

STANDAARD VOOR HET METADATEREN EN PUBLICEREN VAN LINKED OPEN DATA

Versie /// 1.1

Publicatiedatum /// 20 juli 2017

Naam	Rol	Organisatie	E-mailadres
Mathias De Schrijver	Auteur	Informatie Vlaanderen	mathias.deschrijver@kb.vlaanderen.be
Geraldine Nolf	Auteur	Informatie Vlaanderen	geraldine.nolf@kb.vlaanderen.be
Laurens De Vocht	Reviewer	imec – UGent - IDLab	laurens.devocht@ugent.be
Paul Hermans	Reviewer	ProXML	paul@proxml.be
Bert Van Nuffelen	Reviewer	Tenforce	bert.van.nuffelen@tenforce.com
Stijn Goedertier	Reviewer	GIM	stijn.goedertier@gim.be
Dries Beheydt	Reviewer	Informatie Vlaanderen	dries.beheydt@kb.vlaanderen.be

Naam	Organisatie	E-mailadres
Dieter De Paepe	imec – UGent - IDLab	dieter.depaepe@ugent.be
Mathias Van Compernelle	imec	mathias.vancompernelle@imec.be
Ziggy Vanlshout	Informatie Vlaanderen	ziggy.vanlshout@kb.vlaanderen.be
Dirk De Baere	Informatie Vlaanderen	dirk.debaere@kb.vlaanderen.be
Quincy Oeyen	Informatie Vlaanderen	quincy.oeyen@kb.vlaanderen.be
Birgit Bastiaensen	Agentschap Innoveren en Ondernemen	birgit.bastiaensen@vlaio.be
Koen Van Daele	Agentschap Onroerend Erfgoed	koen.vandaele@vlaanderen.be
Geert Van Haute	Agentschap Onroerend Erfgoed	geert.vanhaute@vlaanderen.be
Jelle De Vreese	Cipal / Schaubroeck nv	jelle.devreese@cipalschaubroeck.be
Wim Van Acker	Cipal / Schaubroeck nv	wim.vanacker@cipalschaubroeck.be

Datum publicatie: 20/07/2017

Hoofdzetel

Regionale zetel

informatie.vlaanderen@vlaanderen.be

////////////////////////////////////

INHOUD

Inhoud.....	3
1 OPDRACHT	5
1.1 Aanpak	5
2 Achtergrond & kader	5
2.1 OSLO ²	5
2.2 Technische werkgroep.....	5
3 Hoe LOD-producten metadateren?	6
3.1 Concrete vragen.....	6
3.2 Standaard.....	7
4 Hoe Linked data publiceren als open data?	8
4.1 Concrete vragen.....	8
4.2 Standaard.....	9

////////////////////////////////////

1 OPDRACHT

Binnen het project OSLO² is een technische werkgroep opgericht met als opdracht het uitwerken van standaard voor:

- het metadateren van linked open data; en
- het publiceren van linked data als open data.

1.1 AANPAK

Informatie Vlaanderen zal deze standaard in praktijk brengen voor CRAB-LOD, het linked open dataproduct (LOD) van de authentieke bron met adresgegevens (CRAB¹) zoals binnen het OSLO²-project² ontwikkeld. Uiteraard is de standaard breed toepasbaar en relevant voor alle linked open data van overheden in Vlaanderen.

Een aantal elementen uit de standaard zijn gemeenschappelijk omdat ze zowel gelden voor het metadateren van LOD-producten, als gelden voor het open beleid van LOD-producten. (Deze standaard krijgen een MD/OD-code.) Een aantal elementen uit de standaard zijn specifiek en gaan enkel over of het metadateren of het open beleid van LOD-producten. (Deze specifieke elementen uit de standaard krijgen in dat geval enkel een MD-code of een OD-code.)

De standaard voor het publiceren van linked data als open data zal op termijn onderdeel uitmaken van de open data handleiding³ zoals gepubliceerd op <http://overheid.vlaanderen.be/opendata>.

2 ACHTERGROND & KADER

2.1 OSLO²

Voor meer achtergrondinformatie met betrekking tot het OSLO²-project, de projectdocumentatie en de resultaten wordt verwezen naar <https://overheid.vlaanderen.be/producten-diensten/OSLO2>.

