DCAT-AP-NL, een IPM voor Datasets

Uitwisseling van metadata tussen (open) datacatalogi

# Inhoudsopgave

1. Inhoudsopgave 2

2. Documenthistorie 4

2.1. Wijzigingshistorie 4

2.2. Planning 4

2.3. Bekende issues en todo’s in deze versie 5

3. Inleiding 6

3.1. Scope 7

3.2. Proces 8

3.3. Methodologie 8

3.3.1. Uitgangspunten 8

3.3.2. Gevolgde werkwijze 9

3.4. Gebruikscenario's 9

3.4.1. Dataportaal zoekt publiek 9

3.4.2. Gebruiker zoekt dataset 9

3.4.3. Onderzoekster zoekt over de grens 10

3.5. Terminologie in dit IPM 10

4. Het Nederlandse toepassingsprofiel voor DCAT (DCAT-AP-NL) 11

4.1. Classes 12

4.1.1. Kern-classes uit DCAT-AP-NL 12

4.1.2. Gerefereerde classes uit data.overheid.nl 13

4.1.3. Gerefereerde classes uit OWMS 13

4.2. Datatypen in DCAT-AP-NL 14

4.2.1. Datatypen voor classes uit DCAT-AP-EU 14

4.2.2. Datatype voor regelingen 16

4.3. Properties per class 16

4.3.1. Dataset 16

4.3.2. Distributie 19

4.3.3. Catalogusrecord 20

5. Identificatie en Linked Data 21

5.1. Identificatie van de Dataset 21

5.2. Identificatie van de Verstrekker: verstrekkercode 22

5.3. Identificatie van de Registratiehouder: registratiehoudercode 22

5.4. Identificatie van de Datacatalogus: cataloguscode 23

5.5. Uitwisseling van informatie over datasets: Linked Data 23

5.5.1. Zichtbaarheid van datasets op data.overheid.nl 23

5.5.2. Linked Data omgeving 24

5.5.3. Soorten verwijzingen naar dataset: semantiek 24

5.5.4. Schema.org 25

6. Referenties 26

# Documenthistorie

## Wijzigingshistorie

|  |  |  |  |
| --- | --- | --- | --- |
| Versie | Datum | Omschrijving | Auteur |
| 1.1 | 2018-02-10 | Herijking tegen DCAT-EU 1.1 | Hans Overbeek (KOOP) |

## Over deze versie

De versie van 2018-02-21 van DCAT-AP-NL 1.1 is een conceptversie die ter discussie wordt aangeboden aan de community van belanghebbenden bij de toepassing van DCAT in Nederland. DCAT-AP-NL 1.1 is het Nederlandse toepassingsprofiel op het Europese DCAT-AP-EU 1.1.

Het versienummer van het profiel blijft dus 1.1, de verschillende bewerkingen daarvan hebben geen versienummer maar een versiedatum en een status “concept”, totdat een definitieve specificatie wordt vastgesteld die de status “definitief” krijgt.

## Bekende issues en todo’s in deze versie

1. Bij het opstellen van dit IPM zijn we uitgegaan van de specificaties van DCAT-AP-EU. Daarin wordt niet altijd even consequent omgegaan met verwijzingen naar namespaces en termen uit vocabulaires. Deze onvolkomenheden zijn ook in dit IPM terecht gekomen en moeten nog worden opgelost. Voorbeeld is het gebruik van zowel de adms namespace van de EU als die van het W3C.
2. Uitbreiding OWMS: Dit IPM steunt op nieuwe begrippen die in OWMS moeten worden ondergebracht, zoals nieuwe classes, properties en waardenlijsten. Deze zijn nog niet allemaal in OWMS gedefinieerd
3. Dit IPM, en in het bijzonder het hoofdstuk over Identificatie en Linked Data moet nog worden afgestemd met andere datacatalogi.

## Planning

|  |  |
| --- | --- |
| Datum | Omschrijving |
| Ntb | Concept voorleggen aan community DCAT-NL |
| Ntb | IPM Vaststellen |
| Ntb | IPM publiceren |

# Inleiding

Dit document is opgesteld in de context van het Open Data NEXT programma[[1]](#footnote-1) van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Het beschrijft het Nederlands toepassingsprofiel van de Data Catalogue Vocabulary van het W3C (DCAT)[[2]](#footnote-2). Dit Informatie Publicatiemodel (IPM) voor datasets specificeert een standaard voor de online uitwisseling van metadata tussen Nederlandse data-catalogi, zoals: data.overheid.nl, het Nationaal Georegister (NGR) en bijvoorbeeld ook de (open) data catalogi van gemeenten en provincies.

Wat is daarvoor nodig?

Het IPM voor datasets moet rekening houden met internationale standaarden op dit gebied. Drie standaarden spelen een centrale rol:

1. DCAT **[1]** is een toonaangevende W3C standaard voor metadata van data catalogi. Binnen de EU (DG connect, ISA program en het EU publication office) is een toepassingsprofiel **[2]** voor DCAT opgesteld. Het IPM is hierop gebaseerd. ISA zegt zelf over het EU toepassingsprofiel: "The DCAT Application profile for data portals in Europe (DCAT-AP) is a specification based on the Data Catalogue vocabulary (DCAT) for describing public sector datasets in Europe. Its basic use case is to enable a cross-data portal search for data sets and make public sector data better searchable across borders and sectors. This can be achieved by the exchange of descriptions of data sets among data portals." Dit IPM heeft hetzelfde doel op nationaal niveau.
2. ISO 19115 en ISO 19119 zijn standaarden voor metadata van Geo-informatie en geo-services. Deze standaarden worden gehanteerd door het NGR volgens het Nederlandse toepassingsprofiel van ISO 19115 en ISO 19119 **[3]**.
3. OWMS 4.0, de Overheidbrede Web Metadata Standaard is het Nederlandse toepassingsprofiel van Dublin Core. Het is een standaard voor metadata van informatieobjecten van de Nederlandse overheid op internet.

Dit IPM bevat:

* Een vertaling van de Engelse labels van DCAT naar Nederlandse labels;
* Een keuze voor welke entiteiten en attributen uit het DCAT model wij overnemen;
* Een invulling van het waardebereik van de attributen (zoals bijvoorbeeld OWMS waardenlijsten voor organisaties);
* Een mapping van het metadata model van de Nederlandse metadataprofielen op ISO 19115 en ISO 19119 naar DCAT;
* Een mapping van OWMS (Dublin Core) naar DCAT.

Onderstaand figuur illustreert dit:



## Scope

Het doel van dit IPM is om een toepassingsprofielte definiëren dat gebruikt kan worden voor de uitwisseling van beschrijvingen van Datasets tussen Dataportalen in Nederland.

*IPM* staat voor **Informatie Publicatie Model**, (ook wel contentmodelgenoemd). Een IPM beschrijft de manier waarop een applicatie een bepaalde collectie van Informatieobjecten met metadata kan beschrijven zodat andere applicaties deze informatieobjecten kunnen vinden en in samenhang met andere informatie kunnen ontsluiten.

