitables pour des analyses, des rapprochements de données,...

Quant aux données d'inscription et de diplômés elles sont automatiquement



publiées en open data par le MESRI via son site [#DATAESR](#) (voir-article page 6).

## Des usages plus prospectifs

Poussée à son paroxysme et généralisée à l'ensemble de l'ESR l'ouverture des données pourrait devenir un moyen de répondre aux enquêtes plus simplement, et permettrait sans trop d'efforts administratifs de s'autoévaluer par rapport aux autres établissements ressemblants d'un point de vue organisation et missions.

## Quelques usages hors Enseignement Supérieur et Recherche

Grandes consommatrices d'open data, les applications de mobilité optimisent aujourd'hui des trajets en fonction des places de parking réellement disponibles (données ouvertes en temps réel), indiquent les emplacements des vélos libres services et leurs disponibilités, se mettent à jour en fonction des informations de trafic en temps réel, en vous prévenant de l'arrivée prochaine d'un tramway. Nous parlons ici d'usages

existants dans les métropoles. La logique de villes intelligentes ("Smart cities") qui exploitent le numérique pour améliorer la qualité des services aux usagers est inspirant pour de futurs campus intelligents, qu'ils restent à inventer.

Arrivé récemment en haut du classement des jeux de données les plus utilisés, le DVF "Demandes de valeurs foncières" permet de connaître les valeurs de transactions immobilières par adresse : fortement utilisé par des vendeurs, futurs acheteurs et professionnels pour connaître les prix réels du marché immobilier quartier par quartier, cette ouverture de données a été une petite révolution dans le monde de la vente de biens immobiliers. Enfin des projets importants comme [OpenStreetMap](#) (projet cartographique en ligne et mondial, concurrent ouverte de googlemap) ou [openfoodfacts](#) (répertoire des produits alimentaires du monde entier) allient Open Data et contribution collaborative des citoyens (production participative ou "crowdsourcing").

## Pour conclure

Ces quelques exemples peuvent être complétés par vos témoignages et bonnes idées à adresser à [numerique@amue.fr](mailto:numerique@amue.fr). Tous ces cas d'usages, retours d'expériences pourraient donner lieu à une publication "les belles histoires de l'open data ESR" en mettant en avant les gains pour les usagers par exemple.

### POUR IMAGINER VOS SERVICES

Vous pouvez vous rendre sur [data.gouv.fr](http://data.gouv.fr) et regarder quelques-uns des 38'000 jeux de données

# Hébergement, technologie, format : le côté obscur des données ouvertes

Une recommandation par Amue-DSPSI

## En local...

Dans la construction de sa stratégie open data, chaque établissement pourrait rechercher sa propre plateforme de diffusion des données. Une solution du marché spécialisé dans la diffusion de données, un domaine sur le site institutionnel de l'établissement, un serveur dédié et ouvert au monde, un partenariat avec la métropole, ....

Ce questionnement est tout à fait légitime pour le long, voire moyen, terme mais dans les premières phases d'implémentation de l'open data, il semble important de traiter d'autres aspects que la création d'un environnement technique dédié et individualisé. L'amue a entamé une réflexion pour s'associer avec un établissement afin de proposer un squelette de diffusion, paramétrable par chaque adhérent selon sa stratégie numérique singulière.

## ... ou en national sur le site Data.gouv.fr

Le moyen le plus simple pour valoriser et partager ses données ouvertes est de s'appuyer sur la plate-forme portée par Etalab (voir article page 3) : [data.gouv.fr](http://data.gouv.fr). Il s'agit en France de la plate-forme de référence, lieu où chacun des opérateurs peut déposer ses données et où les consommateurs viennent en premier lieu les collecter. L'utilisation en est aisée, le site est [documenté](#) et le mécanisme de déclaration d'un jeu de données incite à bien les structurer et les documenter, facilitant ainsi leur utilisation et référencement.

## Publier, mais selon quel format ?

Cette plate-forme d'État comprend également toute la documentation

permettant de choisir le format technique le plus adapté à son contexte et à chacun de ses jeux de données. Des formats/standards tels que CSV, JSON, XML sont précisés mais également des formats spécifiques par nature de jeux de données (données de transport, données géographiques, ...).

## Quelques technologies

Hormis les dépôts et mises à jour manuelles des jeux de données, le site [data.gouv.fr](http://data.gouv.fr) propose l'utilisation des technologies de moissonnage ou d'API par exemple. Cela permet d'automatiser les interactions entre son Système d'Information et le point de référence Open Data qu'est [data.gouv.fr](http://data.gouv.fr).

A l'inverse si vous souhaitez présenter sur une page web un ou plusieurs jeux de données, cela est documenté [ici](#). Vous y trouverez même une librairie JavaScript, metacliq, permettant de présenter votre catalogue de données ouvertes issu de [data.gouv.fr](http://data.gouv.fr).



# Le catalogue de données ouvertes du réseau des

## CROUS Par Stéphane Papon, Référent OpenData au CNOUS

Le réseau des CROUS possède depuis plusieurs années plusieurs webservices et flux de données ouvertes mis à la disposition de la communauté de l'enseignement supérieur et de la recherche.

