

/// OSLO Metadatastandaard voor services: Verslag Thematische werkgroep 2



Datum: 13/10/2020

Locatie: 📍 Teams meeting (virtueel)



De **presentatie** is terug te vinden via <https://data.vlaanderen.be/standaarden/standaarden-in-ontwikkeling/metadata-voor-services/index.html>

De **opname van de werkgroep** is terug te vinden via <https://web.microsoftstream.com/video/767e6913-ff89-4e98-930e-42c6e7fe70c5>

AANWEZIGEN

- Informatie Vlaanderen
 - Bert Van Nuffelen
 - Geraldine Nolf
 - Eveline Vlassenroot
 - Martin Vanbrabant
 - Loes Deventer
 - Frédéric Hennequin
 - Sarah Blomme
 - Dirk De Baere
 - Henk Vanderstraeten
 - Eline Deweirdt
 - Barbara Van Broeckhoven
 - Bart Cosyn
 - Tom Van Haute
 - Maximme Durenne
- Rijksinstituut voor Volksgezondheid en Milieu (Nederland)
 - Cor Melse
- GIM
 - Gustaaf Vandeboel
 - Mathieu Chaussier
- MAGDA
 - Bert Van Kets
- CipaI-Schaubroeck
 - Willem Van Hoecke
- Vlaams Instituut voor de Zee (VLIZ)

- Laurian Van Maldeghem
- Departement Omgeving
 - Marleen Van Damme
 - Tom Van Gulck
 - Geert Van Haute
- De Lijn
 - Tim Coninx
- District09
 - Matthias Ghijs
- KULeuven
 - Pieter Fannes
- Fluvius
 - Dries Vlaeminck
- Universiteit Gent
 - Pieter Heyvaert

AGENDA VAN DE WORKSHOP

9u05 – 9u15	Welkom en context
9u15 – 10u15	Discussie (deel 1)
10u15 - 10u30	Pauze
10u30 - 11u30	Discussie (deel 2)
11u30 - 11u50	Demo Geonetwork
11u50 – 12u	Volgende stappen

1. AANLEIDING EN CONTEXT

We verwijzen naar de slides voor meer informatie.

Het initiatief voor dit standaardisatietraject komt vanuit Informatie Vlaanderen waarbij de doelstelling is om een semantisch model op te stellen voor het uitwisselen van metadata over diensten (services) dat is afgestemd met alle betrokken partijen. Dit traject zal de huidige OSLO-standaard DCAT-AP uitbreiden met de mogelijkheid om ook API's te beschrijven.

Het doel is tweeledig:

Als eerste willen we de **metadata van services te standaardiseren** vanuit het perspectief van de beheerder van een informatiesysteem in samenspraak met alle belanghebbenden. Vandaag wordt deze informatie in verschillende vormen en niveaus van kwaliteit voorzien, sterk afhankelijk van de gebruikte technologie en platform om deze datasets en services aan te bieden. Er is nood aan een overzicht.

Op basis van internationale standaarden (DCAT-AP v2.0) bouwen we het vocabularium op dat een uitbreiding zal zijn van de OSLO-standaard DCAT-AP met de mogelijkheid om ook API's te beschrijven.

Ten tweede wordt er tegelijk gewerkt aan een **referentie implementatie (piloot)** voor het metadateren van services. Op deze wijze wordt reeds in een vroeg stadium de toetsbaarheid in het veld getoetst.

2. WERKGROEP

Het doel en de aanpak van de virtuele werkgroep worden hieronder beschreven. Tevens worden de belangrijkste bevindingen uit deze brainstormsessie samengevat.

2.1 Doel van de werkgroep

Feedback verzamelen op eerste voorlopig sneuvelmodel:

- Opgeworpen issues behandelen
- Eigenschappen uit de vorige werkgroep afdrukken en verder concretiseren
- Voorbeelden van de deelnemers proberen inpassen in het sneuvelmodel