Een **Toepassingsprofiel** (Engelse term: *Application Profile*) is een specificatie die termen van een of meer onderliggende standaarden hergebruikt en meer specifiek maakt door verplichte, aanbevolen en optionele elementen aan te wijzen voor een specifiek toepassing of toepassingsgebied en aanbevelingen doet voor te gebruiken waardenlijsten.

Een **Dataset** is een verzameling (ook wel collectie) van data, gepubliceerd of beheerd door een Dataportaal en toegankelijk in een of meer formaten.

Een **Dataportaal** (Engelse term: *Data Portal*) is een systeem op het web dat services aanbiedt om Datasets te verkrijgen. Doorgaans biedt een dataportaal een Datacatalogus aan met beschrijvingen van Datasets.

Een **Datacatalogus** is een verzameling gestructureerde beschrijvingen van Datasets die gebruikers of systemen in staat stelt om datasets te vinden op basis van bepaalde kenmerken.

Het Nederlands toepassingsprofiel van DCAT (DCAT-AP-NL) dat in dit IPM wordt gespecificeerd is gebaseerd op het *DCAT Application Profile for data portals in Europe; Final version 1.00 2013-09-02***[2]** (DCAT-AP-EU) en geeft richtlijnen voor de vertaling van metadata volgens het *Nederlands metadataprofiel op ISO 19115 geografie, 1.3.1* **[3]** (ISO19115) naar DCAT-AP-NL.

Dit IPM:

* Identificeert de essentiële elementen en attributen van DCAT in de context van *Datasets* van de Nederlandse overheid.
* Identificeert de waardenlijsten die gebruikt worden in de Nederlandse context
* Identificeert de minimale set description metadata om uit te wisselen tussen Nederlandse *Dataportalen* en tussen Nederlandse en Europese *Dataportalen*.

Het IPM geeft geen specificaties voor implementatie, zoals mechanisme voor uitwisseling van data en verwacht gedrag van systemen die het IPM toepassen, anders dan hetgeen gedefinieerd in de sectie over conformering.

## Proces

Dit IPM voor Datasets is opgesteld door het *Kennis- en Exploitatiecentrum voor Officiële Overheidspublicaties* (KOOP) van het *Ministerie van Binnenlandse Zaken en Koninkrijksrelaties* (BZK) in het kader van de doorontwikkeling van Data.overheid.nl: het open dataportaal van de Nederlandse overheid. Het IPM zal worden afgestemd met diverse houders van domeinspecifieke Dataportalen.

|  |  |
| --- | --- |
| Houder | Dataportaal |
| Rijkswaterstaat | Open Data RWS |
| Centraal Bureau voor de Statistiek (CBS) | Statline |
| Geonovum/Kadaster | Nationaal Georegister (NGR) |
| KOOP | Data.overheid.nl |

***Tabel: Houders van Dataportalen die bijdragen aan dit IPM.***

## Methodologie

### Ontwerpprincipes

* DCAT-AP-NL moet zo eenvoudig mogelijk toepasbaar zijn door data-portalen, catalogi van datasets en metadata-brokers in Nederland.
* Wijk niet zonder reden af van de specificaties van DCAT en DCAT-AP-EU.
* DCAT-AP-NL volgt OWMS, dus conformiteit aan DCAT-AP-NL betekent automatisch conformiteit met OWMS.
* Het toepassingsprofiel houdt rekening met specificaties in andere standaarden voor Datasets, met name die van het Nederlands toepassingsprofiel voor Geo-data: ISO 19115.
* Op data.overheid.nl, het Open Data portaal van de Nederlandse overheid worden referentiegegevens beschikbaar gesteld. Toepassers van DCAT-AP-NL maken gebruik van deze referentiegegevens indien van toepassing maar kunnen daarvan afwijken als deze referentiegegevens niet toereikend zijn in hun situatie.
* Applicaties die conform DCAT-AP-NL zijn opgezet kunnen onderling beschrijvingen van Datasets aanbieden en ontvangen zoals beschreven in de sectie ‘Terminologie’. In combinatie met informatie op data.overheid.nl kunnen applicaties conform DCAT-AP-EU data aanbieden. Dit betekent dat bijvoorbeeld de datacatalogus op data.overheid.nl wel DCAT-AP-EU compliant kan zijn, maar dat er ook DCAT-AP-NL compliant applicaties kunnen zijn die niet DCAT-AP-EU compliant zijn.

### Gevolgde werkwijze

Bij het opstellen van DCAT-AP-NL zijn we als volgt te werk gegaan:

1. Begin met de classes en properties uit DCAT-AP-EU.
2. Deel alle classes die als datatype geïmplementeerd kunnen worden in als ‘**Datatype’**.
3. Deel alle classes die – reeds of binnenkort – als referentiegegevens op data.overheid.nl of standaarden.overheid.nl beschikbaar zijn, in als ‘**Referentiegegevens’**.
4. Bepaal van de Referentiegegevens welke waardenlijst de instanties van deze Class bevatten die voor DCAT-AP-NL gebruikt worden.
5. Splits classes op die voor verschillende ranges verschillende waardenlijsten hebben. (Bijv skos:Concept voor onder meer Thema en Status) en geef deze een eigen naam
6. Benoem de overgebleven classes van DCAT-AP-EU als ‘**Uitwisselgegevens’** in DCAT-AP-NL.
7. Bepaal van elk referentiegegeven of instanties uit de class worden aangeduid met een URI of met een code. Het datatype van een property die als range referentiegegevens heeft die met een URI worden aangeduid hebben datatype *xsd:anyURI*, als ze met een code worden aangeduid hebben ze datatype *xsd:string*.
8. Bepaal van elke property de range volgens DCAT-AP-NL.
9. Maak DCAT-AP-NL compliant aan OWMS door de OWMS-kern-eigenschappen de juiste cardinaliteit te geven en waar nodig toe te voegen.
10. Loop alle verplichte eigenschappen uit ISO 19115 na en probeer deze te mappen op DCAT-AP-EU. TODO: nalopen voor DCAT 1.1

## Gebruikscenario's

De gebruikscenario's (use cases) illustreren de beoogde toegevoegde waarde van DCAT-AP-NL. Met DCAT-AP-NL moet het eenvoudiger worden om metadata uit verschillende datacatalogi uit te wisselen binnen Nederland en – via data.overheid.nl – met Europa.

### Dataportaal zoekt publiek

Willem beheert een dataportaal met archeologische vondsten van artefacten uit de 2e Wereldoorlog. Het dataportaal biedt een zeer gedetailleerde beschrijving van de vondsten. Met name het onderscheid naar legeronderdeel en fase in de oorlog (capitulatie, bevrijding) is vergaand uitgewerkt. Het dataportaal is echter vooral bekend bij specialisten, terwijl de gegevens bijvoorbeeld relevant kunnen zijn bij een bodem-activiteit of als informatie aan toeristen. Willem wil dus zijn metadata kunnen delen met dataportalen die op deze domeinen opereren.

Zonder DCAT-AP-NL moet hij met elke portaal afspraken maken over de uit te wisselen metadata.

Met DCAT-AP-NL kan hij de meest gangbare metadata aanbieden in een standaard formaat en kan deze door zijn collega's in dat formaat worden afgenomen. Doordat iedereen veldnamen en waarde-bereik uit DCAT-AP-NL gebruikt, hoeft er minder metadata vertaald te worden, wat zowel de kwaliteit als de efficiëntie ten goede komt.