Le début de notre réflexion sur la mise à disposition de nos données est intervenu à la suite d'une plainte d'étudiant ayant vu sur un site universitaire des informations sur un logement qui étaient erronées. Un personnel pourtant consciencieux avait ressaisi manuellement avec exactitude les informations provenant du site internet du CROUS local mais au fil du temps la page du site laissée en plan, sans mise à jour, est devenue diffuseuse d'informations caduques.

Il fallait donc permettre à nos partenaires de disposer de nos données saisies au plus près de l'information.

Cela devrait se faire facilement, librement et avec un processus de mise à jour en temps réel pour que les données soient toujours fiables.

Les premiers flux "restaurants" furent générés par le site pilote du CROUS de Poitiers pour les besoins d'un travail en commun avec l'université de La Rochelle dans le cadre du développement de son ENT (voir article Page 7). Les flux de données permettant ainsi d'afficher dans l'ENT universitaire les informations des restaurants (horaires, géolocalisations, informations & menus).



Résidences universitaires

Source : CNOUS

Dans le prolongement de ce projet nous avons décliné le flux restauration pour les autres services du CROUS sur la région Poitou-Charentes, ainsi furent générés entre autres les flux logements, actualités, annonces de jobs, annonces de logements...

Notre système de diffusion a toujours été pensé et réfléchi pour une portée nationale et pas seulement locale. Donc nous avons pu dupliquer très rapidement ces flux pour l'ensemble des Crous.

Aujourd'hui, ces données ouvertes sont consommées majoritairement par notre système d'information (Crous Mobile, affichage dynamique, site web ...) mais aussi par différents partenaires : des universités pour leur ENT ou leurs applications mobiles, des communes ou des communautés d'agglomération pour leurs sites web, affichage dynamique ou leurs applications mobiles. Nos données sont aujourd'hui générées par des applications spécifiques au plus proche du terrain et de l'information, et stockées sur nos propres serveurs. Nous disposons d'une réplique sur data.gouv.fr de certains flux (n'oubliez pas de vous inscrire comme abonné quand vous consommez nos données, cela nous permet de vous tenir au courant des évolutions).

### Un exemple pour la restauration

Le secrétariat des restaurants dispose d'une application développée par le réseau des Crous qui lui permet de saisir toutes les informations nécessaires au fonctionnement du restaurant (horaires, menus, ouverture/fermeture, évènements ...).

Cette application lui sert entre autres fonctionnalités à générer ses menus en PDF pour envoi de mails pour les sites partenaires, mais surtout elle génère les flux qui sont envoyés sur le serveur open-data des Crous et consommés par le site web, les applications mobiles, l'affichage dynamique...

L'efficacité de ce système permet, par exemple lorsque qu'un défaut d'approvisionnement est constaté en cuisine - impactant de fait le menu du jour - de modifier l'affichage en temps réel sur tous nos canaux par une seule modification de saisie par le secrétariat du restaurant.

### Les ressources disponibles aujourd'hui

- Le catalogue de données disponible en septembre 2019 est composé de flux et de

webservices : Flux restaurant (XML/JSON) géolocalisations, horaires, informations... , Flux menu (XML/JSON), Flux logement (XML) géolocalisations, horaires, informations..., Flux actualité (XML) actualités régionales du CROUS (photo+texte), Flux aides/bourses (XML) (information), Flux événement (XML en cours de finalisation)

- webservice restaurant : Géolocalisations, menus, horaires, informations...

- webservice annonce logement : Affichage des X dernières annonces de logement, possibilité de choisir la zone France/région/département/villes universitaires

- webservice annonce job : Affichage des X dernières annonces de job étudiant, possibilité de choisir la zone France/région/département/villes universitaires

### Les ressources disponibles demain

Les développements et évolutions à venir sur le dernier trimestre 2019/premier trimestre 2020 :

- La migration de l'ensemble des flux XML en JSON sur le même principe que les flux restauration,

- La mise en place de webservices sur l'ensemble des flux,

- La migration du serveur avec la mise en place d'un serveur open data plus en phase avec l'état de l'art et les besoins actuels.

Nos flux sont vivants, n'hésitez pas à vous signaler si vous les exploitez ou souhaitez les consommer, cela nous permet de vous tenir informés des évolutions (sans engendrer une coupure de service de votre côté). Nous sommes aussi à votre écoute si vous avez besoin de plus de contenus, ou que vous souhaitez la mise en place de nouveaux jeux de données Crous. Nous sommes aussi favorables et ouverts aux échanges qui nous permettent de rendre plus performants techniquement et qualitativement nos flux. Nous réalisons à la demande, si vous avez des besoins, de flux spécifiques quand ils peuvent être réutilisables pour l'ensemble de la communauté (exemple : les flux agrégeant l'ensemble des restaurants et logements Crous pour la France entière qui sert d'appui aux cartes du site [etudiant.gouv.fr](http://etudiant.gouv.fr)).