2.2 TECHNISCHE WERKGROEP

Er is bij de start van de technische werkgroep voor gekozen de expertise binnen Informatie Vlaanderen m.b.t. metadata en open data samen te brengen in één werkgroep. Het publiceren van linked data als open data

¹ <https://metadata.agiv.be/zoekdienst/apps/tabsearch/?uuid=f5418a16-61da-4175-8d86-19d3d6e26a4f>

² <https://overheid.vlaanderen.be/producten-diensten/OSLO2>

³ <https://overheid.vlaanderen.be/open-data-handleiding>

gebeurt immers aan de hand van metadata. Omwille van de overlap tussen beide domeinen is gekozen voor een gecombineerde aanpak.

De expertise m.b.t. (het metadateren van) linked open data binnen de overheidssector in Vlaanderen is relatief beperkt. Om die reden is er voor gekozen van bij de start de leveranciers van het OSLO² project i.c. Laurens De Vocht (imec), Bert Van Nuffelen (Tenforce), Paul Hermans (ProXML) en Stijn Goedertier (GIM) nauw te betrekken bij de opmaak van deze standaard. Zij hebben dit rapport nagelezen en opmerkingen geformuleerd.

Na het open evenement rond OSLO² op 9 december 2016 (OSLO² co-creatiedag) werd de technische werkgroep opengesteld voor iedereen met kennis van / interesse voor metadata en open data. In de periode september 2016 – maart 2017 is de technische werkgroep vier maal samengekomen (2 interne sessies en 2 publieke sessies).

Doelstelling van de werkgroep was in onderling overleg goede praktijken neer te schrijven voor het metadateren en publiceren van linked open data. Uiteindelijk is er voor gekozen deze goede praktijken als standaard te formuleren zoals in voorliggend document verder wordt toegelicht.

3 HOE LOD-PRODUCTEN METADATEREN?

Hierna volgt een beknopt overzicht van de concrete vragen van leden van de werkgroep m.b.t. het metadateren van linked open data. Deze vragen zijn thematisch geclusterd. De standaard zal op elk van deze vragen een antwoord formuleren.

3.1 CONCRETE VRAGEN

- *Granulariteit*
 - Is het aangewezen linked open data te metadateren als afzonderlijke dataset / product of als distributie van een dataset / product?
- *Standaard*
 - Welke metadatastandaard gebruiken we voor het metadateren van linked open data? ISO/INSPIRE, (Geo)DCAT-AP of beide?
- *Tooling*
 - Welke tooling gebruiken we voor het metadateren van linked open data? GeoNetwork, CKAN of beide?
- *Andere relevante topics?*
 - Wat met extra meta-informatie, zoals binnen VOID-pagina's, Subject pages, of ev. andere?

////////////////////////////////////

3.2 STANDARD

MD1 / OD1: Je beschrijft linked open data in de metadatastandaard van het domein van de gegevens. Voor linked open data van geografische informatie wordt ISO/INSPIRE en (Geo)DCAT-AP (via harvesting) als metadatastandaard gebruikt. Voor linked open data van niet-geografische informatie is de te gebruiken metadatastandaard DCAT-AP.

Toelichting: DCAT-AP is de de facto standaard voor het metadateren van (linked) open data. Voor geografische data vormen de ISO-standaarden en de Technische Richtlijnen voor INSPIRE metadata de beste garantie om aan de wettelijke verplichtingen te voldoen. GeoDCAT-AP is een extensie van DCAT-AP die toelaat geografische datasets te beschrijven. Idealiter verloopt de omzetting van ISO/INSPIRE naar GeoDCAT-AP geautomatiseerd. Deze omzetting is uni-directioneel en kan gepaard gaan met een (beperkt) verlies aan informatie.

MD2 / OD2: Linked open data wordt als distributie van een dataset gepubliceerd en niet als afzonderlijke dataset.

Toelichting: Wie een *bestaande* dataset als linked data ontsluit, voegt deze verschijningsvorm van hetzelfde product in de bestaande metadata-record als een extra ‘distributie’ toe. Dit is consistent met de definitie van ‘dataset’ en ‘distributie’ die door DCAT-AP gehanteerd wordt: verschillende datasets hebben een verschillende inhoud, eenzelfde dataset wordt aan de hand van één of meerdere distributies / formaten / services ter beschikking gesteld.

