

# Concertation Portails de données ouvertes

Date : 23 juin 2014, 14:00 – 16:00

Lieu : Vlaanderen BZ, Bâtiment Baudouin, Bruxelles

Présents :

- AWT
- ASA
- Fedict
- SPF Économie (BCE + Direction générale Statistique)
- Ville d'Anvers
- Ville de Bruxelles
- Vlaanderen BZ
- ProXML (expert invité)
- TenForce (expert invité)

Excusés :

- Etnic
- Ville de Gand
- Ville de Courtrai
- Ville de Namur
- Intercommunale Leiedal
- WestToer

## Discussion

### Ouverture de la BCE

Depuis quelques mois, la Banque Carrefour des Entreprises (BCE) du SPF Économie est disponible sous 2 formes : en tant que fichiers à télécharger gratuitement et en tant que service web payant (50 euros pour 2000 requêtes en ligne).

L'enregistrement est gratuit, mais obligatoire. En juin, il y avait une cinquantaine d'utilisateurs des fichiers gratuits, et 1 utilisateur du service web payant.

- <http://economie.fgov.be/fr/entreprises/bce/pub/opendata/#.U7pJEbGAzQo>
- <http://economie.fgov.be/fr/entreprises/bce/pub/PuSSW/#.U7pJHrGAzQo>

ProXML a créé un site web afin de proposer les données de la BCE en tant que Linked Open Data.

- <http://kbodata.be>

À l'étranger, il existe des initiatives similaires, comme le projet OpenCorporates qui contient les données d'entreprises de 50 pays.

- <http://opencorporates.com>

### Licences

La Commission européenne a organisé un tour de table afin de déterminer quelles licences *open data* sont utilisées dans les États membres. Sur ce point, la Belgique est unique car de nombreuses licences différentes sont utilisées :

- la ville de Bruxelles utilise une licence d'Etalab (un service créé par le Premier Ministre français) ;
- les licences *open data* flamandes sont dérivées des licences Creative Commons ;
- la ville d'Anvers utilise désormais une licence *open data* flamande ;
- la ville de Gand utilise une licence propre ;
- la BCE utilise également une licence propre ;
- l'AWT est intéressée par les licences modèles développées par l'ASA.

Cela complique la tâche des développeurs lorsqu'ils doivent combiner des données provenant de différentes sources. Une licence européenne pourrait être une solution, bien que cela ne soit peut-être pas possible à cause de différences subtiles dans la législation sur les copyrights.

## Relation entre DCAT, DCAT-AP et la variante belge

Le Data Catalog Vocabulary (DCAT) est une spécification développée par l'Université nationale d'Irlande et qui a ensuite été transférée au W3C. Cette spécification décrit un format de métadonnées pour des ensembles de données ouvertes, basé sur les premiers portails *open data* (data.gov.uk et us).

DCAT est cependant trop général et, en 2013, la Commission européenne a créé un groupe de travail afin de développer un profil d'application : DCAT-AP. Cette variante est compatible avec DCAT, mais contient des directives indiquant quelles données sont obligatoires et quelles données sont optionnelles.

Ce profil peut encore être développé. Il s'agit alors de la variante belge de DCAT-AP, dans laquelle, par exemple, une plus grande attention est portée aux classifications à utiliser.

Remarque : dans le contexte des données ouvertes européennes, les nouvelles initiatives requièrent le support pour DCAT-AP (par exemple, voir la récente adjudication publique du nouveau portail de données ouvertes suédois).

## Portail européens

Il existe actuellement plusieurs portails de données ouvertes à l'échelle européenne :

- Le Portail des données ouvertes de l'Union européenne (open-data.europa.eu), qui contient tous les ensembles de données mis à disposition par toutes les institutions de l'Union européenne. Le développeur était TenForce, en collaboration avec l'OKFN.
- Une plateforme CKAN (publicdata.eu), développée par l'Open Knowledge Foundation (OKFN) dans le cadre du projet LOD2 (lod2.eu). Elle contient des ensembles de données des pouvoirs publics (États, régions, villes). Les données sont collectées (*harvested*) en interrogeant les portails de données ouvertes de ces pouvoirs publics. Les systèmes sources doivent être des systèmes CKAN.
- Une Open Data Interoperability Platform (ODIP), développée par TenForce et PwC dans le cadre du projet Open Data Support (data.opendatasupport.eu). Elle contient des ensembles de données des portails de données ouvertes nationaux, ainsi que de l'Union européenne. L'hypothèse est que chaque État membre de l'Union européenne ait collecté au niveau national toutes les données des niveaux régional et communal.