### Gebruiker zoekt dataset

Jan doet onderzoek naar de verspreiding en bestrijding van de processierups. Hij wil weten in welke leefomgeving de rups goed gedijt en welke plaatsen risico vormen op een rupsenplaag. Daarvoor zoekt hij data over de verspreiding van de rups en over landschap en vegetatie.

Zonder DCAT-AP-NL moet hij op zoek naar betrouwbare bronnen voor deze informatie en bij elke bron op zoek naar een geschikte dataset.

Met DCAT-AP-NL wordt de metadata bij deze datasets uniformer beschreven en met datacatalogi uitgewisseld. Jan kan nu bij enkele centrale datacatalogi op zoek gaan naar deze datasets. De datacatalogi geven een uniforme zoekingang en verwijzen direct door naar de bron.

### Onderzoekster zoekt over de grens

Eva uit Madrid doet onderzoek naar urbanisatie in Europa. Zij is op zoek naar data over bevolkingsopbouw en middelen van bestaan in heel Europa.

Zonder DCAT-AP-NL moet zij in elk land op zoek naar datacatalogi en op elk portaal leren hoe zij daar kan zoeken naar de data op haar onderzoeksgebied.

Met DCAT-AP-NL kan de metadata uit de Nederlandse datacatalogi uitgewisseld worden met Europese catalogi, waardoor zij in 1 taal en met 1 interface metadata uit verschillende landen kan benaderen en sneller de relevante Nederlandse data vindt.

## Terminologie in dit IPM

Deze is gebaseerd op de terminologie die beschreven wordt in Hoofdstuk 5 van DCAT-AP-EU.

In DCAT-AP-NL worden classes en properties gegroepeerd onder de kopjes ‘Mandatory’, ‘Recommended’, ‘Optional’, ‘Referentiegegeven’ en ‘Datatype’. Deze termen hebben de volgende betekenis:

* **Mandatory class**: een ontvanger van een beschrijving van een Dataset MOET informatie over instanties van de class kunnen verwerken; een aanbieder van een beschrijving van een Dataset MOET informatie over instanties van de class verstrekken.
* **Recommended class**: een ontvanger van een beschrijving van een Dataset MOET informatie over instanties van de class kunnen verwerken; een aanbieder van een beschrijving van een Dataset MOET informatie over instanties van de class verstrekken als deze beschikbaar is.
* **Optional class**: een ontvanger van een beschrijving van een Dataset MOET informatie over instanties van de class kunnen verwerken; een aanbieder van een beschrijving van een Dataset MAG informatie over instanties van de class verstrekken, maar is daartoe niet verplicht.
* **Referentiegegeven:** beschrijvingen van datasets verwijzen middels URIs of codes naar instanties van classes die als Referentiegegeven zijn bestempeld. DCAT-AP-NL verwijst naar waardenlijsten met instanties die tot deze classes behoren.
* **Datatype:** beschrijvingen van datasets bevatten de instanties van de class die als Datatype is bestempeld. DCAT-AP-NL beschrijft de syntax waar de instanties aan moeten voldoen.
* **Mandatory property**: een ontvanger van een beschrijving van een Dataset MOET informatie voor die property kunnen verwerken; een aanbieder van een beschrijving van een Dataset MOET informatie voor die property verstrekken.
* **Recommended property**: een ontvanger van een beschrijving van een Dataset MOET informatie voor die property kunnen verwerken; een aanbieder van een beschrijving van een Dataset MOET informatie voor die property verstrekken als deze beschikbaar is.
* **Optional property**: een ontvanger van een beschrijving van een Dataset MOET informatie voor die property kunnen verwerken; een aanbieder van een beschrijving van een Dataset MAG informatie voor die property verstrekken, maar is daartoe niet verplicht.

# Het Nederlandse toepassingsprofiel voor DCAT (DCAT-AP-NL)

Onderstaande figuur is het model voor DCAT-AP-NL. De URIs van de classes en properties komen overeen met de URIs van de corresponderende classes en properties in DCAT-AP-EU. De blauwe rechthoek, gelabeld “Exchange”, duidt de entiteiten aan die tussen datacatalogi worden uitgewisseld. De overige entiteiten zijn de in dit IPM beoogde referentie-entiteiten die gemeenschappelijk gebruikt worden.

[TODO] nieuw figuur maken. Gebruik voor overzicht de excel sheet *[yyyy-mm-dd] DCAT-NL-EU Model.xlsx*.

De URIs in dit IPM zijn gespecificeerd als q-names ([namespace alias]:[term]). Daarbij zijn de volgende namespace aliassen als prefix gebruikt:

| Alias | Namespace |
| --- | --- |
| adms: | http://www.w3.org/ns/adms# |
| dcat: | http://www.w3.org/ns/dcat# |
| dct: | http://purl.org/dc/terms/ |
| foaf: | http://xmlns.com/foaf/0.1/ |
| overheid: | http://standaarden.overheid.nl/owms/terms/ |
| rdfs: | http://www.w3.org/2000/01/rdf-schema# |
| schema: | http://schema.org/ |
| skos: | http://www.w3.org/2004/02/skos/core# |
| vcard: | http://www.w3.org/2006/vcard/ns# |
| xsd: | http://www.w3.org/2001/XMLSchema# |

## Classes

Het model voor DCAT-AP-NL is gebaseerd op het UML-diagram in DCAT-AP-EU **[2]**.

* Meestal heeft een Dataportaal een eigen Catalogus, maar niet altijd.
* Een Metadatabroker verzamelt metadata uit verschillende Catalogi en ontsluit deze verzameling doorgaans weer in een eigen Catalogus. De Catalogus van de Metadatabroker bevat dan beschrijvingen van de andere Catalogi en hun Catalogusrecords. Zo kunnen Datasets in meerdere Catalogi beschreven staan, wat als voordeel heeft dat ze op meerdere manieren beschreven zijn, bijvoorbeeld als statistische data en als geografische data. Gebruikers kunnen de Datasets dan vinden via de metadata die het beste past bij hun eigen domein.

DCAT-AP-NL beschrijft een minimale set metadata die alle Nederlandse Catalogi voor Datasets zouden moeten hanteren. Uiteraard staat het Catalogushouders vrij om naast de metadata in DCAT-AP-NL andere metadata te hanteren.

