

Concertation Portails des données ouvertes

Date : 5 mai 2014, 13h00 – 15:00

Lieu : Fedict, WTC III, Bruxelles

Présents :

- AWT
- ASA
- Etnic (matinée)
- Fedict
- Ville de Bruxelles
- Vlaanderen BZ
- ProXML (expert invité)

Excusés :

- Ville d'Anvers
- Ville de Gand
- Ville de Namur
- WestToer

Discussion

Qualité des ensembles de données/feedback

Il n'est pas toujours aisé de contrôler la qualité d'un ensemble de données. La qualité en elle-même n'est pas non plus facile à définir (s'agit-il d'erreurs de mesure, de la fréquence des mises à jour, ...). Il est donc préférable de parler d'« utilisabilité » pour les citoyens et les entreprises. Il est dès lors important que les citoyens et entreprises (ainsi que d'autres services publics) puissent donner leur feedback concernant les ensembles de données, quel que soit le portail sur lequel l'ensemble de données est publié.

L'objectif est donc de mentionner, pour chaque ensemble de données, les coordonnées (au moins l'adresse e-mail) du service capable de répondre aux questions relatives à l'ensemble de données.

Statut des portails européens

À la Commission européenne, il existe actuellement deux projets visant à créer un portail central dans lequel pourraient être enregistrées les métadonnées de tous les autres portails des données ouvertes en Europe. Il s'agirait d'un portail basé sur le package CKAN d'OKFN et d'un autre portail basé sur un autre logiciel de PWC.

On ne sait pas clairement lequel de ces portails demeurera (c'est-à-dire, celui qui recevra le budget nécessaire), mais il est possible que l'Europe demande, à un certain moment, que les autres portails (nationaux, régionaux et locaux) transmettent leurs métadonnées vers l'un de ces portails.

CKAN et OKFN

De nombreux portails étrangers utilisent le *package open source* CKAN de l'Open Knowledge Foundation (OKFN). En Belgique, il y a plus de variété :

- La Flandre utilise CKAN.
- Gand et Anvers utilisent The Data Tank avec un *front-end* Drupal.
- Bruxelles(-Ville) a récemment lancé un nouveau site avec OpenDataSoft.
- Fedict (data.gov.be) utilise Drupal 6.
- L'ASA (publicdata.belgium.be) utilise Drupal et Sharepoint.

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.

Format standard de métadonnées (variante belge de DCAT)

DCAT est un format de métadonnées international qui est apparu sur les premiers portails (entre autres aux États-Unis et au Royaume-Uni) et qui continue désormais à être développé par le W3C.

DCAT n'est pas suffisamment spécifique pour nos objectifs, car la plupart des champs sont facultatifs et parce qu'il n'existe aucune recommandation quant aux catégorisations à utiliser. Un groupe de travail européen a affiné ce format pour en faire le DCAT-AP. L'objectif est de continuer à le développer afin d'arriver à une variante belge. À cette fin, la Flandre a déjà ébauché une amorce dans le manuel Open Data.

L'objectif est que les portails belges soutiennent cette variante belge et puissent exporter des métadonnées vers ce format. Le but n'est cependant pas de copier les ensembles de données eux-mêmes. Il est en effet préférable de laisser les données où elles se trouvent, et de simplement prévoir un lien de retour.

Échange de métadonnées

Il est probable qu'un système de ce type voie le jour permettant aux portails locaux de transférer des métadonnées aux portails régionaux, aux portails régionaux de transférer des métadonnées aux portails nationaux et aux portails nationaux de transférer des métadonnées aux portails européens. Les mouvements contraires doivent d'ailleurs également être possibles.

En outre, il n'est pas impensable qu'une tierce partie commence à développer elle-même des portails, mais selon une répartition en thèmes/secteurs.

Traductions/plurilinguisme

Le plurilinguisme est un point auquel il convient de prêter attention. En effet, un certain nombre d'autorités veulent ou doivent mettre des données à disposition dans plusieurs langues.

De nombreux logiciels de portail (tels que CKAN) permettent de travailler avec plusieurs langues mais ne permettent pas d'indiquer qu'une description précise est une traduction d'une autre description, même si l'ensemble de données sous-jacent est le même. En d'autres termes, on constate l'existence de doublons. Cela n'est pas un souci mais peut causer des problèmes en cas de remaniement des traductions.

