

# Open Data Portals overleg

Datum : 23 juni 2014, 14:00 - 16:00

Locatie: Vlaanderen BZ, Boudewijngebouw, Brussel

Aanwezig:

- AWT
- DAV
- Fedict
- FOD Economie (KBO + Algemene Directie Statistiek)
- Stad Antwerpen
- Stad Brussel
- Vlaanderen BZ
- ProXML (invited expert)
- TenForce (invited expert)

Verontschuldigd:

- Ethnic
- Stad Gent
- Stad Kortrijk
- Stad Namur
- Intercommunale Leiedal
- WestToer

# Bespreking

## Ontsluiten KBO

Sinds enkele maanden is de Kruispuntbank Ondernemingen (KBO) van de FOD Economie beschikbaar in 2 vormen: als bestanden die gratis te downloaden zijn en als betalende webservice (50 EUR voor 2000 online requests).

Registratie is gratis, maar wel verplicht. In juni waren er een 50-tal gebruikers van de gratis bestanden, en 1 gebruiker van de betalende webservice.

- <http://economie.fgov.be/nl/ondernemingen/bce/pub/opendata>
- <http://economie.fgov.be/nl/ondernemingen/bce/pub/PuSSW>

ProXML heeft een website gemaakt om de KBO-gegevens te ontsluiten als Linked Open Data.

- <http://kbodata.be>

In het buitenland zijn er soortgelijke initiatieven, zoals het OpenCorporates project dat gegevens van bedrijven uit 50 landen bevat.

- <http://opencorporates.com>

## Licenties

De Europese Commissie heeft een rondvraag georganiseerd om te kijken welke open data licenties er gebruikt worden in de lidstaten. België is wat dat betreft uniek, omdat er zoveel verschillende licenties worden gebruikt:

- de Stad Brussel gebruikt een licentie van Etalab (een dienst die opgericht werd door de Franse Eerste Minister);
- de Vlaamse Open Data licenties zijn afgeleid van de Creative Commons licenties;
- de Stad Antwerpen gebruikt nu een Vlaamse Open Data licentie;
- de Stad Gent gebruikt een eigen licentie;
- het KBO gebruikt ook een eigen licentie;
- AWT is geïnteresseerd in de modellicenties die de DAV heeft ontworpen.

Dit maakt het voor ontwikkelaars moeilijk om gegevens van verschillende bronnen te combineren. Een Europese licentie zou misschien een oplossing zijn, al is dat misschien niet mogelijk door subtiele verschillen in copyright-wetgeving.

## Relatie tussen DCAT, DCAT-AP en de Belgische variant

Data Catalog Vocabulary (DCAT) is een specificatie die ontwikkeld werd door de National University van Ierland, en die nadien overgedragen werd aan het W3C. Deze specificatie beschrijft een meta-data formaat voor open data sets, gebaseerd op de eerste open data portalen ([data.gov.uk](http://data.gov.uk) en [us](http://data.gov.us)).

DCAT is echter te algemeen, en de Europese Commissie heeft in 2013 een werkgroep opgericht om een Application Profile te ontwikkelen: DCAT-AP. Deze variant is compatibel met DCAT, maar bevat richtlijnen over welke gegevens verplicht in te vullen zijn, en welke gegevens optioneel zijn.

Ook dit profiel kan nog verder uitgewerkt worden, dit is dan de Belgische variant van DCAT-AP, waarin bijvoorbeeld nog meer aandacht wordt besteed aan de te gebruiken classificaties.

Opmerking: in de Europese Open Data context is het zo dat nieuwe initiatieven ondersteuning voor DCAT-AP vereisen (bv. zie de recente openbare aanbesteding van de nieuwe Zweedse Open Data Portal).

## Europese portalen

Op dit moment zijn er verschillende open source portals op Europees niveau:

- De Europese Unie open data portal ([open-data.europa.eu](http://open-data.europa.eu)), welke alle datasets die beschikbaar worden gesteld door alle Europese Unie instellingen omvat. TenForce was hier de ontwikkelaar, samen met OKFN.
- Een CKAN-platform ([publicdata.eu](http://publicdata.eu)), ontwikkeld door de Open Knowledge Foundation (OKFN) in het kader van het LOD2 project ([lod2.eu](http://lod2.eu)). Het bevat datasets van openbare besturen (landen, regio's, steden). De gegevens worden verzameld (harvested) door de open data portals van die openbare besturen te ondervragen. De bronsystemen moeten CKAN systemen zijn.
- Een Open Data Interoperability Platform (ODIP) ontwikkeld door TenForce en PwC in het kader van het Open Data Support project ([data.opendatasupport.eu](http://data.opendatasupport.eu)). Het bevat datasets van nationale open data portals met die van de Europese Unie. De veronderstelling is dat elke EU lidstaat op nationaal niveau alle gegevens van de regionale of stedelijk niveau verzameld heeft.