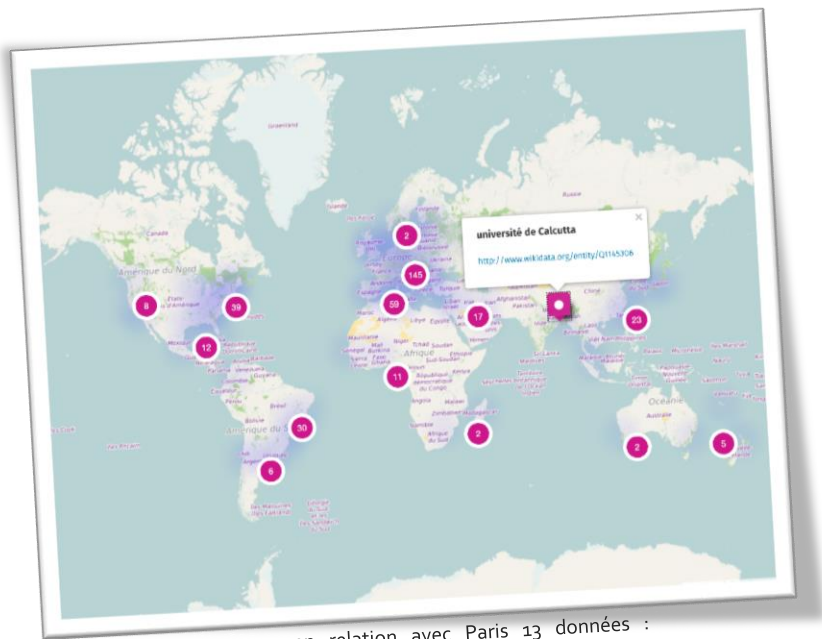
Pour contacter le CNOUS A CE SUJET

opendata@crous.fr  
stephane.papon@crous-poitiers.fr



# Référentiels collaboratifs en données ouvertes par Pierre Boudes, VP SI

université Paris 13



Établissements étrangers en relation avec Paris 13 données :  
Wikidata, Carte : contributeurs OpenStreetMap

Nous avons tous dans nos systèmes d'informations des références à des données dont nos établissements ne maîtrisent pas la production. Par exemple, des références à des lycées. Si les attributs de ces lycées (nom, adresses, etc.) changent, nous ne sommes pas prévenus. C'est pour cela que nous préférons consommer un référentiel national plutôt que d'accepter la charge d'entretenir ces données. En interne, il est également très inconfortable pour un service d'être responsable de la maintenance d'une donnée sans être partie prenante du processus qui modifie l'information. Améliorer les processus, ouvrir les silos, interconnecter les applications sont autant de façons de traiter ces frictions

## Des référentiels sans autorité centrale ?

Mais comment faire lorsque les données ne sont pas produites sous le contrôle d'une autorité centrale ? Le cas se pose pour les établissements supérieurs étrangers que nous devons référencer, par exemple, pour la composition des jurys de thèse. Une bonne solution est alors de passer à un référentiel international ouvert et collaboratif.

## Les infrastructures essentielles du libre savoir

Wikipedia, créé en 2001, se révèle particulièrement efficace pour produire collectivement des informations raisonnablement fiables et de bonne qualité. Une partie de ces informations sont semi-structurées et DBpedia en tire profit pour les rendre disponibles sous forme de données ouvertes liées (voir encadré), constituant ainsi depuis 2007 une brique centrale dans la révolution silencieuse du web sémantique.

Wikidata complète le processus en permettant aux contributeurs et contributrices de directement produire et mettre en qualité les données structurées, dans une logique de référentiel. Les assertions au sujet d'un élément y sont souvent étayées par un lien vers la source d'information et contribuer est particulièrement facile.

À l'université Paris 13, dans le cadre du déploiement de [Sinaps](#) (solution Amue de gestion des données de référence), il fallait complètement réviser nos données sur les établissements étrangers. Nous avons fait le choix d'utiliser Wikidata comme référentiel externe. Ainsi nous n'endossons pas la responsabilité de suivre les changements concernant ces établissements (nom,

adresse, fusion, etc.) et une simple recherche dans [Qwant](#) assortie du mot clé &wd permet de (re)trouver un établissement rapidement. Auparavant, il fallait produire la donnée complète à partir de documents en langue étrangère ce qui occasionnait des erreurs comme une université British de code pays Colombie (sauriez-vous retrouver l'erreur ?). Au moment de construire des universités européennes, Wikidata fournit un référentiel commun multilingue et de confiance.

### WIKIDATA



Chaque donnée est un triplet sujet-propriété-objet qui, dans une table, correspondrait à la cellule à la ligne "sujet" et à la colonne "propriété" contenant la valeur "objet". Mais sans table ! Cela permet d'organiser des données hétérogènes et d'associer au sujet plusieurs valeurs pour une même propriété. Ces triplets sont communiqués de façon ouverte et interopérable rendant possible des requêtes multi-bases arbitrairement complexes. Les identifiants (des URI) sont requêttables en HTTP et désignent les sujets mêmes et non leurs codifications dans une base. Autour de [Wikidata](#) (60 millions de sujets), beaucoup de données liées sont directement exploitables avec les outils produits par la communauté, à retrouver sur <https://tools.wmflabs.org/>.



