OSLO²: Thematische Werkgroep Adres & Gebouw 1

# Details

|  |  |
| --- | --- |
| Datum | 2016-06-08 |
| Locatie | AIV Gent, Koningin Maria Hendrikaplein, Zaal Albert Servaes |
| Gerelateerde documenten | MappingOSLO-VBRv03.xlsx |
| Notulen | Dieter De Paepe, Dries Beheydt |

## Aanwezigen

|  |  |  |
| --- | --- | --- |
| **Naam** | **Afkorting** | **Organisatie** |
| Paul Hermans | PH | ProXML |
| Laurens De Vocht | LDV | iMinds |
| Dieter De Paepe | DDP | iMinds |
| Barbara Van Broeckhoven | BVB | AIV |
| Tony Vanderstraete | TV | AIV |
| Ruben Capiau | RC | AIV |
| Jan Laporte | JL | AIV |
| Hendrik Van Hemelryck | HVH | AIV |
| Raf Buyle | RB | AIV |
| Ziggy Vanlishout | ZV | AIV |
| Geert Thijs | GT | AIV |

## Verontschuldigd

Veerle Beyaert, Liesbet De Wolf, Hans Vliebergh

# Agenda

* OSLO en OSLO²
  + Historiek OSLO
  + Het OSLO²-traject
  + Wat gaat OSLO² doen?
* CRAB-LOD
  + Intro: Linked Open Data sterrenmodel
  + Toelichting status POC
* Semantisch kader
  + Startpunt
  + Over de mapping zelf
  + Conclusies
* Project gebouwenregister en OSLO²
  + Waar moeten we het over hebben tussen beide projecten
  + Vaststellingen en afgeleide actiepunten

Dit document is een beknopte samenvatting van de thematische werkgroep: Adres & Gebouw 1, die plaats vond op 2016-06-08.

# Inleiding

Er wordt een korte intro gegeven over OSLO². (Link naar “Over OSLO²”)

Gezien het feit dat de geplande CRAB-LOD proof of concept zich focust op het adres domein, is het behandelen van dit domein iets dringender dan de ander. Op zich zou dit geen probleem mogen vormen aangezien er reeds een stabiele basis is en dit het meest gebruikte product van AIV is tot op heden.

# Proof of concept: CRAB-LOD

|  |
| --- |
| Discussie |

Linked Open Data wordt typisch geclassificeerd volgens het [5 sterren model](http://5stardata.info/en/) van Tim Berners-Lee:

* 1 ster: zet de data online met een open licentie
* 2 sterren: vorige + gebruik een gestructureerd formaat
* 3 sterren: vorige + gebruik een niet-gepatenteerd, open formaat
* 4 sterren: vorige + gebruik RDF en SPARQL als ontsluitingsvorm
* 5 sterren: vorige + link naar de data van anderen

Een voorbeeld van een 5 sterren implementatie is de data van de Kruispuntbank van Ondernemingen (KBO) gepubliceerd als Linked Open Data op <http://data.kbodata.be/>. De web-pagina’s zijn een simpele visualisatie van de achterliggende RDF triples. Een gelijkaardige aanpak zal gevolgd worden voor deze proof of concept (POC).

In CRAB-LOD is het doel het centraal referentieadressenbestand (CRAB) open te stellen via een SPARQL endpoint (SPARQL is de querytaal voor RDF), de data te ontsluiten via een 5 sterren classificatie en alle identifiers resolvable te maken, zoals bij het voorbeeld van de KBO het geval is.

De ontwikkeling gebeurt op een kopie van de adressen van het gebouwenregister gehost op Microsoft Azure. Voor de beta versie is het de bedoeling om met een ontsluitingsdatabase te werken. Via een [R2RML](https://www.w3.org/TR/r2rml/) mapping kunnen de waarden uit deze relationele databank omgezet worden naar RDF. Omdat de OSLO² ontologie nog volop in ontwikkeling is, wordt er voorlopig gewerkt met een voorbeeld ontologie. Er zullen in de POC 2 aanpakken vergeleken worden:

* Virtualisatie, door middel van een translatie van SPARQL naar SQL queries wordt een virtuele SPARQL endpoint gemaakt.
* De opzet van een echte SPARQL endpoint, waarbij via synchronisatie de data in de triple store up to date wordt gehouden met die uit de relationele databank.

**Opmerking GT:** Het is een vreemde vaststelling om te zien dat we onze eigen ontologie maken maar dat in de resultaten van een query bijvoorbeelde een W3C ontologie voorkomt.

