

Open Data Portals overleg

Datum : 5 mei 2014, 13:00 - 15:00

Locatie: Fedict, WTC 3, Brussel

Aanwezig:

- AWT
- DAV
- Ethic (voormiddag)
- Fedict
- Stad Brussel
- Vlaanderen BZ
- ProXML (invited expert)

Verontschuldigd:

- Stad Antwerpen
- Stad Gent
- Stad Namur
- WestToer

Bespreking

Kwaliteit van datasets / feedback

Het is niet altijd eenvoudig om de kwaliteit van een dataset te controleren. Ook kwaliteit is niet eenvoudig te definiëren (gaat het over meetfouten, frequentie van updates, ...), beter is om te spreken van “bruikbaarheid” voor burgers en bedrijven. Het is dan ook belangrijk dat burgers en bedrijven (en andere overheidsdiensten) feedback kunnen geven betreffende de datasets, ongeacht op welk portaal de dataset gepubliceerd wordt.

Het is dan ook de bedoeling om bij elke dataset de contactgegevens (minstens een e-mailadres) te vermelden van de dienst die vragen over de dataset kan beantwoorden.

Status van de Europese portalen

Momenteel bestaan er twee projecten bij de Europese Commissie om een overkoepelend portaal te maken waarin de meta-data van alle andere open data portalen in Europa opgenomen kan worden: een portaal gebaseerd op het CKAN-pakket van OKFN, en een ander portaal gebaseerd op andere software van PWC.

Het is niet duidelijk welk portaal er zal overblijven (= voldoende budget zal krijgen), maar wellicht zal Europa op een bepaald moment toch vragen dat andere (nationale, regionale en lokale) portalen hun meta-data doorsturen naar een van deze portalen.

CKAN en OKFN

Veel buitenlandse portalen gebruiken het open source pakket CKAN van de Open Knowledge Foundation (OKFN), in België is er meer variatie te merken:

- Vlaanderen gebruikt CKAN.
- Gent en Antwerpen gebruiken The Data Tank met een Drupal front-end.
- Brussel (Stad) heeft onlangs een nieuwe site met OpenDataSoft gelanceerd.
- Fedict (data.gov.be) gebruikt Drupal 6.
- DAV (publicdata.belgium.be) gebruikt Drupal en Sharepoint.

Het is niet de bedoeling dat alle portalen verplicht worden om dezelfde software te gebruiken, het is echter wel de bedoeling dat voor de uitwisseling van meta-data hetzelfde formaat gebruikt wordt.

Standaard meta-data formaat (Belgische variant van DCAT)

DCAT is een internationaal meta-data-formaat dat ontstaan is uit de eerste open data portalen (van o.a. USA en UK), en nu verder ontwikkeld wordt door W3C.

DCAT is voor onze doeleinden niet specifiek genoeg, omdat de meeste velden optioneel zijn en omdat er geen aanbevelingen zijn in verband met de te gebruiken categorisaties.

Een Europese werkgroep heeft dit formaat dan ook verder verfijnd naar DCAT-AP, en het is de bedoeling om dit nog verder uit te werken naar een Belgische variant. Vlaanderen heeft hiervoor al een eerste aanzet gegeven in het Open Data handboek.

Het is de bedoeling dat de portalen in België deze Belgische variant ondersteunen, en meta-data kunnen exporteren naar dat formaat. Het is echter niet de bedoeling om de datasets zelf te gaan kopiëren: het is beter om de data te laten staan waar ze staat, en gewoon terug te linken.

Uitwisselen van meta-data

Waarschijnlijk zal er zo een systeem ontstaan waarbij lokale portalen meta-data doorsturen naar de regionale portalen, regionale portalen naar nationale en de nationale portalen naar de Europese. Ook de omgekeerde beweging moet overigens mogelijk zijn.

Bovendien is het niet ondenkbaar dat een derde partij zelf portalen begint te bouwen, maar dan ingedeeld per thema / sector.

Vertalingen / meertaligheid

Een aandachtspunt is meertaligheid: verschillende overheden willen of moeten gegevens ter beschikking stellen in meerdere talen.

In veel portaalsoftware (zoals CKAN) is het wel mogelijk om met meerdere talen te werken, maar is het niet mogelijk om aan te geven dat een bepaalde beschrijving een vertaling is van een andere beschrijving, ook al is de onderliggende dataset dezelfde. Met andere woorden, er treden dubbels op. Dit hoeft op zich geen probleem te zijn, maar kan soms aanleiding geven tot problemen bij het onderhouden van de vertalingen.

