

ARCHIVES OUVERTES COMME DONNÉES OUVERTES ? L'EXEMPLE DE LA SUISSE

*Dr Jean-Luc Cochard
Archives fédérales suisses*

1 Introduction

Depuis le début de 2013, les Archives fédérales suisses (AFS) jouent un rôle moteur dans la mise en place d'un portail de données ouvertes publiques, communément appelées OGD (Open Government Data). La version pilote de ce portail est opérationnelle depuis le 14 septembre 2013 (opendata.admin.ch). Ce portail n'est cependant que la partie visible d'un projet dont l'objectif est la mise en place d'une stratégie nationale autour des données ouvertes publiques.

En quoi les AFS peuvent-elles trouver un intérêt pour leurs activités propres à participer si activement à un projet s'intéressant essentiellement aux données de base produites par les administrations publiques, et en particulier à des données très actuelles dont la durée de vie est réduite, comme des mesures de température ou les retards des trains du réseau ferroviaire suisse ?

Les raisons sont multiples, comme nous allons le présenter ci-dessous, mais avant cela on peut tirer plusieurs parallèles intéressants entre le fonctionnement d'un projet OGD et celui des AFS. Ainsi dans les deux cas, des informations produites par tous les services de l'administration sont collectées et indexées sur la base de métadonnées fournies aussi par ces services. OGD s'intéresse uniquement aux données numériques et les formats de ces données sont au centre des discussions. Deux aspects qu'on retrouve aussi dans les préoccupations des AFS et des archives en général. Et pour terminer cette comparaison, la publication en ligne de leurs données avec un outil de recherche efficace est une préoccupation de OGD et des AFS.

OGD est un projet complexe qui requiert la collaboration de nombreux acteurs et un contexte politique favorable. Investir dans OGD pour les AFS a un coût non négligeable comme le démontre la section qui suit.

2 Le projet Open Government Data suisse

Le projet OGD suisse a eu deux volets principaux qui se sont enchaînés. D'abord il s'est agi de mettre sur pied un portail pilote avec l'objectif d'expérimenter cette approche de collecte et de publication de données de base de différents services de l'administration suisse. Ensuite afin d'obtenir une décision politique permettant de travailler sur le moyen terme, une stratégie a été établie et soumise au Conseil fédéral.

2.1 La réalisation d'un portail pilote

Le projet de développement d'un portail pilote OGD a débuté en mars 2013 avec le regroupement autour des AFS de plusieurs offices de l'administration fédérale, à savoir l'Office fédéral de la statistique (OFS), Swisstopo, Météosuisse, la Bibliothèque nationale (BN) et l'Unité de pilotage de l'informatique de la Confédération (UPIC). Quelques mois plus tard, le Canton de Zürich a rejoint le projet et au début 2014, deux autres offices fédéraux, l'Office fédéral de l'environnement (OFEV) et l'Office fédéral de la santé publique (OFSP), ont aussi pris part au projet.

Pour la réalisation de ce prototype, le choix s'est porté sur l'environnement CKAN (www.ckan.org) qui est très répandu en Europe dans la réalisation de portails de données ouvertes publiques des pays européens.

L'éventail des administrations associées du projet était un garant de la diversité des données, des environnements techniques et juridiques à traiter durant cette phase pilote. Deux aspects majeurs ont ainsi été particulièrement étudiés :

1. La collecte des informations de référencement (métadonnées) et la gestion de données primaires très variées telles que des statistiques, du contenu de bases de données bibliographiques, des mesures de température en temps réel ou des cartes géographiques.
2. La mise en place d'une solution multilingue – allemand, français, italien et anglais, tant au niveau de l'interface utilisateur qu'à niveau du contenu des métadonnées qui devaient être rédigées dans ces différentes langues.