### Kern-classes uit DCAT-AP-NL

In het model staan de classes Dataset, Distributie en Catalogusrecord centraal. Datasets worden aangeboden op Dataportalen, zoals het NGR en de Dataportalen van gemeenten en provincies. Een Catalogusrecord in een Catalogus beschrijft een Dataset.

| **Classnaam inDCAT-AP-NL / *DCAT-AP-EU*** | **URI** | **Toelichting** |
| --- | --- | --- |
| Dataset*Dataset* | dcat:Dataset | Beschrijving van een verzameling gestructureerde gegevens, gepubliceerd door één partij en toegankelijk in één of meer formaten. |
| Distributie*Distribution* | dcat:Distribution | Fysieke implementatie van een Dataset in een bepaald formaat.Bij voorkeur bevatten de verschillende Distributies van dezelfde Dataset dezelfde data en verschilt alleen het formaat. Maar het is ook mogelijk om een Dataset te laten bestaan uit bijvoorbeeld een Distributie per periode of per regio. Zie Aanbeveling over Dataset-series. |
| Catalogusrecord*Catalogue record* | dcat:CatalogRecord | Herkomstinformatie (*provenance*) bij een beschrijving van een Dataset in een Catalogus. |

### Referentiegegevens

De volgende classes uit DCAT-AP-EU zijn in DCAT-AP-NL opgenomen als referentiegegevens. Dat wil zeggen dat de instanties uit deze classes worden opgenomen in een waardenlijst.

| **Classnaam inDCAT-AP-NL / *DCAT-AP-EU*** | **URI** | **Toelichting** |
| --- | --- | --- |
| Agent*Agent* | dcat:Agent | Persoon of organisatie die een Dataset technisch beheert en/of publiceert. (Een Catalogus is ook op te vatten als een Dataset).In DCAT-AP-NL de URI van een organisatie uit een waardenlijst met organisaties. Indien de organisatie voorkomt in de waardenlijst voor organisaties op data.overheid.nl dan moet deze  |
| Catalogus*Catalogue* | dcat:Catalog | Een verzameling (beschrijvingen van) Datasets. |
| Thema*Category* | skos:Concept | Een onderwerp van een Dataset. Een waarde uit een Rubricering. |
| Rubricering*Category scheme* | skos:ConceptScheme | Een verzameling definities van Thema’s.Het gaat dus om de aanduiding van de thema-indeling in zijn geheel. |
| Filetype*Media type or extend* | dct:MediaType-OrExtent | Bestandformaat van een Dataset.DCAT-AP-EU benoemt maar 1 class *Media type or extend* en gebruikt die zowel voor Filetype als Mediatype. |
| Mediatype*Media type or extend* | dct:MediaType-OrExtent | Mediatype van een Dataset.DCAT-AP-EU benoemt maar 1 class *Media type or extend* en gebruikt die zowel voor Filetype als Mediatype. |
| Frequentie*Frequency* | dct:Frequency  | Het aantal maal per periode dat een bepaalde gebeurtenis zich voordoet. |
| Licentie*Licence document* | dct:LicenseDocument | Een juridisch document dat officieel toestemming geeft tot bepaald gebruik van een resource.In DCAT-AP-NL een URL van een web-document. |
| Licentietype*Licence type* | skos:Concept | Indeling van Licenties in de mate van openbaarheid. |
| Taal*Linguistic system* | dct:LinguisticSystem | Een systeem van tekens, symbolen, geluiden, gebaren of regels, gebruikt voor communicatie, bijvoorbeeld een taal. |
| Locatie*Location* | dct:Location | Een plaats op aarde. Deze kan direct worden aangeduid met coördinaten of indirect met een identifier (naam of code). |
| Soort verstrekker*Publisher type* | skos:Concept | Type Agent dat optreedt als verstrekker. |
| Changetype*N/A* | skos:Concept | Een indicatie van de soort van de laatste wijziging van een Catalogusrecord.Geen class in DCAT-AP-EU, maar in toelichting bij change type van *Catalogue Record* worden *hard coded* drie waarden voorgeschreven: “:created”, “:updated”, “:deleted”. |
| Status*Status* | skos:Concept | Een indicatie van de levensfase van een Dataset.Op basis van adms:status (http://purl.org/adms/status/1.0)CompletedDeprecatedUnderDevelopmentWithdrawn |
| openbaarheidsniveau*rights statement* | dct:RightsStatement | Een indicatie van het niveau van openbaarheid. |

## Datatypen in DCAT-AP-NL

### Datatypen voor classes uit DCAT-AP-EU

De volgende classes uit DCAT-AP-EU zijn in DCAT-AP-NL geïmplementeerd als vrije tekst of datatype en niet als Class.

| **Classnaam inDCAT-AP-NL / *DCAT-AP-EU*** | **URI inDCAT-AP-EU** | **Toelichting** | **Datatype in DCAT-AP-NL** |
| --- | --- | --- | --- |
| Literal*Literal* | rdfs:Literal | Een tekenreeks. Een literal met voor de mens leesbare tekst kan een taal-code hebben. In DCAT-AP-NL wordt dit vrije tekst | xsd:string |
| Datum*Literal as a date* | rdfs:Literal | Een tekenreeks met datum-tijd-formaat volgens W3CDTF: “*yyyy-mm-ddThh:mi:ss*” | xsd:dateTime |
| Decimal*Literal as a decimal* | rdfs:Literal | Een tekenreeks met getalformaat. | xsd:decimal |
| Resource*Resource* | rdfs:Resource | Alles wat met RDF beschreven wordt.In DCAT-AP-NL wordt dit de URI van de resource. | xsd:anyURI |
| Document*Document* | Foaf:Document | Informatie voor menselijke consumptie.In DCAT-AP-NL is dit een URL van een web-document. | xsd:anyURI |
| Identificatie*Identifier* | adms:Identifier | Identificatie in een bepaalde context, bestaande uit:* De referentie
* Het gebruikte schema
* Schema-versie
* Schema-beheerder

In DCAT-AP-EU gebruikt voor samengestelde identifiers. DCAT-AP-NL ondersteunt alleen URI. TODO: bepalen wat we hiermee willen. | xsd:anyURI ???? |
| Periode*Period of time* | dct: PeriodOfTime | Een tijdsinterval.DCAT-AP-EU ondersteunt benoemde tijdsintervallen en tijdsintervallen aangeduid met start- en einddatum. DCAT-AP-NL alleen de laatste. | Complex type Periode |
| Rechtenclaim*Rights statement* | dct: RightsStatement | Een statement over de intellectuele eigendomsrechten (IPR) die rusten op een resource.In DCAT-AP-EU een zelfstandige entiteit. In DCAT-AP-NL geïmplementeerd als URL die kan verwijzen naar een pagina met een beschrijving van de rechten die van toepassing zijn op de distributie. | xsd:anyURI |
| Standaard*Standard* | dct:Standard | Specificatie waar een Dataset aan voldoet.In DCAT-AP-EU een zelfstandige entiteit. In DCAT-AP-NL geïmplementeerd als URL die kan verwijzen naar een pagina met een beschrijving van de standaard waar de distributie aan voldoet. | xsd:anyURI |
| Contactgegevens*Kind* | vcard:Kind | Een beschrijving volgens de vCard specificatie. |  |
| Regeling*N/A* | overheid:Regeling | Link naar wet- of regelgeving. | Complex type Regeling |