2.2 Oefening werkgroep

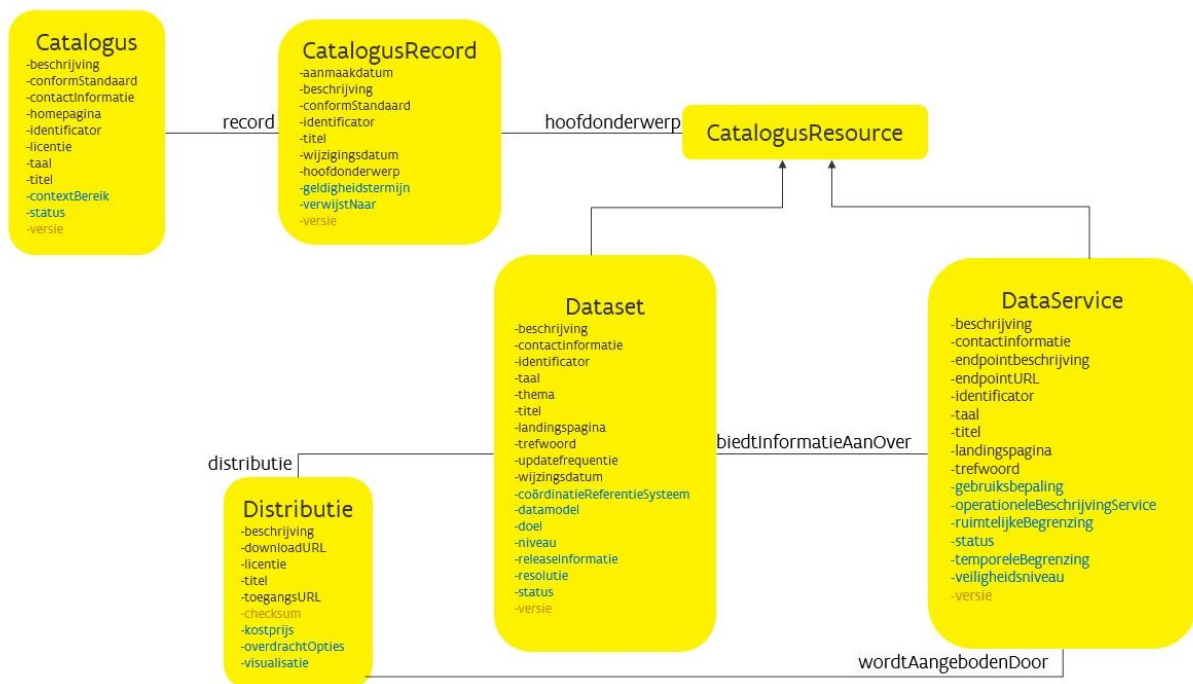
De werkgroep bestond uit slides waarbij er definities, attributen en relaties werden voorgesteld. Bij elke slide werden verschillende vragen gesteld en werd gevraagd aan de deelnemers om te reageren in de chat of via de microfoon.

Daarbij werd gevraagd om de voorgestelde definities, attributen en relaties te bekijken vanuit:

- Vanuit je eigen perspectief
- Vanuit de use cases
- Vanuit je eigen service en of deze hierin beschreven kan worden

2.2.1 Sneuvelmodel applicatieprofiel: Metadatastandaard voor services

Vanuit DCAT en de input uit de eerdere werkgroepen werd het eerste sneuvelmodel voor het applicatieprofiel opgesteld. Dit sneuvelmodel werd kort besproken en toegelicht.



2.2.1 Generieke elementen

In het bovenstaande applicatieprofiel werden enkele generieke elementen zoals DataService, Dataset, Distributie, Catalogus, CatalogusRecord en CatalogusResource geïdentificeerd. Deze werden vervolgens besproken en eventuele opmerkingen werden gecapteerd.

3. OPMERKINGEN UIT DE WERKGROEP

Vragen en opmerkingen bij het model

- Een catalogus wordt toch ook aanzien als een dataset? Het is toch ook belangrijk dat je uw catalogus zelf publiceert?
 - De complexe relatie tussen Catalogus als een subset van Dataset is expres niet opgenomen om het model of het zicht erop nu niet te ingewikkeld te maken. Maar inderdaad, een Catalogus is een speciale soort Dataset. De eigenschappen van Dataset druppelen door naar Catalogus in DCAT, maar deze staan nu apart omdat we focus willen nemen.
 - We willen ook alle catalogi harvesten zodat deze samen in de hoofdcatalogus geïntegreerd kunnen worden.
- CatalogusRecord en CatalogusResource dat zijn eigenlijk twee dingen die wij niet gebruiken. Wij hangen onze datasets onder de catalogus.
 - Dat is ook correct. CatalogusResource werd ingevoerd bij DCAT om een klasse te hebben waarin ze allemaal eigenschappen kunnen zetten met dezelfde semantiek binnen DCAT. Dit is bedoeld als abstracte klasse waarbij er gevraagd wordt om geen concrete elementen van CatalogusResource te gaan uitwisselen. CatalogusResource is een hulpmiddel om CatalogusRecord, Catalogus, Dataset en

DataService en Distributie te gaan uitwisselen. Dit laat ook toe om dat er naast Dataset en DataService andere dingen zijn die we expliciet willen gaan beheren.

- Onze Catalogus heeft meerdere Distributies, je kan die zowel op een SPARQL-endpoint gaan querien maar je hebt daar ook RDF-distributies van of een HTML-voorstelling.
 - Dat is inderdaad ook compatibel met hoe wij het zien, echter nemen we in dit sneuvelmodel focus op Dataservice.
- Wat is de relatie tussen Dataset en DataService?
 - De betekenis van deze relatie ligt nog volledig open, we behandelen dit later tijdens de sessie.

Vragen en opmerkingen bij concept “DatSERVICE”

Voorstel definitie: Een verzameling van operaties die ofwel toegang tot één of meer datasets voorzien ofwel een dataverwerkingsdoel hebben.”