La collecte des métadonnées (harvesting) s'appuie sur un processus automatique journalier qui va lire les contenu de fichiers de données déposés sur des serveurs accessibles via internet avec des URL prédéfinis (un fichier par office – fournisseur de données). Chaque fichier contient les métadonnées des jeux de données mis en libre accès ainsi que la référence aux données primaires (URLs permanentes). Dès le départ, il était convenu que chaque fournisseur de données était responsable de ses données primaires, qu'il s'agisse de fichiers de type tableaux de données ou qu'il s'agisse de services web. Le portail `opendata.admin.ch` n'a comme objectif que de référencer les jeux de données et d'indiquer où se trouvent les données, sans en prendre en charge la gestion. Cet aspect est contradictoire avec le fonctionnement usuel des archives à qui l'on verse des dossiers analogiques ou numériques et qui de ce fait acquièrent la responsabilité de les préserver et de les rendre publics sous certaines conditions. Mais dans le cas des données OGD, le fournisseur de données les met à disposition du public tout en continuant à les utiliser pour ses besoins propres.

La Figure 1 1 décrit schématiquement l'organisation d'un portail OGD typique avec la séparation des responsabilités entre OGD qui est en charge de collecter et publier des informations sur des jeux de données et le fournisseur de donnée qui établit son catalogue et qui gère ses données primaires.

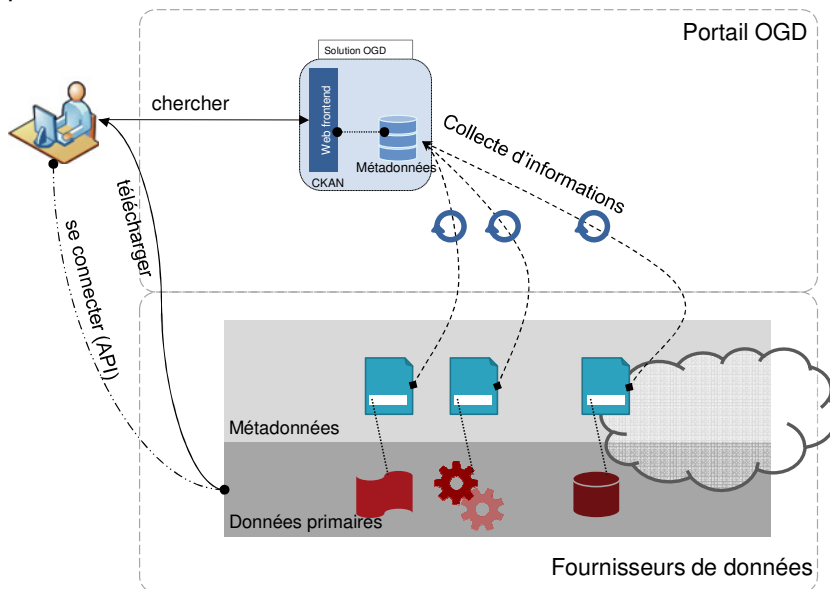


Figure 1 : Schéma général d'un portail OGD avec la séparation entre domaines de responsabilité des acteurs du projet.

Pour les utilisateurs des données ouvertes publiques, un aspect crucial est le format des données qui doit être tel qu'il en facilite l'utilisation informatique. Ainsi un rapport financier en PDF n'est pas idéal comme donnée primaire si l'on souhaite visualiser l'évolution des dépenses d'une administration sur plusieurs années. L'utilisateur préférerait une copie des tableaux comptables qui sont reproduits dans le rapport financier, au détriment des explications, des graphiques et des analyses.

Un travail a donc été réalisé par les différents partenaires du projet pour identifier des sources de données qui étaient le plus possible compatibles avec les dix principes de publication des données ouvertes publiques (The Sunlight Foundation, 2010) et pour tester leur acceptation auprès des usagers du portail.