**Antwoord:** Dit is een manier om interoperabiliteit te bekomen in de Linked Data wereld. Door reeds gedefinieerde termen te hergebruiken, is het makkelijker om data samen te voegen of te interpreteren. Ook voor OSLO² zullen we moeten kijken of er bestaande termen zijn die overeenkomen met onze definities.

|  |
| --- |
| Actiepunten |

Na de POC moeten we bekijken of geoSPARQL geschikt is voor onze nodes op gebied van geografische zaken, dit wordt immers de industrie standaard.

# Semantisch kader

|  |
| --- |
| Discussie |

Door het opstellen van een vergelijking tussen de definities van OSLO1 (adres en gebouw) en de definities uit de Vlaamse Basisregistraties (VBR) kunnen we achterhalen wat de mismatch is tussen de twee. Aangezien de definities van VBR/[GRB](https://www.agiv.be/producten/grb) (Grootschalig Referentie Bestand) gebruikt waren bij het opstellen van OSLO1 zijn er geen grote verschillen, echter zijn er wel enkele kleine verschillen zoals bijvoorbeeld:

* velden die gebundeld zitten in andere velden (zo ontbreekt “straatnaam” als apart veld in OSLO1)
* velden die ontbreken (zo ontbreekt “aantalVerdiepingen” in OSLO1)
* verschillen in het datatype (“postcode” is geen codelijst in OSLO1)

**Opmerking:** Meertaligheid is inherent ondersteund in RDF door gebruik van “language tags”, bijvoorbeeld: “Belgium”@en

Ook zijn er verschillen door het streven naar internationale ondersteuning. Zo moet een buitenlands adres ook mogelijk zijn in OSLO², een Vlaams bedrijf kan immers een buitenlandse zetel hebben, maar dit zou een te open formaat zijn voor de Vlaamse bronnen met enkel Vlaamse adressen. Als oplossing hiervoor kunnen er extra restricties gebruikt worden door instanties waar gepast, zodat de interne data kwaliteit veilig gesteld is, terwijl het algemene uitwisselformaat flexibel genoeg blijft.

|  |
| --- |
| Actiepunten |

De specifieke verschillen opgemerkt in de mismatch zullen behandeld worden in (aanloop naar) de volgende werkgroep.

Een volledig overzicht van de conflicten tussen OSLO1 en VBR is te vinden in de spreadsheet “MappingOSLO-VBRv03.xlsx”, tabblad “MappingOSLO2VBR”.

# Project gebouwenregister

|  |
| --- |
| Beslissing |

Er is reeds een project lopend over het modelleren binnen het gebouwenregister. Hierdoor is het moeilijk om af te stemmen voor OSLO². Om de vooruitgang van het gebouwregister project niet te verstoren zal OSLO² zelf de gevonden issues uitklaren en achteraf (vanaf oktober) afstemmen zodat zij als klankbord kunnen dienen. Ook voor de CRAB proof of concept zal er afstemming nodig zijn.

# Besluit en acties

## Besluit

We moeten eerst de oefening intern houden om dan achteraf met het team van het Gebouwenregister af te stemmen. Zo kunnen de project resources voor het Gebouwregister gefocust blijven. Pas vanaf oktober is afstemming mogelijk omdat ook het datamodel daar nog gewijzigd wordt. De groep van het Gebouwenregister zal vooral als klankbord dienen. Dit betekent ook dat oktober de richtdatum wordt om een ontwerpmodel klaar te hebben voor OSLO² om te kunnen voorleggen aan de groep van het Gebouwenregister.

Ook de infrastructuur/architectuur van het Gebouwenregister is nog niet finaal, er zal afstemming nodig blijven tussen het team van CRAB-LOD en het Gebouwenregister.

## Acties

* Mikken op oktober 2016 als eerste moment waarop de OSLO²-modellering van gebouw en adres wordt voorgelegd aan de bestaande werkgroep Gebouwenregister. Deze zal fungeren als klankbord om de uitkomst van de interne workshops te laten valideren.
* Samen met de betrokken consultants van Tenforce en ProXML wordt voor de piloot van CRAB-LOD een planning opgemaakt van de bouw en oplevering. Deze planning kunnen we vervolgens toetsen aan de releaseplanning van het gebouwenregister. Verder is het ook belangrijk dat voor de uiteindelijke CRAB-LOD – vanuit de non-functional requirements – een zicht wordt gegeven op de benodigde infrastructuur. Dit heeft namelijk ook impact op de database van het Gebouwenregister. Deze oefening loopt momenteel: de POC die PH nu doet met een aantal tools heeft als doel om de non-functionals zinvol te kunnen formuleren.