Het ODIP-systeem kan dataset informatie van andere portalen verzamelen en via SPARQL transformeren naar DCAT-AP. Momenteel gaat het vooral over CKAN-portalen, maar ook RDF dumps en CSV dumps worden ondersteund. De import kan worden uitgebreid naar eender welk input formaat.

Het opzetten van zo'n transformatieproces is ongeveer een week werk, daarna kan de basisinformatie (meta-data) van een portaal automatisch opgehaald en verwerkt worden. Nadien kan gekeken worden om de transformatie verder te verbeteren en te verfijnen.

De datasets zelf worden overigens niet gekopieerd, die blijven gewoon staan waar ze stonden. Het platform is een (open-source) uitbreiding van de open source tool LODMS (nu Unified Views).

Mogelijk zullen de twee Europese platformen samensmelten tot 1 geheel (waarschijnlijk wordt hiervoor een lastenboek gepubliceerd), waarbij het CKAN-deel zal dienen als front-end (= website voor eindgebruikers) en het ODIP-deel als back-end (die de meta-data van de portalen van de lidstaten verder zal blijven verzamelen en transformeren naar het DCAT-AP formaat).

## Belgische portalen, export naar DCAT-AP

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: de Belgische variant van DCAT-AP.

Om dit doel te bereiken zouden alle portalen en sites dus een DCAT-AP bestand moeten aanleveren, zodat dit bijvoorbeeld door het Europese ODIP systeem verwerkt kan worden (of door om het even welk ander systeem dat DCAT kan verwerken).

Een stand van zaken:

- Brussel (Stad) heeft onlangs een nieuwe site met OpenDataSoft gelanceerd. Dit systeem genereert al een volledige export van meta-data naar DCAT, en is vrij eenvoudig te transformeren naar DCAT-AP.
- Vlaanderen BZ/Corve en AWT gebruiken allebei CKAN, maar de structuur van de Vlaamse en Waalse CKAN verschilt, en geen van beide portalen exporteert naar DCAT omdat CKAN de voorkeur geeft aan een eigen API. Sommige buitenlandse portalen hebben dan maar zelf een DCAT-module geschreven, maar deze zijn portaal-specifiek en/of werken niet op de laatste versie van CKAN. Misschien zouden Corve en AWT kunnen samenwerken om een DCAT-AP module te (laten) ontwikkelen.
- Gent en Antwerpen gebruiken een oudere versie van The Data Tank met een Drupal front-end. Hierbij zit een deel van de meta-data in TDT en een andere deel (vooral dan de beschrijving) in Drupal. Momenteel is er geen DCAT-AP export voorzien. Binnenkort neemt de stad Antwerpen wel een nieuwe versie van TDT in gebruik, maar het is nog afwachten of dat iets aan de situatie verandert.
- DAV ([publicdata.belgium.be](http://publicdata.belgium.be)) gebruikt Drupal en Sharepoint, er is momenteel geen export naar DCAT-AP voorzien.
- Fedict ([data.gov.be](http://data.gov.be)) gebruikt Drupal 6, er is momenteel geen export naar DCAT-AP voorzien.
- FOD Economie (statbel) is een Tridion site, er is momenteel geen export naar DCAT-AP voorzien. De Directie Statistiek van de FOD Economie stuurt ook gegevens door naar EuroStat, maar dat zijn niet noodzakelijk dezelfde gegevens die (ook) op [statbel.fgov.be](http://statbel.fgov.be) gepubliceerd worden.

Als tussenoplossing kunnen er scraping-tools gebouwd worden (Fedict werkt aan een voorbeeld) die gegevens van open data portalen omzetten naar DCAT-AP.

## Hackathons

De Stad Kortrijk organiseerde voor de tweede maal een hackathon ("AppLab 2").

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

De Stad Brussel organiseert op 17 en 18 oktober een hackathon.

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

## Actiepunten

- (Allen) Bekijken om thema's / keywords van eigen portaal te mappen naar Eurovoc termen.
- (Allen) Contactpersonen zoeken van het federaal en de regionale parlementen om te kijken of er interesse is om het open data systeem van het Vlaams parlement te hergebruiken.
- (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 ...).