Ook voor de ISO/INSPIRE-standaard wordt dit principe doorgetrokken. Linked open data metadateren van geografische gegevens betekent dat je deze beschrijft als een distributie-vorm van de dataset waarop dit product gebaseerd is. Je hoeft geen afzonderlijke dataset te voorzien voor het linked open dataproduct zelf.

Bovenstaande gaat zowel op voor bestaande als nieuwe datasets / producten. Echter, nieuwe datasets / producten kunnen ook gebaseerd zijn op één of meerdere bestaande datasets en een gereduceerde, uitgebreide of gecombineerde inhoud bevatten ten opzichte van de oorspronkelijke dataset(s). In dat geval ontstaat een *nieuwe* dataset / product en is de LOD-distributie die wordt aangeboden een distributie van de nieuwe dataset en niet van (één of meerdere) oorspronkelijke datasets waarvan men vertrokken is.

MD3 / OD3: Je ontsluit linked open data via alle portalen die geschikt zijn om dat soort gegevens te ontsluiten en gebruikt hiervoor de bijhorende tooling.

Toelichting: Net omdat je de aanbeveling MD1 / OD1 volgt, zullen je gegevens op alle portalen terecht komen waar de soort gegevens thuis horen. Dit betekent bijvoorbeeld dat geografische linked open data zowel op het geografische portaal Geopunt als op het Vlaamse Open data portaal (via harvesting) ontsloten worden. Voor deze datasets geldt het once only principe, zij dienen niet opnieuw van metadata te worden voorzien in het Vlaams Open data portaal.

Alsook dat de niet-geografische linked open data via het Vlaamse Open data portaal en eventueel eigen open data portalen van Vlaamse, provinciale en lokale overheden ontsloten zijn. Overheden met een eigen portaal zorgen er voor dat hun (linked) open datasets ook op het Vlaams Open data portaal terug te vinden zijn. Voor het metadateren van linked open data op deze portalen maak je gebruik van bijhorende tooling, conform standaard MD1 / OD1.

Het Vlaams Open data portaal is de gids voor alle open data van overheden in Vlaanderen. Het portaal geeft bovendien invulling aan de decretale verplichting⁴ om een portaal op te zetten met voor hergebruik beschikbare bestuursdocumenten.

MD4: Je voorziet voor je linked open data, waar je dit kan, bijkomende meta-informatie binnen VOID-pagina's.

Toelichting: VOID is een vocabularium om linked open datasets te beschrijven met extra informatie, die machine-leesbaar is.

Beschrijvende metadata zijn reeds aanwezig in DCAT- of ISO/INSPIRE-metadata, enkel het toevoegen van technisch/structurele metadata cfr. VOID biedt meerwaarde. VOID-pagina's zijn bijgevolg geen must, maar spreken wel het typische linked open data publiek aan.

Als jouw technische oplossing standaard VOID kan aanbieden, dan bevelen we aan om die VOID beschrijving te voorzien.

MD5: Je voorziet voor je linked open data bijkomende meta-informatie als onderdeel van de subject-pagina's voor persistente URI's.

Toelichting: Subject-pagina's beschrijven het object (de eigenschappen en zijn relaties) en de metadata die het object beschrijft. Wij bevelen aan om als onderdeel van de subject-pagina's voor persistente URI's een link op te nemen naar de metadata-record van de brondataset waar dit object toe behoort.

Voor geografische data is dit de ISO/INSPIRE-metadatarecord, aangezien we steeds willen doorverwijzen naar de meest volledige beschrijving van de brondataset (de oorspronkelijke metadata). Zo krijgt de lezer meer informatie over het grotere geheel waar het object in thuis hoort: een algemeen kader, de contactgegevens van de eigenaar, de licentie, om welke versie van de dataset het gaat, enz.

4 HOE LINKED DATA PUBLICEREN ALS OPEN DATA?

Hierna volgt een beknopt overzicht van de concrete vragen van leden van de werkgroep m.b.t. het publiceren van linked open data. Deze vragen zijn thematisch geclusterd. De standaard zal op elk van deze vragen een antwoord formuleren.