Des données géographiques sous forme de cartes numériques mais aussi sous forme de services web (WFS – Web Feature Service et WMS – Web Map Service) constituent les données les plus consultées. Des fichiers de données en temps différé de MétéoSuisse avec une mise à jour toutes les 15 minutes constituent une autre approche. Le contenu de plusieurs bases de données est aussi mis à disposition selon des modèles différents. La BN exporte chaque jour des fichiers au format MARCXML (www.loc.gov/standards/marcxml) avec le contenu de plusieurs de ses bases de données. Les AFS ont choisi une autre option qui consiste à créer une copie de son catalogue comme base de données MySQL. Quant aux autres offices, ils livrent en majorité des tableaux de données au format ODS ou CSV.

La variété des formats est grande mais la tendance est de retenir des formats non propriétaire et explicites. On notera qu'une telle tendance est tout à l'avantage des archives qui doivent garantir la pérennité de ces données, une fois livrées, sur le très long terme.

Le portail pilote est donc opérationnel depuis septembre 2013 et est régulièrement alimenté par de nouvelles données. Il va être maintenu tel quel en exploitation jusqu'à ce qu'il soit remplacé dans le courant de l'année 2015 par son successeur qui sera le portail définitif des administrations publiques suisses.

2.2 L'établissement d'une stratégie nationale

La mise en place d'un portail pilote a certes joué un rôle dans la décision prise par le gouvernement suisse d'adopter la Stratégie OGD pour la Suisse le 16 avril 2014 pour la période 2014-2018 mais le portail n'est qu'une des treize mesures qui ont été élaborées afin de garantir une adoption aussi large que possible de cette stratégie (CF, 2014).

Bien que le projet soit fédéral, son succès se mesurera entre autre à l'adhésion non seulement des offices de l'administration fédérale mais aussi et surtout des cantons et des communes, qui n'ont aucune obligation d'y participer. C'est la raison pour laquelle plusieurs mesures vont dans ce sens comme *l'adaptation du cadre légal*, la mise en place d'un *processus de libération des données*, le développement de *moyens auxiliaires et instruments pour la publication des données*, l'adoption de *normes pour les OGD* et *l'élaboration d'un modèle de coopération relatif aux OGD*, afin d'inclure les cantons et les communes dans cette dynamique.

Un autre critère majeur du succès d'une telle stratégie réside dans l'utilisation qui en sera faite des données ainsi disponibles, à la fois par les acteurs externes à l'administration que sont les développeurs d'applications, les spécialistes en visualisation de données et les institutions de recherche et développement, mais aussi par les administrations elles-mêmes qui peuvent améliorer leur efficacité en utilisant des données produites en dehors de leurs propres champs d'activités. Pour ce faire, il est nécessaire de disposer de *conditions d'utilisation uniformes* afin de simplifier la réutilisation des données. Il faut aussi créer une culture du libre accès aux données en établissant un *dialogue avec les utilisateurs des OGD* pour en comprendre les besoins et les attentes, et en *présentant l'offre de données au public*, car la manipulation de ces

données n'est ni immédiate, ni forcément simple. Dans certains cas, des compétences techniques approfondies sont même requises. L'accès facilité aux données ne signifie pas forcément que les données faciles à interpréter.

En mettant finalement sur pied un portail efficace et fiable ainsi qu'un *plan de libération des données*, on se donne les moyens de transformer une initiative soutenue par quelques responsables au sein de l'administration fédérale en une institution stable, reconnue et indiscutablement utile.

2.3 Les sujets en cours

La mise en œuvre des différentes mesures mentionnées plus haut va se faire progressivement et sur la durée des quatre prochaines années. Certains éléments sensibles doivent cependant trouver rapidement des réponses afin de constituer un contexte clair et rassurant pour les fournisseurs de données actuels et futurs.

2.3.1 Les conditions d'utilisation

Un élément clé des principes de publication des OGD est de connaître les conditions d'utilisation des données référencées sur le portail. Le monde anglo-saxon utilise la notion de licence pour définir et décrire ces notions et c'est sous l'égide de l'association Creative Commons (CC) (creativecommons.org) qu'une série de licences CC qu'on retrouve très régulièrement dans les portails OGD de nos voisins, a vu le jour.