# Amue et Open Data à l'unisson dorénavant

Par Amue-DSPSI



La page OpenData de l'Amue

L'Amue s'est positionnée sur l'Open Data sur deux types d'actions : sensibilisation et anticipation

## Sensibilisation

Le premier axe consiste à accompagner les établissements par des opérations de sensibilisation, de partage d'expérience dont ce numéro de la collection numérique est la première concrétisation. Une journée "Open Data" est programmée (voir encart) le 14 novembre 2019 à la maison des universités. La matinée sera consacrée à une vue d'ensemble de l'Open Data adaptée au contexte de l'Enseignement Supérieur et de la Recherche. Des ateliers de travail seront organisés l'après-midi. Ils visent à produire collectivement une 1<sup>ère</sup> liste de jeux de données à ouvrir en analysant plus-value,

### POUR VOUS INSCRIRE....

.. à la journée Open Data organisée par l'Amue : le 14 novembre 2019 > [s'inscrire](#)

facilité de mise en œuvre, sensibilité, ...

## Anticipation par mutualisation via les outils et ...

Le second axe vise à anticiper les demandes d'ouverture de données contenues dans les solutions logicielles Amue. En effet, une partie du patrimoine informationnel des établissements est géré dans ces outils et en mutualiser l'extraction et format sera un gain

d'efficacité pour les établissements qui mettent en œuvre leur politique Open Data. Ses travaux permettront aux solutions existantes (Apogée, Sifac, Siham, Sinaps, ...) de proposer, sous le contrôle des experts métiers, des publications de données ; ils permettront également pour les solutions en cours de construction (Pc-Scol/Pégase, Caplab) de le prévoir dans leur feuille de route.

## ...des schémas de données

Les données ouvertes se caractérisent par des schémas de données documentés (quels sont les différents champs, comment sont représentées les données, quelles sont les valeurs possibles etc.). Ceux-ci seront évidemment mutualisés pour ce qui concerne les travaux autour des solutions logicielles Amue ou co-construite avec des partenaires (Cocktail et CNRS). Il semble également intéressant que les établissements mutualisent et partagent les schémas de données qui ne seraient pas directement dans les domaines de gestion couverts par les solutions Amue. Enfin, le projet SI décisionnel communautaire sera également partie prenante pour partager son expertise en qualité de données, en mutualisation de schémas de données et surtout dans la cohérence transverse de la donnée.

### POUR PLUS D'INFORMATIONS

site de référencement des schémas de données publiques pour la France porté par Etalab <https://schema.data.gouv.fr/>

```

Fields:
  0:
    name: "id_lieu"
    description: "Identifiant du lieu de covoiturage, délivré par le point d'accès national"
    exemple: "35238-C-001"
    type: "string"
    constraints:
      required: true
      pattern: "^[013-9]\\d{2}[01-9]\\d{3}-C-\\d{3}$"
  1:
    name: "com_lieu"
    description: "Le commune / le lieu-dit du covoiturage"
    exemple: "Rouen"
    type: "string"
    constraints:
      required: false
  4:
    name: "insee"
    description: "Le code INSEE de la commune d'implantation"
    exemple: "76540"
    type: "string"
    constraints:
      required: true
      pattern: "^[013-9]\\d{2}[01-9]\\d{3}$"

```

### Exemple de schéma de données

Au plus les schémas sont communs et partagés sur l'ensemble du territoire, au mieux les consommateurs peuvent proposer des services à valeur ajoutée par la consommation de ces données. Cette démarche de mutualisation des schémas rejoint la démarche d'Etalab par son site <https://schema.data.gouv.fr/> qui permet de référencer des schémas de données (voir encart).

Cette logique de travail mutualisé par et pour les établissements permettrait de constituer un socle commun aux établissements pour faciliter leur démarche Open Data. Des groupes de travail pourraient co-construire ce socle.

## Un nouveau site web : Data.amue.fr

Comme tout établissement public de plus de 50 agents, l'Amue est concernée par l'Open Data pour ses propres données. Aussi, dans une logique progressive, 3 premiers jeux de données sont ouverts. Il s'agit de la liste de nos adhérents, de la collection numérique et enfin d'un jeu de données proposant les documents de plan d'action et rapports d'activités. Ces jeux de données sont publiés sur le site de l'état [data.gouv.fr](https://data.gouv.fr/). Une page sur le sujet de l'open data a été mise en ligne : [data.amue.fr](https://data.amue.fr/) : elle présente succinctement le sujet et référence les jeux de données publiés par l'Amue.



# Retour d'expérience sur un projet Open Data en établissement

Par Lionel Maurel, ancien chef de projet, Université Paris Nanterre

## Open Data : n'oublions pas les données administratives !

Lorsqu'on évoque la question de l'Open Data dans le champ des établissements d'enseignement supérieur et de recherche, c'est généralement au sujet des données de recherche. Le [Plan National pour la Science Ouverte](#), publié par le Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation en juillet 2018, contient ainsi un axe intitulé "Structurer et ouvrir les données de la recherche", tandis qu'un nombre croissant d'établissements se dotent de politiques de partage et d'entrepôts institutionnels visant à les stocker et à les diffuser.