Le système ODIP peut collecter des informations d'ensembles de données d'autres portails et les transformer, via SPARQL, pour DCAT-AP. Actuellement, il s'agit principalement de portails CKAN, mais les *dumps* RDF et *dumps* CSV sont également supportés. L'importation peut être étendue à n'importe quel type de format d'entrée.

La mise en place d'un tel processus de transformation représente environ une semaine de travail. Ensuite, les informations de base (métadonnées) d'un portail peuvent être automatiquement collectées et traitées. On peut alors envisager d'améliorer et d'affiner la transformation.

Les ensembles de données eux-mêmes ne sont par ailleurs pas copiés. Ils restent simplement là où ils étaient. La plateforme est une extension (*open source*) de l'outil *open source* LODMS (aujourd'hui Unified Views).

Il est possible que les deux plateformes européennes fusionnent pour former un ensemble (à cet effet, un cahier des charges sera probablement publié), dans lequel la partie CKAN servira de *front-end* (= site web pour les utilisateurs finaux) et la partie ODIP de *back-end* (qui continuera à collecter les métadonnées des portails des États membres et à les transformer au format DCAT-AP).

## Portails belges, export vers DCAT-AP

L'objectif n'est pas d'obliger tous les portails à utiliser le même logiciel, mais bien d'utiliser le même format pour l'échange de métadonnées : la variante belge de DCAT-AP.

Pour atteindre cet objectif, tous les portails et sites devraient donc fournir un fichier DCAT-AP, afin qu'il puisse, par exemple, être traité par le système européen ODIP (ou par n'importe quel autre système capable de traiter le format DCAT).

État de la situation :

- Bruxelles(-Ville) a récemment lancé un nouveau site avec OpenDataSoft. Ce système génère déjà un export complet de métadonnées vers DCAT et est assez facilement transformable au format DCAT-AP.

- Vlaanderen BZ/Corve et l'AWT utilisent CKAN, mais la structure CKAN flamande et wallonne diffère. En outre, aucun de ces deux portails n'exporte au format DCAT car CKAN privilégie les API propres. Certains portails étrangers n'ont développé qu'un seul module DCAT, mais ceux-ci sont spécifiques au portail et/ou ne fonctionnent pas sur la dernière version de CKAN. Corve et l'AWT devraient pouvoir collaborer afin de (faire) développer un module DCAT-AP.

- Gand et Anvers utilisent une ancienne version de The Data Tank avec un *front-end* Drupal. En outre, une partie des métadonnées se trouve dans TDT et une autre partie (principalement la description) dans Drupal. Actuellement, aucun export au format DCAT-AP n'est prévu. D'ici peu, la ville d'Anvers utilisera bien une nouvelle version de TDT, mais il faut encore attendre de voir si cela change quelque chose à la situation.

- L'ASA ([publicdata.belgium.be](http://publicdata.belgium.be)) utilise Drupal et Sharepoint et aucun export vers DCAT-AP n'est actuellement prévu.

- Fedict ([data.gov.be](http://data.gov.be)) utilise Drupal 6 et aucun export vers DCAT-AP n'est actuellement prévu.

- Le SPF Économie (statbel) est un site Tridion et aucun export vers DCAT-AP n'est actuellement prévu. La Direction Statistique du SPF Économie envoie également des données vers EuroStat, mais qui ne sont pas nécessairement les mêmes que celles qui sont (également) publiées sur [statbel.fgov.be](http://statbel.fgov.be).

Une solution intermédiaire peut être de développer des outils de *scraping* (Fedict travaille sur un exemple) qui transforment au format DCAT-AP des données de portails de données ouvertes.

## Hackathons

La ville de Courtrai a organisé pour la deuxième fois un hackathon (« AppLab 2 »).

- <http://www.kortrijk.be/opendata/evenementen/applab2>

La ville de Bruxelles organise un hackathon les 17 et 18 octobre.

- <http://checkthis.com/hackathonbxl>
- <http://agenda.brussels.be/4458>

## Points d'action

- (Tous) Examiner le *mapping* des thèmes/mots-clés des propres portails vers les termes d'Eurovoc.
- (Tous) Rechercher des personnes de contact des parlements fédéral et régionaux afin d'examiner s'il existe un intérêt pour la réutilisation du système *open data* du Parlement flamand.
- (Tous) Examiner ce qu'il faut faire précisément pour exporter les métadonnées des propres portails vers une version belge de DCAT-AP (y a-t-il suffisamment d'expertise, de budget, etc.?).