## Properties per class

### Dataset

| **Man** | **Property** | **URI** | **Toelichting** |  **Card** | **Range** |
| --- | --- | --- | --- | --- | --- |
| Man | global identifier*identifier* | dct: identifier | Uniforme identificatie van de dataset.Verplicht vanwege het grote belang voor ontdubbelen.Globaal unieke identificatie van de dataset, toegekend door de eerste catalogus waarin de dataset wordt beschreven. Op basis van dedze identifier worden duplicaten opgespoord. (Zie https://joinup.ec.europa.eu/release/dcat-ap-how-manage-duplicates).Range is niet URI maar vrije tekst omdat het hier om de letters van string gaat (de literal) en niet om de link.Zie de sectie over Identificatie elders in dit document. | 1..1 | Type – Vrije tekst |
| Man | titel*title* | dct:title | Naam waaronder de dataset bekend staat. Herhaalbaar voor versies in verschillende talen. | 1..n | Type – Vrije tekst |
| Man | omschrijving*description* | dct: description | Beschrijving van de dataset op basis waarvan een gebruiker kan bepalen of de dataset bruikbaar is en die de dataset onderscheidt van andere datasets. Herhaalbaar voor versies in verschillende talen | 1..n | Type – Vrije tekst |
| Man | wijzigingsdatum*update/ modification date* | dct: modified | Laatste datum waarop de dataset zelf is gewijzigd. Deze datum zegt niets over de wijzigingsdatum van de metadata van de dataset.In DCAT-AP-NL verplicht vanwege OWMS. | 1..1 | Type – xsd:date |
| Man | taal*language* | dct: language | Taal van de inhoud van de dataset.In DCAT-AP-NL verplicht vanwege OWMS. Herhaalbaar indien meerdere talen voorkomen. | 1..n | Type – xml:lang |
| Rec | contact*contact point* | dcat: contactPoint | Contactinformatie voor informatie over de dataset | 0..n | Type – vcard:Kind |
| Rec | registratiehouder*N/A* | overheid: authority | De organisatie die eindverantwoordelijk is voor de inhoud van de dataset. De registratiehouder is ook verantwoordelijk voor de relatie met de verstrekker.Toegevoegd aan DCAT-AP-NL vanwege OWMS. | 0..1 | Waardelijst – overheid: Agent |
| Rec | verstrekker*publisher* | dct:publisher | Persoon of organisatie die verantwoordelijk is voor het technisch beheer en publicatie van de dataset. | 0..1 | Waardelijst – overheid: Agent |
| Rec | trefwoord*keyword/tag* | dcat:keyword | Trefwoord dat de dataset karakteriseert. | 0..n | Type – Vrije tekst |
| Rec | thema*theme/category* | dcat:theme | Thema waar de Dataset onder valt. | 0..n | Waardenlijst – overheid: Taxonomie­Beleids­agenda |
| Rec | landings­pagina*landing page* | dcat: landingPage | Webpagina, meestal bij de verstrekker, die toelichting geeft bij de dataset, de distributies en metadata die de verstrekker geeft.In DCAT-AP-ANL beperkt tot maximaal 1 en recommended | 0..1 | Type – URL |
| Rec | locatie*spatial/ geographical coverage* | dct:spatial | Geografisch gebied waarop de dataset betrekking heeft.In DCAT-AP-NL recommended vanwege OWMS. | 0..n | Type – overheid: Spatial |
| Rec | dekking in tijd*temporal* | dct:temporal | Periode in de tijd waarop de dataset betrekking heeft.In DCAT-AP-NL recommended vanwege OWMS. | 0..n | Complex type – Periode |
| Opt | openbaarheid*access rights* | dct: accessRights | De mate waarin een dataset openbaar beschikbaar is. | 0..1 | Waardenlijst – overheid:Openbaarheidsniveau |
| Opt | conform*conforms to* | dct: conforms­To | Standaard waar de dataset aan voldoet. | 0..n | Type – Vrije tekst |
| Opt | documentatie*documentation* | foaf:page | Een document of pagina over de Dataset. | 0..n | Type - URL |
| Opt | wijzigings-frequentie*frequency* | dct: accrual­Periodicity | Frequentie waarmee de dataset wordt gewijzigd. | 0..1 | Waardenlijst – overheid: Frequentie |
| Opt | heeft versie*has version* | dct: hasVersion | Verwijzing naar een Dataset die een versie is van deze dataset | 0..n | Waardenlijst – donl:Dataset |
| Opt | is versie van*is version of* | dct: isVersionOf | Verwijzing naar een Dataset waar deze dataset een versie van is | 0..n | Waardenlijst – donl:Dataset |
| Opt | alternatieve identificatie*frequency* | adms: identifier | Secundaire identifier waarmee de Dataset wordt aangeduid | 0..n | Complex type - adms-identifier |
| Opt | totstandkoming*provenance* | dct: provenance | Link naar een beschrijving over de totstandkoming van de dataset | 0..n | Type – URL |
| Opt | heeft relatie met*related resource* | dct:relation | Link naar een gerelateerde resource | 0..n | Type – URL |
| Opt | uitgiftedatum*release date* | dct:issued | Datum waarop de dataset formeel is uitgegeven. | 0..1 | Type – xsd:date |
| Opt | voorbeeld*sample* | adms:sample | Voorbeelddata | 0..n | Type – URL |
| Opt | brongegevens*source* | dct:source | Dataset waaruit deze dataset is afgeleid | 0..n | Type – URL |
| Opt | type*type* | dct:type | Nadere aanduiding van het soort dataset | 0..1 | Type – xsd:anyURI |
| Opt | versie*version* | owl: versionInfo | Versieaanduiding. | 0..1 | Type – Vrije tekst |
| Opt | versie­toelichting*version notes* | adms: versionNotes | Toelichting bij de versieaanduiding.Herhaalbaar voor versies in verschillende talen | 0..n | Type – Vrije tekst |
| Opt | grondslag*N/A* | overheid: grondslag | Link naar wettelijke grondslag op basis waarvan de dataset is opgesteld. | 0..n | Type – overheid: Regeling |

### Distributie

| **Man** | **Property** | **URI** | **Toelichting** | **Card** | **Datatype** |
| --- | --- | --- | --- | --- | --- |
| Man | toegangs-URL*access URL* | dcat:accessURL  | Webpagina, meestal bij de verstrekker, die toegang geeft tot een distributie. De pagina kan – naast links naar de distributie – ook nadere informatie over de distributie bevatten. | 1..n | Type – URL |
| Rec | omschrijving*description* | dct: description  | Beschrijving van de distributie op basis waarvan een gebruiker kan bepalen of de distributie bruikbaar is en die de distributie onderscheidt van andere distributies.Herhaalbaar voor versies in verschillende talen | 0..n | Type – Vrije tekst |
| Rec | formaat*format* | dct:format  | Fileformaat van de distributie. | 0..1 | Type - MDR File type NAL |
| Rec | licentie*licence* | dct:license  | Licentie waaronder de distributie beschikbaar wordt gesteld. | 0..1 | Waardenlijst –overheid: Licentie |
| Opt | grootte*byte size* | dcat:byteSize | Omvang van de distributie in bytes. | 0..1 | Type – Getal |
| Opt | checksum*checksum* | spdx:checksum | Verificatiemiddel om te bepalen of de inhoud van de distributie is gewijzigd. | 0..1 | Type – Vrije tekst |
| Opt | documentatie*documentation* | foaf:page | Een document of pagina over de distributie. | 0..n | Type - URL |
| Opt | download-URL*download URL* | dcat:downloadURL | Directe link naar het downloadbare bestand in het gegeven formaat. | 0.n | Type – URL |
| Opt | taal*language* | dct: language | Taal van de inhoud van de dataset. | 0..n | Type – xml:lang |
| Opt | voldoet aan*linked schemas* | dct: conformsTo | Standaard waaraan de distributie voldoet | 0..n | Type – Vrije tekst of URL |
| Opt | media type*media type* | dcat: mediaType | Mediatype volgens de IANA media type register | 0..1 | Waardenlijst – IANA Media type |
| Opt | uitgiftedatum*release date* | dct:issued | Datum waarop de distributie formeel beschikbaar is gesteld. | 0..1 | Type – xsd:date |
| Opt | rechten*rights* | dct:rights | De rechten die rusten op een distributie | 0..1 | Type – Vrije tekst of URL |
| Opt | status*status* | adms:status | Levensfase van de distributie (afgerond, in ontwikkeling, niet langer ondersteund, teruggetrokken). Let op: Dit is iets anders dan de adms:status van het catalogusrecord. | 0..1 | Waardenlijst – adms: Distributie­status |
| Opt | titel*title* | dct:title | Naam waaronder de distributie bekend staat. Herhaalbaar voor versies in verschillende talen. | 0..n | Type – Vrije tekst |
| Opt | wijzigingsdatum*update/ modification date* | dct:modified | Laatste datum waarop de distributie zelf is gewijzigd. Deze datum zegt niets over de wijzigingsdatum van de metadata van de distributie. | 0..1 | Type – xsd:date |