Ce mouvement en faveur de l'ouverture est louable, mais il se déploie en laissant de côté ce qui constitue pourtant la base de la politique d'Open Data, telle qu'on la voit se développer en France depuis maintenant plusieurs années. Ce sont en effet d'abord les données administratives des établissements d'enseignement supérieur et de recherche qui devraient être diffusées en ligne et rendues librement réutilisables. Depuis 2016 et l'adoption de la [Loi République Numérique](#), c'est même devenu une obligation légale, puisque ce texte soumet les administrations publiques à un "principe d'ouverture par défaut" des informations qu'elles produisent dans le cadre de leurs activités.

En focalisant l'attention sur elles, les données de la recherche jouent un peu le rôle de l'arbre qui cache la forêt des données académiques et universitaires. Par "données administratives", on entend les données courantes de fonctionnement des établissements d'enseignement supérieur et de recherche : données du budget et des marchés publics ; statistiques des populations étudiantes ; listes des formations proposées ; données issues du bilan social relatives au personnel administratif ; informations liées à l'implantation des bâtiments et à la géolocalisation des services ; données documentaires liées à l'activités des bibliothèques, des centres de documentation et des presses universitaires ; informations sur les partenariats avec les entreprises et les actions internationales, etc.

Cette liste pourrait être encore longue, car tous les services d'une université – jusqu'au restaurant universitaire ! - produisent en réalité des données. Pour autant, et de manière assez surprenante, les projets d'Open Data embrassant les données administratives dans les universités se comptent sur les doigts de la main et les établissements qui se lancent dans cette démarche font figure d'exception.

## Un projet pionnier à l'Université Paris Nanterre

Avant même l'adoption de la loi République Numérique, l'université Paris Nanterre a initié un projet [d'Open Data administratif](#), lié au contexte local particulier auquel elle appartient. L'université est en effet implantée sur le territoire du département des Hauts-de-Seine, qui fait partie des collectivités locales ayant développé très tôt une politique volontariste en matière d'ouverture des données publiques. Les deux entités étaient de surcroît liées par une convention de partenariat comportant un volet numérique dans lequel fut inséré à partir de 2015 un point relatif à l'Open Data. Le département proposait de mettre à disposition son portail Open Data, ce qui permettrait de dispenser l'université d'avoir à mettre en place un site internet dédié. En retour, l'université s'engageait à mettre à disposition chaque année un certain nombre de jeux de données.

Pour être menée à bien, cette démarche a été conduite en mode projet, en rassemblant une équipe mêlant diverses compétences - personnel administratifs, chercheurs, bibliothécaires - et en s'appuyant sur la participation d'étudiants dans le cadre de stages. La tâche de ce groupe consistait à inventorier les jeux de données susceptibles d'être ouverts et à collaborer avec les services producteurs de manière à pouvoir les retravailler et les documenter en vue de leur diffusion. Un des aspects importants consistait également à obtenir la validation des instances politiques de l'université avant la publication des données.

Le portail Open Data des Hauts-de-Seine diffuse à ce jour [19 jeux de données](#) de l'université, placés sous la Licence Ouverte permettant leur libre réutilisation et

téléchargés chacun plusieurs centaines de fois. Dans la sélection des données, l'accent a été mis sur des informations relatives aux formations et diplômes délivrés à l'université, à l'activité des écoles doctorales, aux structures de recherche liées à l'établissement, ainsi que des statistiques fournies par le Service Commun de la Documentation.

Un tel projet a révélé également les difficultés et obstacles auxquels peut se heurter une démarche d'Open Data dans un contexte universitaire. Plus qu'une question technique, l'ouverture est avant tout un enjeu de gouvernance, car elle nécessite l'implication de nombreux services liés à toutes les composantes de l'université qui n'ont pas nécessairement l'habitude de travailler ensemble. Le coût en travail humain à fournir est également important, non seulement pour la collecte, mais aussi pour la mise à jour annuelle des données. Certains impératifs ne doivent pas être négligés, comme celui de la protection des données à caractère personnel et de la vie privée, qui nécessite une anonymisation des jeux, parfois complexe à obtenir.

Ces considérations plaideraient pour que les universités se concertent pour développer conjointement des politiques d'Open Data. A l'image de ce qui a été réalisé pour les collectivités territoriales dans le cadre de l'association [Open Data France](#), on pourrait imaginer par exemple que soit défini un "Socle commun des données universitaires" qui pourrait constituer la première marche à atteindre pour les établissements souhaitant se lancer dans des projets d'ouverture.

### FOCUS SUR DATA.HAUTS-DE-SEINE.FR



Les données sont ouvertes sur le site [opendata.hauts-de-seine.fr](http://opendata.hauts-de-seine.fr)





# Annuaire-education.fr, une expérience des données ouvertes du monde de l'éducation

par Julien Delmas, enseignant

Enseignant depuis une quinzaine d'années, développeur web à mes heures perdues, j'ai découvert récemment la quantité importante d'open data disponibles sur la plateforme du Ministère de l'éducation nationale [data.education.gouv.fr](http://data.education.gouv.fr).