4.1 CONCRETE VRAGEN

- *Vragen m.b.t. de granulariteit*
 - Is het aangewezen linked data te publiceren als afzonderlijke dataset of als distributie van een bestaande dataset?

⁴ <https://overheid.vlaanderen.be/omzetting-psi-richtlijn>

- *Vragen m.b.t. de metadata*
 - Welke metadatastandaard gebruiken we voor het publiceren van linked open data? DCAT-AP⁵ is de de facto standaard in Europa, wat met INSPIRE-data⁶? Hoe linked datasets metadateren op het Vlaams open data portaal?
- *Vragen m.b.t. portalen*
 - Welk portaal of portalen gebruiken we voor het publiceren van linked open data? Het Vlaams Open data portaal⁷ en eventueel eigen open data portalen bij Vlaamse bestuursinstanties liggen voor de hand, maar wat met Geopunt⁸?
- *Vragen m.b.t. formaten / services*
 - Welk(e) formaat / formaten of services zijn aangewezen / haalbaar bij het publiceren van linked open data? Kiezen we hierbij voor een RDF-dump of een SPARQL⁹-endpoint of beide?

4.2 STANDAARD

Voor een antwoord op de vragen m.b.t. de metadatastandaard, de granulariteit en de portalen worden de elementen van de standaard uit paragraaf **Error! Reference source not found.** herhaald. Voor deze topics vallen de elementen uit de standaard voor het metadateren / publiceren van linked (open) data samen. Andere elementen uit de

standaard hebben specifiek betrekking op het publiceren van linked open data.

MD1 / OD1: Je beschrijft linked open data in de metadatastandaard van het domein van de gegevens. Voor linked open data van geografische informatie wordt ISO/INSPIRE en (Geo)DCAT-AP (via harvesting) als metadatastandaard gebruikt. Voor linked open data van niet-geografische informatie is de te gebruiken metadatastandaard DCAT-AP.

Toelichting: DCAT-AP is de de facto standaard voor het publiceren van (linked) open data. Voor geografische data vormen de ISO-standaard en de Technische Richtlijnen voor INSPIRE metadata de beste garantie om aan de wettelijke verplichtingen te voldoen.. GeoDCAT-AP is een extensie van DCAT-AP die toelaat geografische datasets te beschrijven. Idealiter verloopt de omzetting van ISO/INSPIRE naar GeoDCAT-AP geautomatiseerd. Deze omzetting is uni-directioneel en kan gepaard gaan met een (beperkt) verlies aan informatie.

MD2 / OD2: Linked open data wordt als distributie van een dataset gepubliceerd en niet als afzonderlijke dataset.

Toelichting: Wie een *bestaande* dataset als linked data ontsluit, voegt deze verschijningsvorm van hetzelfde product in de bestaande metadata-record als een extra ‘distributie’ toe. Dit is consistent met de definitie van ‘dataset’ en ‘distributie’ die door DCAT-AP gehanteerd wordt: verschillende datasets hebben een verschillende

⁵ https://joinup.ec.europa.eu/asset/dcat_application_profile/description

⁶ Conform de INSPIRE-kaderrichtlijn: <http://inspire.ec.europa.eu/>

⁷ <http://opendata.vlaanderen.be/>

⁸ <http://www.geopunt.be/>

⁹ <https://www.w3.org/TR/spargl11-query/>

inhoud, eenzelfde dataset wordt aan de hand van één of meerdere distributies / formaten / services ter beschikking gesteld.

Ook voor de ISO/INSPIRE-standaard wordt dit principe doorgetrokken. Linked open data metadateren van geografische gegevens betekent dat je deze beschrijft als een distributie-vorm van de dataset waarop dit product gebaseerd is. Je hoeft geen afzonderlijke dataset te voorzien voor het linked open dataproduct zelf.