En Suisse, le terme même de « licence » n'a pas d'existence juridique dans le droit public. Dans ce contexte reprendre les licences CC en Suisse pose problème. Sur le fond cependant, il est possible de reprendre les principes même des licences CC pour les inscrire dans les notions du droit suisse. C'est une des tâches en cours qui devrait aboutir dans les prochains mois.

2.3.2 La standardisation des métadonnées

Les données référencées par un portail OGD en général et par le portail pilote actuel proviennent de sources très diverses et sont en lien avec une très grande diversité de domaines d'activités. Le défi d'un portail OGD est de fournir un outil de recherche efficace à travers tous les jeux de données répertoriés. Ainsi donc les métadonnées qui décrivent chaque jeu de données sont au centre de ces préoccupations.

Comment décrire précisément chaque jeu de données, en gardant les spécificités de chaque domaine, sans toutefois créer une telle surcharge de travail qui démotiverait les administrations à ne pas publier leurs données sur un portail OGD ?

Une piste possible vient de la Commission européenne qui a identifié ce problème et propose des normes ayant pour but de consolider le contenu des différents portails OGD des membres de l'Union européenne au sein d'un portail européen supranational. L'utilisation de vocabulaires contrôlés tels que le Data Catalog Vocabulary (DCAT – www.w3.org/TR/vocab-dcat) et Dublin Core (dublincore.org/documents/dcmi-terms), qui s'appuient sur le modèle de données RDF (en.wikipedia.org/wiki/Resource_Description_Framework) constitue une initiative très intéressante pour garantir une cohésion des métadonnées produites par différents secteurs d'activités et dans différents pays.

Cette solution est certainement amenée à évoluer mais le fait qu'elle s'appuie sur le modèle RDF, offre une flexibilité et une puissance expressive qui pourrait s'avérer centrale à son succès.

2.3.3 Les conséquences financières

En parallèle au développement du portail pilote, une étude sur le potentiel économique de OGD pour la Suisse a été réalisée (Bürgi-Schmelz, 2013). Cette étude arrive à la conclusion que la réutilisation par les milieux économique de OGD pourrait engendrer une création de richesse d'environ 1 milliard de francs suisses par an. Ce montant est relativement prudent si on le met en relation avec d'autres études européennes citées dans le rapport.

Une question posée à l'auteure de cette étude était de déterminer si en renonçant à prélever des taxes sur certaines données publiques dans la perspective de les publier sur le portail OGD, cette perte fiscale aurait un impact sur les finances publiques. La conclusion est que le renoncement à ces taxes va être compensé à terme par un accroissement de la recette fiscale en provenance des entreprises.

Des mécanismes de compensation doivent cependant être envisagés car même si les prévisions économiques se réalisent, il va forcément y avoir une période de latence durant laquelle les effets de OGD sur l'économie ne seront pas visibles alors que la perte de revenu sera elle effective pour certaines administrations.

3 L'évaluation du portail pilote

Le portail pilote est maintenant opérationnel depuis plus d'un an, sans défaillance majeure. La fréquentation est stable en dehors de quelques événements qui stimulent sa fréquentation, essentiellement lors de la publication de nouvelles données (voir Figure 2).