## Agréger des données ouvertes

J'ai constaté que de nombreux sites utilisaient ces données, mais de façon très parcellaire, comme l'[annuaire des établissements](#), les [indicateurs de résultats des lycées](#) ou l'[organisation du temps scolaire](#) réalisés par le ministère de l'Éducation Nationale. La Caisse des Dépôts et Consignations propose quant à elle [eCarto "l'observatoire des territoires et du numérique éducatif"](#) et beaucoup d'entreprises privées s'intéressent aux résultats des examens, pour mieux classer les lycées.

J'ai voulu réaliser un annuaire de tous les établissements scolaires français : écoles, collèges, lycées, EREA, qu'ils soient publics, privés, sous et hors contrat. [Annuaire-education.fr](#) agrège toutes les informations existantes sur plus de 65 000 établissements scolaires : des coordonnées aux résultats aux examens, en passant par les programmes scolaires, les cartes des établissements d'une même ville, la proportion d'enseignants homme-femme ou l'équipement informatique.

l'Éducation Nationale qui en ont compris l'intérêt. Ce site n'est pas là pour proposer un comparatif ou un classement, mais uniquement des informations brutes que chacun interprétera comme il le souhaite.

Parmi les 10 000 visiteurs de la première semaine, j'ai eu beaucoup de retours de personnels de l'Éducation Nationale qui m'ont indiqué l'importance de toutes ces informations lors du choix d'un établissement dans lequel demander sa mutation. Ils en ignoraient jusqu'à ce jour l'existence.

## Côté développement

D'un point de vue plus technique, cet annuaire fonctionne avec le système de gestion de contenu franco-français, SPIP. Il n'interroge pas moins de quinze bases de données dupliquées depuis le site Open Data du ministère de l'Éducation Nationale. Cela représente près de deux millions de lignes et 400 Mo de données.



La vue développeur : image par Julien Delmas

J'ai été confronté à plusieurs difficultés dans le développement de cet outil. Certaines bases de données n'étaient pas réellement mises à jour, malgré des opérations de maintenance quotidiennes. Après signalement de ma part, le problème a été résolu. Le deuxième problème, et pas des moindres, est la médiocre qualité de structuration des bases de données avec des champs en double et des intitulés particulièrement obscurs.

## Valoriser des données ouvertes

Cet annuaire est certainement le site le plus complet sur les établissements scolaires français et l'ensemble de ses informations provient uniquement des données ouvertes du Ministère. Au vu des réactions suscitées, on peut espérer que cette modeste initiative personnelle permette de faire connaître l'existence de ces données publiques et qu'elle incite les établissements à actualiser leurs informations auprès du Ministère et à répondre plus régulièrement et plus précisément aux différentes enquêtes, pour fiabiliser encore plus ces données ouvertes.

### FOCUS SUR UNE ANALYSE

Ces Open Data permettent de mettre en avant certaines informations et sont très significatives de la dynamique d'un établissement scolaire. Quand on constate par exemple que certains collèges accueillent 30% d'élèves latinistes alors que d'autres, bien plus favorisés, ne convainquent que 5% d'élèves à choisir le latin, on peut s'interroger sur les raisons de ces différences. Il en est de même avec le pourcentage d'élèves de 3ème en retard scolaire qui est un indicateur objectif des difficultés réelles des élèves d'un établissement.

Le site a été lancé le 26 août 2019 et quelques semaines après son ouverture au public, nombreux sont les enseignants ou cadres de



Une page de l'annuaire de l'éducation



# Comment la mise en circulation des données pourrait induire la gouvernance du numérique de l'université ou de l'établissement ?

Par Bertrand Mocquet, expert numérique à l'Amue

Dans son ouvrage consacré à rendre compte du succès d'internet, Benjamin Loveluck nous affirme que *"la libre circulation de l'information (sous forme de code) [peut] devenir le principal socle de l'autonomie individuelle et collective"*. Il définit ainsi une nouvelle idéologie, celle du *"libéralisme informationnel"*, une sorte de *"reformulation de la théorie politique libérale fondée sur les flux d'informations au sein des réseaux informatiques"* (Loveluck, 2015). Il ajoute que cette modalité en réseau offre des nouvelles formes d'autorégulations sociales et économiques au sein de la société.

## Libre circulation de l'information

L'open-data s'inspire de cette finalité de circulation de l'information. Ouvertes, les données deviennent exploitables, exploitées et il y a une création de valeur autour des services s'appuyant sur ces données. Par exemple, l'application "Trouve ton campus", créée en 2014 par l'association StrasWeb formée d'étudiants et de jeunes diplômés, permettait jusqu'en 2017 aux étudiants de l'Université de Strasbourg de visualiser les moyens de transports dont ils disposent pour se rendre sur leur campus. C'est un bon exemple de *"mise en circulation des données [qui] renouvelle les modes de faire (...) en s'appuyant sur des acteurs tiers qui produisent des services à partir de ses données"* (Courmont, 2016), ici une association d'étudiants.