Het is niet de bedoeling om alle meta-data (beschrijvingen, titels ...) altijd te gaan vertalen naar 2, 3 of zelfs 4 talen (drie landstalen + Engels), dat hangt onder meer af van de taalwetgeving en de beschikbaarheid / betaalbaarheid van vertaaldiensten.

Het is wel de bedoeling om expliciet aan te geven in welke taal een titel, beschrijving... geschreven is, zodat de geautomatiseerde verwerking van de meta-data makkelijker wordt. Dit is overigens voorzien in DCAT.

Mappen naar bestaande thesauri

Zowat elk portaal gebruikt een eigen indeling in thema's, eigen keywords, ... Dit is zeer flexibel, maar maakt het lastiger om gelijkaardige data sets te vinden in andere portalen.

Het is dan ook de bedoeling om de eigen lijsten bijkomend te "mappen" naar reeds bestaande, internationaal gebruikte thesauri. Op deze manier kan de eigen indeling - die vaak iets verfijnder is - behouden blijven, en wordt het tegelijk eenvoudiger om de meta-data automatisch te verwerken.

Om een thema / categorie aan te duiden, kan bijvoorbeeld naar Eurovoc verwezen worden. Eurovoc is een thesaurus die door het Europese Publication Office wordt beheerd en gebruikt, en die vertaald is naar alle officiële talen van de EU. Onlangs werd deze thesaurus vrijgegeven onder een open licentie, en deze kan dan ook gratis gedownload worden.

Om aan te geven over welk geografisch gebied het gaat, kan eventueel de Europese NUTS classificatie gebruikt worden. Nadeel is dat NUTS enkel tot op arrondissement-niveau gaat, en bijvoorbeeld niet tot op het niveau van gemeentes of straten. Misschien is Geonames een oplossing.

Bij voorkeur wordt voor het mappen het SKOS-formaat gebruikt (een W3C specificatie), maar desnoods kan een gewone spreadsheet gebruikt worden, waarbij men kan aangeven of een term van het portaal hetzelfde (skos:exactMatch) of specifieker (skos:narrowMatch) is dan een term van een gekende thesaurus.

Handboek open data

Vlaanderen heeft een tweede versie van het Handboek Open Data gepubliceerd, die voorlopig enkel in het Nederlands beschikbaar is. Suggesties (en vertalingen) zijn steeds welkom.

Open data project Vlaams parlement

Vlaanderen geeft subsidies aan enkele open data projecten ("VIP"), onder andere aan een project om de website van het Vlaams parlement te vernieuwen. Hierbij zullen ook gegevens als open data gepubliceerd worden onder andere de lijst van de volksvertegenwoordigers.

Het zou interessant zijn als ook de andere parlementen in België een gelijkaardig project zouden kunnen realiseren, eventueel met hergebruik van de kennis en/of software van het Vlaams Parlement. Op regionaal en Europees niveau zijn er namelijk al wat privé-initiatieven ontstaan die dit soort gegevens willen gebruiken.

Enkele linken

Projecten

EU Open Data Support (gratis online training etc.)

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

EU Linked Open Data project (tools)

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

EU Location Framework

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

(niet te verwarren met het European Location Framework <http://www.elfproject.eu/>)

Vlaamse Open Data projecten "VIP"

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

Portalen

EU Open Data Portal (enkel van de Europese instellingen)

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

EU Open Data Portal, gebaseerd op CKAN (gemaakt door OKFN, beta)

- <http://publicdata.eu/>

EU Open Data Portal ODIP (gemaakt door PriceWaterhouseCoopers)

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

Specificaties en thesauri

Eurovoc Thesaurus

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

EU NUTS classificatie (Eurostat)

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

Vlaamse Open Data handleiding 2.0 (en meer specifiek, deel 8.1 Metadata, p 36 - 39)

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

Actiepunten

- (Allen) Bekijken om thema's / keywords van eigen portaal te mappen naar Eurovoc termen
- (Allen) Bekijken wat er precies moet gedaan worden om meta-data van eigen portaal te exporteren naar Belgische versie van DCAT-AP (is er genoeg expertise, budget ...)
- (Fedict) Uitsuren Doodle om datum van de volgende vergadering vast te leggen, Stad Kortrijk en FOD Economie ook uitnodigen
- (Vlaanderen) Contactinformatie i.v.m. VIP project Vlaams Parlement doorsturen naar DAV