Bovenstaande gaat zowel op voor bestaande als nieuwe datasets / producten. Echter, nieuwe datasets / producten kunnen ook gebaseerd zijn op één of meerdere bestaande datasets en een gereduceerde, uitgebreide of gecombineerde inhoud bevatten ten opzichte van de oorspronkelijke dataset(s). In dat geval ontstaat een *nieuwe* dataset / product en is de LOD-distributie die wordt aangeboden een distributie van de nieuwe dataset en niet van (één of meerdere) oorspronkelijke datasets waarvan men vertrokken is.

MD3 / OD3: Je ontsluit linked open data via alle portalen die geschikt zijn om dat soort gegevens te ontsluiten en gebruikt hiervoor de bijhorende tooling.

Toelichting: Net omdat je de aanbeveling MD1 / OD1 volgt, zullen je gegevens op alle portalen terecht komen waar de soort gegevens thuis horen. Dit betekent bijvoorbeeld dat geografische linked open data zowel op het geografische portaal Geopunt als op het Vlaamse Open data portaal (via harvesting) ontsloten worden. Voor deze datasets geldt het once only principe, zij dienen niet opnieuw van metadata te worden voorzien in het Vlaams Open data portaal.

Alsook dat de niet-geografische linked open data via het Vlaamse Open data portaal en eventueel eigen open data portalen van Vlaamse, provinciale en lokale overheden ontsloten zijn. Overheden met een eigen portaal zorgen er voor dat hun (linked) open datasets ook op het Vlaams Open data portaal terug te vinden zijn. Voor het metadateren van linked open data op deze portalen maak je gebruik van bijhorende tooling, conform standaard MD1 / OD1.

Het Vlaams Open data portaal is de gids voor alle open data van overheden in Vlaanderen. Het portaal geeft bovendien invulling aan de decretale verplichting¹⁰ om een portaal op te zetten met voor hergebruik beschikbare bestuursdocumenten.

OD4: Bij publicatie op het Vlaams Open data portaal en eigen open data portalen van Vlaamse bestuursinstanties van linked open data van geografische informatie wordt naast metadata conform GeoDCAT-AP ook de link toegevoegd van de originele beschrijving van de brondataset, conform de ISO/INSPIRE-metadatastandaard.

Toelichting: Om voor geografische linked open data ontsloten via het Vlaams Open data portaal of eigen open data portalen bij Vlaamse bestuursinstanties steeds zicht te hebben op alle relevante metadata wordt naast GeoDCAT-AP conforme metadata ook een link toegevoegd naar de volledige ISO/INSPIRE-conforme metadata-beschrijving (de oorspronkelijke metadata).

De meest eenvoudige manier om dit te doen, is door gebruik te maken van de RDF property *rdfs:seeAlso*.

OD5: Bij de publicatie van linked open data worden minimaal dereferenceable URI's aangeboden. Je moet er bovendien voor zorgen dat je aan je eindgebruikers toelaat dat ze URI's makkelijk terug kunnen vinden.

¹⁰ <https://overheid.vlaanderen.be/omzetting-psi-richtlijn>

Toelichting: Bij de publicatie van linked open data is het absolute minimum het ontsluiten van dereferenceable URI's met een RDF-representatie. Persistente URI's moeten worden opgemaakt conform de Vlaamse URI-standaard¹¹. Bovendien moet je de eindgebruiker de mogelijkheid bieden URI's makkelijk terug te vinden.

Dit kan aan de hand van een RDF-dump, SPARQL-endpoint, triple pattern fragments, reconciliation API, fulltekst search, LOD-enabled API¹², ... In dit verband zijn een aantal goede praktijken ontwikkeld binnen het OSLO²-project voor de publicatie van CRAB als linked open data.

¹¹ <https://overheid.vlaanderen.be/OSLO-URI-standaard>

¹² Een bestaande Web-API kan worden uitgebreid om Linked Data aan te bieden. Bijvoorbeeld de bestaande API van het Organisatieregister biedt naast de bestaande reST-API ook JSON-LD aan volgens de Linked Data-principes en de OSLO²-dataspecificaties. Of nog een voorbeeld, de bestaande CRAB-webservice voor adres-geocoding wordt uitgebreid met een LD-conforme interface en biedt OSLO² JSON-LD aan. Een laatste voorbeeld, een WFS-service wordt via een proxy ook als Linked Data ontsloten, zie ook <https://github.com/interactive-instruments/ldproxy>.