### Catalogusrecord

| **Man** | **Property** | **URI** | **Toelichting** | **Card** | **Datatype** |
| --- | --- | --- | --- | --- | --- |
| Man | dataset*primary topic* | foaf:primary­Topic | Identifier van de dataset waar dit catalogusrecord betrekking op heeft. | 1..1 | Waardenlijst –dcat:Dataset |
| Man | wijzigingsdatum*update/ modification date* | dct:modified | Laatste datum waarop het catalogusrecord zelf , dus de metadata, is gewijzigd. Deze datum zegt niets over de wijzigingsdatum van de dataset. | 1..1 | Type – xsd:date |
| Rec | voldoet aan*application profile* | dct: conformsTo | Standaard waaraan de metadata van de dataset voldoet. (Bijvoorbeeld DCAT-AP-NL) | 0..1 | Type – Vrije tekst of URL |
| Rec | md-soort wijziging*change type* | adms:status | Indicatie van de soort van de laatste wijziging: :created, :updated of :deleted. Let op: Dit is iets anders dan de adms:status van de distributie. | 0..1 | Type – adms: changetype |
| Rec | uitgiftedatum*issued* | dct:issued | Datum waarop de beschrijving van de dataset voor het eerst in de catalogus is opgenomen. | 0..1 | Type – xsd:date |
| Opt | omschrijving*description* | dct:description  | Omschrijving van het catalogusrecord. Herhaalbaar voor versies in verschillende talen | 0..n | Type – Vrije tekst |
| Opt | taal*language* | dct:language | Taal van de metadata | 0..n | Type – xml:lang |
| Opt | brongegevens*source metadata* | dct:source | Verwijzing naar de originele metadata waaruit deze metadata is afgeleid | 0..1 | Type – URL |
| Opt | titel*title* | dct:title  | Naam voor het catalogusrecord.Herhaalbaar voor versies in verschillende talen | 0..n | Type – Vrije tekst |

### Periode

| **Man** | **Property** | **URI** | **Toelichting** | **Card** | **Datatype** |
| --- | --- | --- | --- | --- | --- |
| Rec | startdatum*start date* | schema: startDate | Eerste datum van de periode | 0..1 | Type – xsd:dateTime |
| Rec | einddatum*end date* | schema: endDate | Laatste datum van de periode | 0..1 | Type – xsd:dateTime |

### Contactgegevens

In DCAT-AP-EU wordt de vCard class “Kind” genoemd, maar er worden geen properties gegeven. In DCAT-AP-NL worden hier wel keuzes in gemaakt.

| **Man** | **Property** | **URI** | **Toelichting** | **Card** | **Datatype** |
| --- | --- | --- | --- | --- | --- |
| Man | emailadres*N/A* | vcard: hasEmail | Email adres voor contact | 1..1 | Type – Vrije tekst |
| Rec | postadres *N/A* | vcard: Address | Postadres voor contact | 0..1 | Type – Vrije tekst |
| Rec | naam *N/A* | vcard:fn | Naam van contact | 0..1 | Type – Vrije tekst |
| Opt | telefoonnummer*N/A* | vcard: hasTelephone | Telefoonnummer voor contact | 0..1 | Type – Vrije tekst |
| Opt | website*N/A* | vcard:hasURL | Website voor contact | 0..1 | Type – Vrije tekst |
| Opt | titel*N/A* | dcterms:title | Titel van contact | 0..1 | Type – Vrije tekst |

### Regeling

| **Man** | **Property** | **URI** | **Toelichting** | **Card** | **Datatype** |
| --- | --- | --- | --- | --- | --- |
| Man | juriconnect-verwijzing*N/A* | overheid: juriconnect­verwijzing | Logische verwijzing naar een regeling conform een Juriconnect standaard | 1..1 | Type – Vrije tekst |
| Rec | url*N/A* | foaf:page | Link naar een publicatie van de regeling. Bijvoorbeeld naar een bekendmaking in het Staatsblad of Staatscourant of naar een geconsolideerde regeling in het Basiswettenbestand (BWB) | 0..1 | Type – URL |
| Rec | linktekst*N/A* | overheid: linktekst | Voor de mens leesbare tekst die gebruikt wordt om de link leesbaar te maken.Bijvoorbeeld een tekstuele verwijzing naar een artikel | 0..1 | Type – Vrije tekst |

### Agent

| **Man** | **Property** | **URI** | **Toelichting** | **Card** | **Datatype** |
| --- | --- | --- | --- | --- | --- |
| Man | naam*name* | foaf:name | Naam waaronder de agent bekend is | 1..1 | Type – Vrije tekst |
| Rec | type*type* | dct:type | Soort organisatie | 0..n | Waardenlijst – adms-publisher type |

# Identificatie en Linked Data

Om te kunnen refereren aan datasets, verstrekkers, registratiehouders, catalogi en andere entiteiten (resources) en metadata over deze resources met andere partijen te kunnen uitwisselen is het prettig om standaard identificatie van deze resources af te spreken. Met deze afspraak sluiten we aan bij de URI-strategie van het Platform Linked Data Nederland (PLDN)[[3]](#footnote-3).

Data.overheid.nl publiceert de URIs van datasets, verstrekkers, registratiehouders en datacatalogi die als referentiedata zouden kunnen dienen voor andere datacatalogi.

**Hierover zijn nog geen afspraken gemaakt. De identifiers zijn ook nog niet geïmplementeerd zoals hier beschreven. Dit hoofdstuk is bedoeld als discussiestuk.**

## Identificatie van de Dataset

Elke dataset krijgt een unieke, duurzame ID. De ID moet duurzaam zijn in de zin dat de ID behouden blijft bij het updaten van informatie over de dataset of het toevoegen van nieuwe versies van links naar onderliggende datasets.