Figure 1 : Exemple d'écrans (application aujourd'hui fermée)

## Trois effets pour ce sujet stratégique

Pour nos gouvernances, au sens de *"gouvernance ou ensemble de dispositifs mis en œuvre par la firme pour réaliser des coordinations efficaces"* (Côme, 2013), à quel moment cela deviendrait-il un enjeu stratégique du numérique de l'université ou établissement ? Pour nous, il s'agit de prendre conscience que *"les données [ouvertes] sont devenues un enjeu de gouvernement à partir du moment où elles ont été mises en circulation"*. (Courmont, 2016). C'est donc la conséquence de l'ouverture des données que nous conseillons de saisir d'un point de vue gouvernance du numérique de l'université ou de l'établissement. Regardons-en quelques effets.

Un des premiers effets est de focaliser son attention stratégique sur la qualité des données de l'organisation et de percevoir leur richesse au sein de sa propre organisation : *"attachées à des producteurs, conservées dans des silos "métiers", les données sont des instruments au service d'une politique publique sectorielle"*. (Courmont, 2016). Les ouvrir, dans un premier temps en interne, permet de déssectoriser son activité interne, et de la penser de manière plus systémique et globale : on devient ainsi facilement le consommateur de ses propres données ouvertes.

Un des deuxièmes effets, est la *"capacité de contrôle de l'usage qui est fait de ses données"*. (Courmont, 2016), et une des conséquences, la mise en place d' *"une stratégie de régulation des flux de données. Par le choix des modalités de mise en circulation des données (licences, formats, etc.), elle encourage l'attachement de ses données à certains utilisateurs ou au contraire en limite d'autres"*. (Courmont, 2016)

Enfin, un des troisièmes effets serait de redonner du pouvoir aux citoyens, certaines personnes de plus en plus nombreuses souhaitant que *"de plus en plus d'actions et de comportements soient régulés par des flux*

*d'informations"* persuadés que *"si l'acteur public ne contrôle pas ces flux, il perd la maîtrise de ses politiques publiques"*. (Courmont, 2016)

Ces trois effets, mais la liste n'est pas exhaustive, de l'ouverture des données sur la gouvernance du numérique nous montrent qu'il est peut-être opportun aujourd'hui de bien percevoir les contours de l'open-data, de s'en saisir d'un point de vue stratégique plutôt que d'un point de vue législatif.

### POUR PLUS D'INFORMATIONS

Courmont, A. (2016). Politiques des données urbaines. Ce que l'open data fait au gouvernement urbain. *Sciences Po, Paris*.

Côme, T. (2013). La gouvernance des universités. *Gestion et management public, Volume 2/n°1(3), 15*.

Loveluck, B. (2015). *Réseaux, libertés et contrôle: Une généalogie politique d'internet*. Armand Colin.



# Expérimentation d'ouverture des données sur les parcours d'étude

Par Pierre Boudes, président du comité stratégique de l'Unif

Aux côtés de la Recherche, qui s'est saisie de l'ouverture des données pour servir à la fois une plus grande qualité (disponibilité des sources et reproductibilité) et une meilleure diffusion des connaissances (publication ouverte), l'enseignement supérieur ouvre aussi plus timidement ses données, par exemple sur l'offre de formation, les effectifs, les taux de réussite, les diplômes poursuivis, et plus rarement les débouchés. Pour mieux remplir nos missions, il est effectivement important de mieux connaître notre activité et nos publics et d'en partager largement cette connaissance. D'autant que l'ouverture des données favorise l'amélioration des processus internes et le dialogue entre les services.

## Un grand absent : le suivi des parcours d'étude réels

Pourtant, dans toutes les données disponibles librement, il manque un suivi des parcours d'études individuels qui rende compte de la succession des étapes réellement franchies, d'année d'étude en année d'étude, diplôme après diplôme et qui tienne compte de la mobilité inter-établissements, voire des interruptions et reprises d'études.

Ce suivi est simple à construire en s'appuyant sur le numéro d'identification national étudiant (INE). Il apporte des informations importantes lorsqu'il est mis en place à l'intérieur de l'établissement et il éclairerait nettement les relations inter-établissements s'il était disponible à ce niveau. On peut par exemple découvrir en interne, que sur les dix dernières années, 376 primos (arrivants l'année de leur bac) ont validé telle licence en 3 ans et combien ont choisi de poursuivre leur études dans la même université. En croisant nos données, on pourra aussi savoir combien ont poursuivi leurs études dans un autre établissement.

Les suivis de cohortes généralement réalisés par les observatoires de la vie étudiante, fournissent des indicateurs pour le pilotage. Systématiser et automatiser ces suivis, sans filtre statistique donnera de nouvelles informations sur le fonctionnement des formations et leurs relations. Il ne s'agit pas de se substituer au travail expert d'analyse à mener ensuite, mais de fournir des données

reproductibles et vérifiables, résultant directement des processus internes.