Le but n'est pas de toujours traduire toutes les métadonnées (descriptions, titres ...) en 2, 3 voire 4 langues (les trois langues nationales et l'anglais – cela dépend notamment de la législation sur l'emploi des langues et de la disponibilité/du caractère payable des services de traduction).

L'objectif est bien d'indiquer explicitement la langue de rédaction d'un titre, d'une description, etc., afin de faciliter le traitement automatisé des métadonnées. Cet objectif est d'ailleurs prévu dans le DCAT.

Mapping vers les thésaurus existants

Quasi tous les portails utilisent une répartition propre en thèmes, mots-clés, etc. Ce système est très flexible mais rend complexe la recherche d'ensembles de données similaires dans d'autres portails.

L'objectif est dès lors de prévoir un *mapping* supplémentaire des listes propres vers des thésaurus existants, utilisés à l'échelle internationale. La répartition propre, qui est souvent un peu plus affinée, peut dès lors être conservée, ce qui permet en outre de simplifier le traitement automatique des métadonnées.

Pour indiquer un thème/une catégorie, on peut par exemple se référer à Eurovoc. Eurovoc est un thésaurus géré et utilisé par l'Office des publications de l'Union européenne qui a été traduit dans toutes les langues officielles de l'UE. Récemment, ce thésaurus a été publié par le biais d'une licence ouverte et peut donc être téléchargé gratuitement.

Pour indiquer de quelle région géographique il s'agit, la classification européenne NUTS peut éventuellement être utilisée. Le désavantage est que NUTS se limite au niveau de l'arrondissement et ne s'étend par exemple pas jusqu'au niveau des communes ou des rues. Geonames pourrait être une solution.

Pour le *mapping*, on utilisera de préférence le format SKOS (spécification W3C) mais, au besoin, on peut utiliser un tableur, permettant d'indiquer si un terme du portail est identique (skos:exactMatch) ou plus spécifique (skos:narrowMatch) qu'un terme d'un thésaurus connu.

Manuel Open Data

La Flandre a publié une seconde version du manuel Open Data, qui n'est provisoirement que disponible en néerlandais. Les suggestions (et traductions) sont toujours les bienvenues.

Projet Open Data du Parlement flamand

La Flandre subventionne quelques projets Open Data (« VIP »), notamment un projet visant à renouveler le site web du Parlement flamand. Dans ce cadre, les données seront aussi publiées comme données ouvertes, notamment la liste des députés.

Il serait intéressant que les autres parlements belges puissent également réaliser un projet similaire, en réutilisant éventuellement les connaissances et/ou le logiciel du Parlement flamand. Aux niveaux régional et européen, de nombreuses initiatives privées désirant utiliser ce type de données ont en effet déjà vu le jour.

Quelques liens

Projets

EU Open Data Support (formation gratuite en ligne etc.)

- <https://joinup.ec.europa.eu/community/ods/description>

Projet EU Linked Open Data (outils)

- <http://lod2.eu/Welcome.html>

EU Location Framework

- http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-13action_en.htm

(à ne pas confondre avec le European Location Framework <http://www.elfproject.eu/>)

Projets flamands Open Data « VIP »

- <http://www.corve.be/projecten/corve-vip/>

Portails

Portail des données ouvertes de l'Union européenne (uniquement des institutions européennes)

- <http://open-data.europa.eu/fr/data/>

EU Open Data Portal, basé sur CKAN (réalisé par l'OKFN, bêta)

- <http://publicdata.eu/>

EU Open Data Portal ODIP (réalisé par PriceWaterhouseCoopers)

- <http://data.opendatasupport.eu/>

Spécifications et thésaurus

Thésaurus Eurovoc

- <http://eurovoc.europa.eu/>

Classification NUTS UE (Eurostat)

- http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction

Manuel flamand Open Data 2.0 (et plus spécifiquement partie 8.1 Métadonnées, p 36 - 39)

- http://opendataforum.info/files/Open_Data_Handboek.pdf

Points d'action

- (Tous) Examiner le *mapping* des thèmes/mots-clés des propres portails vers les termes d'Eurovoc
- (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.?)
- (Fedict) Envoyer un Doodle pour fixer la date de la prochaine réunion ; inviter également la ville de Courtrai et le SPF Économie
- (Flandre) Envoyer à l'ASA les informations de contact concernant le projet VIP du Parlement flamand