Een ID van een dataset is als volgt opgebouwd:

{ID dataset} = DSET:{verstrekkercode}:{invulling door verstrekker}

De verstrekkercode is de code die door data.overheid.nl wordt toegekend aan organisaties die datasets publiceren en vervolgens aanmelden bij data.overheid.nl of bij een catalogus die metadata uitwisselt met data.overheid.nl. Zodra de verstrekkercode bekend is, kan een verstrekker zelf zijn/haar eigen IDs voor datasets genereren.

Voorbeeld:

Stel dat Rijkswaterstaat de verstrekkercode ‘RWS’ krijgt en elke dataset een nummer van achtcijfers geeft. Een voorbeeld ID voor een Dataset van Rijkswaterstaat zou dan zijn: “DSET:RWS:56838902”.

Een ID voor een CBS dataset zou kunnen zijn: “DSET:CBS:STATLINE4053\_50qw90”.

En een ID voor een Dataset van het Ministerie van BZK zou kunnen zijn: “DSET:BZK:2014-salarisschalen”.

De URI-strategie van PLDN dringt aan op het gebruik van http-URIs. Een http-URI is ‘resolvable’ (je kunt er op klikken). Van een Dataset-ID kan eenvoudig een http-URI gemaakt worden door er het domein van een datacatalogus voor te zetten, bijvoorbeeld:

http://data.overheid.nl/id/dataset/

Een datacatalogus moet een aanroep ondersteunen in het volgende formaat:

{http-URI dataset} = {domein datacatalogus}/{ID dataset}

Bijvoorbeeld:

“http://data.overheid.nl/id/dataset/DSET:RWS:56838902/”

Deze URI wordt met een http-303 geredirect naar de pagina met metadata over de dataset in het data.overheid.nl portaal[[4]](#footnote-4). Deze pagina met metadata heeft een doc-URI:

“http://data.overheid.nl/doc/dataset/DSET:RWS:56838902/”

Gebruik van de ID en http-URI buiten het portaal.

* het moet mogelijk zijn dat bijvoorbeeld Rijkswaterstaat op haar website bij elke dataset een link maakt naar de beschrijving van de dataset op data.overheid.nl. De link wordt gevormd door de http-URI. Deze kan door RWS zelf worden gegenereerd conform bovengenoemd schema.
* Ook ontwikkelaars, portalen en andere hergebruikers van datasets kunnen op deze manier verwijzen naar de dataset. Zij kunnen hierbij kiezen om de http-URI of alleen de ID te gebruiken en zelf in een resolver te voorzien. Dit zal bijvoorbeeld voor portalen interessant zijn die discussie en co-creatie faciliteren waarbij meerdere datasets worden ingezet.

Eventuele ID’s die van andere portalen worden binnengehaald (zoals NGR) blijven behouden en verbonden naar de betreffende datasets. De binnengehaalde ID’s worden de ID’s van een dataset. We gaan er hierbij vanuit dat een organisatie als Rijkswaterstaat een ID doorgeeft aan NGR, zodat deze uniek geïdentificeerd blijft. Rijkswaterstaat geeft in bovenstaand voorbeeld dan DSET:RWS:56838902 door als ID van hun dataset aan NGR.

Indien NGR geen ID ontvangt van een aanbieder, maakt zij zelf een ID aan (UUID). Deze UUID wordt doorgegeven aan data.overheid.nl. De ID en de http-uri van de betreffende dataset op data.overheid.nl is in dit geval:

* de ID (DSET:NGR:44bfd8c9-8b99-4bce-8161-ef4b02324362)
* de http-URI (http://data.overheid.nl/id/dataset/DSET:NGR:44bfd8c9-8b99-4bce-8161-ef4b02324362)

## Identificatie van de Verstrekker: verstrekkercode

Elke overheidsorganisatie die datasets publiceert kan een verstrekkercode krijgen. Om te voorkomen dat dezelfde verstrekkercode aan meerdere organisaties wordt toegekend is een registratie van verstrekkercodes nodig. Data.overheid.nl kent verstrekkercodes toe en beheert een registratie voor de verstrekkercodes. Het streven is om deze verstrekkercodes pragmatisch en dus kort te houden.

Ook verstrekkers krijgen een http-URI op data.overheid.nl. Deze URI wordt gevormd door voor de verstrekkercode de volgende string te plaatsen:

http://data.overheid.nl/id/verstrekker/

De http-URI van CBS wordt dan bijvoorbeeld:

http://data.overheid.nl/id/verstrekker/CBS

Deze URI wordt met een http-303 geredirect naar de pagina met metadata over de verstrekker in het data.overheid.nl portaal. Deze pagina met metadata heeft een doc-URI:

http://data.overheid.nl/doc/verstrekker/CBS

## Identificatie van de Registratiehouder: registratiehoudercode

Een organisatie die eindverantwoordelijkheid draagt is voor het tot stand komen en de inhoud van de dataset noemen we een registratiehouder. Doorgaans is dit de opdrachtgever voor de verstrekker om de dataset samen te stellen en te publiceren. Ook registratiehouders krijgen een http-URI op data.overheid.nl. Deze URI wordt gevormd door voor de registratiehoudercode de volgende string te plaatsen:

http://data.overheid.nl/id/registratiehouder/

De http-URI van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties wordt dan bijvoorbeeld:

http://data.overheid.nl/id/registratiehouder/MINBZK

Deze URI wordt met een http-303 geredirect naar de pagina met metadata over de verstrekker in het data.overheid.nl portaal. Deze pagina met metadata heeft een doc-URI:

http://data.overheid.nl/doc/registratiehouder/MINBZK

## Identificatie van de Datacatalogus: cataloguscode

Elke catalogus die metadata over datasets publiceert kan een cataloguscode krijgen. Om te voorkomen dat een cataloguscode aan meerdere catalogi wordt toegekend is een registratie van cataloguscodes nodig. Data.overheid.nl onderhoudt een registratie van datacatalogi en geeft elke datacatalogus een cataloguscode. Het streven is om deze cataloguscodes pragmatisch en dus kort te houden.

Ook datacatalogi krijgen een http-URI op data.overheid.nl. Deze URI wordt gevormd door voor de cataloguscode de volgende string te plaatsen:

http://data.overheid.nl/id/catalogus/

De http-URI van het Nationaal Georegister (NGR) wordt dan bijvoorbeeld:

http://data.overheid.nl/id/catalogus/NGR

Deze URI wordt met een http-303 geredirect naar de pagina met metadata over de catalogus in het data.overheid.nl portaal. Deze pagina met metadata heeft een doc-URI:

http://data.overheid.nl/doc/catalogus/NGR

## Uitwisseling van informatie over datasets: Linked Data

De uniforme ID en http-URI per dataset, verstrekker en catalogus maakt het mogelijk om duurzaam naar deze resources te verwijzen.