Pour l'orientation, un accès à de telles données via des applications adaptées permettrait de défricher le sujet en préparation d'un entretien individuel et fournirait une ossature de référence sur laquelle appuyer un conseil, qui doit rester personnalisé pour être pertinent.

## Faisons-le !

Le comité stratégique de l'UNIF propose à toutes les universités intéressées (franciliennes ou non) de croiser leurs données sur les inscriptions et la réussite des étudiants et étudiantes.

L'objectif est de réaliser rapidement une première publication sous licence ouverte, à titre expérimental, dans la perspective de pérenniser la démarche sous la forme d'une participation pérenne à l'infrastructure essentielle des données dont ont besoin les services numériques modernes.

Nous pensons en particulier à des services tiers consommant différentes sources de données libres d'accès. Par exemple, pour un projet de GPS de l'orientation proposant une vue à la première personne sur les continuités d'études, des débouchés réels et des mises en relation avec des personnes en poste, l'UNIF envisage de fournir à un partenaire les données nécessaires de façon non exclusive et les rendant libre d'accès, en open data.

Mieux connaître les mobilités entre établissements est également important pour produire des services numériques adaptés, notamment pour ce qui concerne la portabilité des données ou le choix plus radical du self data où chaque usager porte et contrôle ses données.

Les données que nous envisageons de publier prennent leur source dans les SI de scolarité (Apogée). Nous pouvons reconstruire le parcours de chaque individu (identifié par une empreinte cryptographique de son INE) à partir de ses inscriptions annuelles. En groupant les parcours identiques, nous pouvons les compter et oublier les individus. La spécialité du bac d'origine ou d'autres données peuvent faire partie du parcours. On

peut également décider d'oublier l'année d'inscription et de ne tenir compte que de la succession temporelle. On obtient ainsi des traces qui peuvent chevaucher plusieurs établissements.

Pour éviter de publier des informations ré-identifiantes, lorsqu'il y a trop peu de parcours identiques, on peut ne pas en donner la cardinalité exacte (anonymisation par généralisation). Par exemple, lorsque moins de 10 personnes ont effectué le même parcours, on pourra publier la trace sans en donner le nombre (ou en donnant un nombre d'une personne comme dans les traces publiées en open data par l'université Paris 13). Des données privées plus fines, incluant les cardinalités exactes, seront également communiquées aux établissements participants à l'expérimentation.

Pour participer vous pouvez nous contacter sur le mail : <mailto:opendata@unif.fr>.

## FOCUS SUR L'UNIF



UNIVERSITÉ NUMÉRIQUE ÎLE-DE-FRANCE

L'université numérique Île-de-France (anciennement université numérique Paris Île-de-France) est un acteur de la mutualisation sur le sujet du numérique, ayant pris la forme d'un service inter-universitaire actuellement porté par l'université Paris 13. Ses membres sont les établissements supérieurs publics de la région Île-de-France, dont une quinzaine d'universités et 4 Comues.



# L'Open Data ESR vous intéresse, vous souhaitez contribuer ?

Par Amue-DSPSI

Dans une logique de mutualisation, de partage d'expériences, de co-construction de solutions, la contribution des établissements est nécessaire pour l'Amue.

Vous pouvez contacter l'équipe numérique de l'Amue par mail à [numerique@amue.fr](mailto:numerique@amue.fr) ou participer à la journée Open Data du 14 novembre 2019.

[>S'inscrire](#)



Source : data.amue.fr

## Open data et ESR : Opportunité de créer de nouveaux services

Contactez-nous

Amue-DSPSI  
[numerique@amue.fr](mailto:numerique@amue.fr)



### Septembre 2019

Directeur général de la publication : Stéphane Athanase  
Rédacteurs en chef : Bertrand Mocquet et David Rongeat. Secrétaire de rédaction : Mire.  
ISSN 2650-8494

La collection numérique est sous Licence Creative Commons CC [BY-NC-SA 4.0](#)

Ont collaboré à ce numéro : (un grand merci)

Stéphane Athanase, Marin Dacos, Romain Tales, Emmanuel Weisenburger, Thierry Cheminade, Philippe Berger, Jérémie Valentin, Jean-Marie Bourgogne, Stéphane Papon, Pierre Boudes, Lionel Maurel et Julien Delmas.

Remerciements à la société coopérative [Datactivist](#)

Editeur : Amue – 103 boulevard Saint-Michel - 75005 Paris  
Fabriqué en France

Toutes les images et photos de ce numéro sont <Copyright> et libres de droit, droits réservés autorisation d'usage spécifique à cette publication.

### TOUS LES NUMEROS DE LA COLLECTION SONT EN TELECHARGEMENT

Amue. (2019). La collection numérique, à l'adresse  
<http://www.amue.fr/systeme-dinformation/metier/la-collection-numerique/>



### PROCHAIN NUMERO DE LA COLLECTION NUMERIQUE (NOVEMBRE 2019)

Et si nous regardions l'ESR par le prisme de la donnée ? : vos propositions de témoignage et retour d'expérience à [numerique@amue.fr](mailto:numerique@amue.fr)