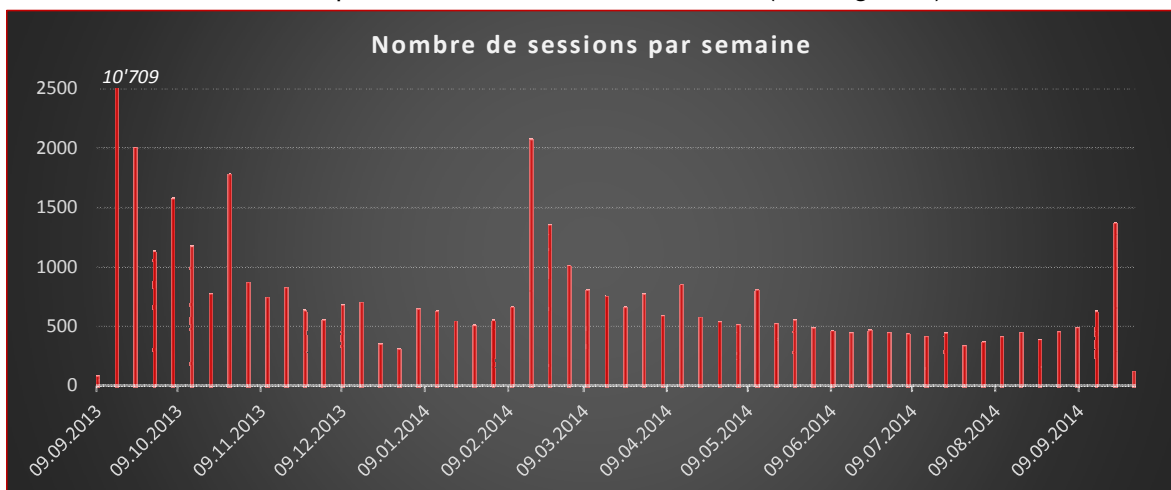


Figure 2 : Statistique hebdomadaire d'utilisation du portail se septembre 2013 à septembre 2014. Le lancement du portail a généré une telle attention que le nombre de visiteurs (10'709) dépasse largement l'échelle retenue pour représenter l'évolution des visites sur le reste de l'année.

Un rapport est un cours d'élaboration qui rassemble de nombreux enseignements sur l'expérience du projet pilote, tant selon la perspective des utilisateurs que selon celle des fournisseurs de données. Un résultat parmi d'autres démontre la diversité des attentes des usagers concernant OGD (voir Figure 3).

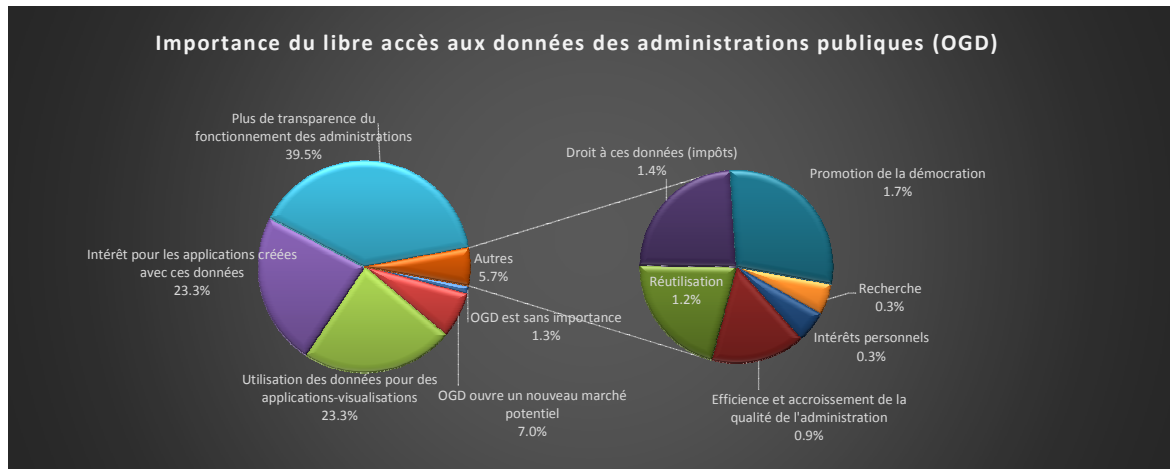


Figure 3 : Eventail des raisons qui motivent les usagers à s'intéresser à OGD.

4 Les perspectives pour les AFS

4.1 Les AFS comme fournisseur de données

Une option immédiate pour les AFS est de considérer OGD comme un canal supplémentaire de publication de ses données. Cette voie a déjà été exploitée en publiant plusieurs jeux de données. Elle le sera certainement encore à l'avenir pour y publier d'autres données primaires telles que l'historique des comptes de la Confédération. Nous pensons que OGD offre un canal idéal pour valoriser des données brutes en s'adressant à un public qui dépasse celui des chercheurs et des journalistes.