Voor portals, community sites en andere platformen voor hergebruikers en ontwikkelaars is het interessant om gestandaardiseerd naar de gebruikte datasets te verwijzen die de basis vormen van een toepassing of waarover gediscussieerd wordt. Zo kan ondubbelzinnig worden aangegeven welke dataset wordt hergebruikt of bediscussieerd. Het bij elkaar brengen van vraag en aanbod (hergebruikers en verstrekkers) begint bij het goed begrijpen van de beschikbare data. Het helpt verstrekker en hergebruiker om vragen en antwoorden over de datasets uit te wisselen. Hoe meer duidelijkheid een hergebruiker krijgt over de context, toepassingsmogelijkheid en samenhang van een dataset, des te groter is de kans op toegevoegde waarde uit de dataset.

Begrip van de datasets is de basis om co-creatie, toepassingen en uiteindelijke toegevoegde maatschappelijke en economische waarde te genereren.

### Zichtbaarheid van datasets op data.overheid.nl

De rol van data.overheid.nl bestaat vooral uit het zichtbaar maken, waar meer informatie over een specifieke dataset kan worden gevonden, zowel voor verstrekker als voor de hergebruiker. Op deze manier kan de verstrekker uitvinden waar hij/zij zich in een discussie kan mengen, en vergelijkbaar kunnen hergebruikers uitvinden waar al kennis over een dataset te vinden is.

Voor de verstrekker biedt dit de mogelijkheid om vragen over de dataset eenmalig te beantwoorden, en eventueel met nieuw inzicht de beschikbaarheid en/of de beschrijving van de dataset aan te passen.

Data.overheid.nl zal een actieve rol nemen om te achterhalen waar datasets van de Nederlandse overheid worden gebruikt, bediscussieerd en toegepast; maar alleen als de hergebruiker dit wenst. Alleen wanneer de hergebruiker expliciet verwijst naar een dataset (met de duurzame ID) EN wanneer de hergebruiker dit doet op een site, community of platform dat bekend is bij data.overheid.nl, kan de link worden gelegd tussen dataset en toepassing/discussie.

### Linked Data omgeving

Om de links naar datasets op sites te herkennen, zullen deze sites actief moeten worden “geharvest”. De overheid kan bekende sites nalopen of er links naar datasets worden genoemd en deze links vervolgens opnemen.

In het project Linked Data Overheid doet de overheid dit al voor links naar formele overheidsinformatie zoals wet- en regelgeving, jurisprudentie of beleid. Het zoeken naar verwijzingen naar datasets lijkt veel op het zoeken naar links naar formele overheidsinformatie. Ook de links naar formele overheidsinformatie worden op gestandaardiseerde wijze gemaakt en zijn daardoor herkenbaar.

Om de links naar datasets te verzamelen zal de Linked Data Overheid omgeving worden ingezet. De omgeving biedt al voorzieningen om links te verzamelen, te herkennen en vervolgens de links op te vragen. Aan de Linked Data Overheid omgeving kan op elk moment worden gevraagd welke “links” er wijzen naar een specifieke dataset.

Indien een dataset wordt opgevraagd op data.overheid.nl, kan tegelijkertijd bij LiDO worden opgevraagd welke gerelateerde links er zijn bij de dataset. De relaties hoeven niet alleen links naar sites te zijn waar de dataset zelf wordt genoemd. De dataset bevat mogelijk ook relaties naar wet- en regelgeving die de grondslag vormde voor het verzamelen van de data, of door beleidsmakers of de rechter is verwezen naar een dataset. Deze relaties worden verzameld door LiDO en kunnen worden geleverd op basis van de ID van een dataset.

LiDO biedt diverse services en widgets die vanuit data.overheid.nl kunnen worden aangeroepen om de gerelateerde links te tonen. Een voorbeeld van zo’n lijstje:



### Soorten verwijzingen naar dataset: semantiek

Een mogelijkheid die nog moet worden onderzocht op haalbaarheid, is het toekennen van “semantiek” aan de relaties naar datasets. Het soort gebruik van de dataset zou bijvoorbeeld kunnen worden vastgelegd door de hergebruiker bij het linken:

Linken naar een dataset:

1. Toepassing : de dataset wordt toegepast [dset-id]
2. Discussie : gesprek over de dataset [dset-id]
3. Gerelateerd: een verwijzing naar een gerelateerde dataset [dset-id]

Linken vanuit een dataset:

1. Grondslag: een verwijzing naar de regeling [juriconnect] op basis waarvan de dataset is gecreëerd.
2. Gerelateerd: een verwijzing naar een gerelateerde dataset [dset-id], bijvoorbeeld met vergelijkbare gegevens, of een vorige versie.

Het vastleggen van de betekenis van de link naar een dataset kan met een microformat:

### Schema.org

Schema.org is een metadata-standaard die door de grote zoekdiensten als Google, Bing en Yahoo wordt gebruikt om Linked Data te harvesten. We zouden kunnen onderzoeken of het voordelen biedt om de metadata van datacatalogi ook in schema.org beschikbaar te stellen. Vooralsnog is er geen mapping van DCAT naar schema.org bekend en zien we onvoldoende aanleiding om hier zelf veel tijd aan te besteden.

# Praktische aanbevelingen

In de praktijk bestaat er natuurlijk een breed spectrum van soorten datasets die allemaal hun eigen bijzonderheden kennen. Zowel DCAT, alsook DCAT-AP-EU en DCAT-AP-NL zijn bewust zo eenvoudig mogelijk gehouden om de implementatie en de interoperabiliteit te verggroten. Dus zijn er verschillende situaties voorkomen waar enige creativiteit gevraagd wordt om een bruikbare vertaling naar het DCAT-model te maken. Dit hoofdstuk probeert voor een aantal van die situaties aanbevelingen te doen.

De aanbevelingen gaan uit van die van DCAT-AP-EU op <https://joinup.ec.europa.eu/solution/dcat-application-profile-implementation-guidelines>.

## Datasetseries

<https://joinup.ec.europa.eu/discussion/mi2-dataset-series>.

# Referenties

**[1]** **Data Catalog Vocabulary (DCAT);
W3C Recommendation 16 January 2014:**<http://www.w3.org/TR/2014/REC-vocab-dcat-20140116/>.

**[2] DCAT Application Profile for data portals in Europe;
Final version 1.00 2013-09-02:** <https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-application-profile-data-portals-europe-final>.

**[3] Nederlands metadataprofiel op ISO 19115 geografie, 1.3.1:**<http://www.geonovum.nl/wegwijzer/standaarden/nederlands-metadataprofiel-op-iso-19115-geografie-131>.

**[4] Waardenlijsten op standaarden.overheid.nl:**[standaarden.overheid.nl/owms/4.0/doc/waardelijsten](file:///C%3A%5CUsers%5CGebruiker%5CDownloads%5Cstandaarden.overheid.nl%5Cowms%5C4.0%5Cdoc%5Cwaardelijsten).

1. Open Data NEXT programma: <https://data.overheid.nl/opendataNEXT> [↑](#footnote-ref-1)
2. W3C. Data Catalogue Vocabulary (DCAT). W3C Recommendation, 16 January 2014 <http://www.w3.org/TR/2014/REC-vocab-dcat-20140116/> [↑](#footnote-ref-2)
3. PLDN; Zie: [pilod.nl](http://pilod.nl/) [↑](#footnote-ref-3)
4. Zie: [Cool URIs for the Semantic Web](http://www.w3.org/TR/cooluris/#r303gendocument) [↑](#footnote-ref-4)