En publiant des données OGD, on peut dans certains cas faire l'économie de développer un environnement de valorisation de ses propres données sous forme de site internet par exemple, en permettant à d'autres de le faire.

OGD peut aussi servir à publier des documents dans des formats non optimaux et susciter la collaboration des usagers pour produire des versions plus utilisables des données qu'ils contiennent.

Le format même des données référencées sur OGD peut être un sujet de recherche. Le web des données, aussi appelé web sémantique ou Linked Data, offre des perspectives très intéressantes pour publier des données qui seront reprises par d'autres et combinées avec d'autres sources, comme l'illustre très bien (van Hooland & Verborgh, 2014). OGD est encore une fois un excellent canal pour déterminer si une approche a plus ou moins de potentiel qu'une autre en publiant des données dans différents formats et en analysant les utilisations qui en sont faites.

4.2 Les AFS comme exploitant du portail national

La gestion du projet OGD et l'exploitation de son portail sont officiellement sous la responsabilité des AFS à partir de 2015. En quoi une telle décision est-elle cohérente avec les tâches usuelles des AFS ?

On peut d'abord y voir le fait que les AFS sont des spécialistes des catalogues de données et le portail OGD n'est autre qu'un catalogue qui référence des sources de données dispersées au sein de l'administration fédérale. Les AFS ont un rôle transversal dans l'administration. Elles sont en contact avec l'ensemble des offices qui y livrent régulièrement des dossiers d'archives.

Le projet OGD a aussi l'ambition de s'adresser à l'ensemble des offices fédéraux et même au-delà, à savoir aux cantons et aux communes.

La mise en place d'une solution d'archivage de documents électroniques ainsi qu'une plateforme sécurisée de livraison de ces a permis de créer une prestation d'archivage pour des tiers. Les cantons et les communes sont les clients potentiels de cette nouvelle prestation. Qui dit offrir une prestation, dit acquérir des clients, donc convaincre ces clients. Que la prestation soit gratuite comme dans le cas de OGD ou payante comme pour l'archivage numérique, le principe est assez similaire et les AFS travaillent en ce sens.

4.3 Les données OGD comme futures archives

Une dernière raison à plus long terme est de considérer que les données OGD encapsulées dans des jeux de données avec des métadonnées riches ont toutes les chances de devenir un jour des archives sous la responsabilité des AFS. En étant aujourd'hui directement impliqué dans le processus de sélection, de publication de ces données et de contrôle de la qualité des métadonnées qui leur sont associées, on se positionne idéalement pour disposer à l'avenir de données d'archives beaucoup plus riches et mieux documentées que si elles sont publiées individuellement par chaque office de l'administration.

OGD est donc une chance pour une institution comme les AFS qui peuvent en tirer un énorme profit aussi bien à court terme qu'à long terme, sans compter que cette offre donne une visibilité aux AFS auprès d'un public peu coutumier des salles de lecture.

5 Bibliographie

- Bürgi-Schmelz, A. (2013). *Wirtschaftliche Auswirkungen von Open Government Data*. Consulté le 30.9.2014, sur http://www.bar.admin.ch/themen/01648/01651/index.html?lang=de&download=NHZLpZeg7t,Inp6I0NTU042I2Z6ln1acy4Zn4Z2qZpnO2Yuq2Z6gpJCDeIR4e2ym162epYbg2c_JiKbNoKSn6A--
- CF. (2014). Stratégie en matière de libre accès aux données publiques en Suisse pour les années 2014 à 2018. *La Feuille fédérale*, 2014(18). Récupéré sur <http://www.admin.ch/opc/fr/federal-gazette/2014/3347.pdf>
- The Sunlight Foundation. (2010). *Ten Principles for Opening up Government Information*. Consulté le 30.9.2014, sur <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>.
- van Hooland, S., & Verborgh, R. (2014). *Linked Data for Libraries, Archives and Museums*. Facet Publishing.