Feedback op de definitie:

- De werkgroep noemt Metadata voor service en plots spreekt met hier over datSERVICE. Is datSERVICE geen specifieke vorm van een service? Kunnen we uitklaren waarom we dit hier reduceren tot DataService? En misschien moeten we dan ook de naam van de werkgroep veranderen?
 - *De term is overgenomen uit het DCAT-model en die gaat zeer breed, daar valt bv. Ook een SPARQL-endpoint onder, een wisselkoersberekenaar onder. Dit is zeer breed, we namen dit over.*
- Een loginservice of een authenticatieservice zijn services waarbij er niet echt data wordt teruggegeven maar waarvoor een bepaalde actie ondernomen wordt. Bijv. Voldoet iemand aan een bepaalde voorwaarde, dan krijg je een ja/nee terug en wordt er niet echt naar een bepaalde dataset gekeken of teruggeleverd. Kan het onderscheid tussen Service en DataService worden gemaakt in het model?
 - *We kunnen dit onderscheid maken maar wat wordt er dan bedoeld met Service, dit is niet hetzelfde als Dienstverlening? Een ja/nee terugkrijgen is ook een vorm van verwerking van data. Dit zit ook zo in de definitie als “verwerkingsdoel”.*

Suggestie om definitie te wijzigen: “Een verzameling van operaties die toegang tot één of meer datasets voorzien en/of een dataverwerkingsdoel hebben.

De werkgroep is akkoord met deze suggestie.

- De toegang tot de Datasets kan ook het gevolg zijn van een verwerking.

Voorstel eigenschappen:

- We spreken hier nog niet over kardinaliteiten. Bijv. EndpointURL kunnen één of meerdere zijn.
 - Dat is inderdaad waar, dit komt later in de volgende werkgroep nog aan bod. Op dit moment is er nog geen enkele kardinaliteit vastgelegd. We willen eerst focussen op de definities en de eigenschappen.
- **Landingpagina:** er staat geen verwijzing naar een DataService. Ik neem aan dat dit iets is langswaar je kunt navigeren om toegang te krijgen tot een Catalogus, Dataset, Distributie maar ook tot een Dataservice?
 - Nemen we deze op bij Landingpagina of bij de endpointURL?

Suggestie: De Landingspagina zou bij de Distributie kunnen staan, afhankelijk van de relatie tussen Dataset, DataService en Distributie.

- **Endpointbeschrijving:**
 - Heel concreet over de specifieke technische toegang tot het endpoint, operaties, errorboodschappen, WSDL op SOAP etc.
 - Gaat technisch beschrijven over welke graph je kan queren in welke dataset, geeft technische informatie weer over de dataset die je kunt beschrijven (=Beschrijving over het endpoint)
 - bv. ook de link naar de Open API beschrijving, een swagger beschrijving, de WSDL, de GetCapabilities, ...
 - Geen URL maar een invoerveld waarin je vanalles beschrijft?
 - *Momenteel is dit nog open, er bestaan machineleesbare formaten bij bv. Swagger of OPENAPI die dit heel mooi invullen. We kunnen definiëren hoe we dit willen ingevuld krijgen maar dit is een vervolgvraag. We willen hier focussen op de semantiek en niet op de implementatie.*
 - Een endpoint aanbieden is meestal om via machine tot machine te kunnen communiceren.
- **Landingspagina:**
 - Een bredere pagina met context waarbij dat er algemene afspraken rond het endpoint en de dataservice beschreven worden. Bijv. Een link naar de algemene documentatie, kan ook ook over de machtigingen gaan om verder te kunnen
 - Geeft gebruiker letterlijk toegang tot SPARQL-endpoint service via een formulier. (=Gebruik of richting van gebruik van informatie van endpoint weergeeft.)
 - Link naar de handleiding in de gebruikersomgeving
 - Het kan ook zo zijn dat deze samenvallen in de implementatie van de service.
 - En misschien is er ook een verschil in de open en gesloten community?
 - Met een landingspagina wil je de mens bedienen.
 - Indien we ze beide willen behouden, wat verwachten we in een landingspagina en wat verwachten we in een endpointbeschrijving?
- **Gebruiksbeplating:**
 - Zit de Gebruiksbeplating niet beschreven op de Landingspagina? Beschrijven we dit apart of bij een andere eigenschap bij?
 - Authorisatie of toelating staat niet altijd op de landingspagina. Nuttig om ook op serviceniveau een toelating te voorzien.
 - De info over wie er toelating heeft staat ook op de landingspaginas van services bij een deelnemer .
 - Kunnen we dit definiëren en onderscheid tussen maken
 - Machtigingen zijn gebruiker gebonden en niet in detail beschreven in de gebruikershandleiding. Geen directe relatie tussen machtiging en endpointomschrijving.
 - Kan deze Gebruiksbeplating op het niveau van de CatalogusResource opgenomen worden?
 - *Model is nu plat en omdat DataService een subklasse is van CatalogusResource (zie pijl in schema) geldt deze eigenschap ook voor CatalogusResource.*
 - *Model is nu platgeslagen voor het doel van de discussie. Indien dit een eigenschap is die ook voor alle entiteiten in de catalogus moet gelden dan kunnen we die toevoegen aan CatalogusResource.*

- **Veiligheidsniveau:**
 - Hangt van uw implementatie af.
 - Is dit geen gebruiksbepaling?
 - Hoeft dit wel op de dataservice te zitten? Bestaat hier een eigenschap? Dit wordt meestal op het data of op het informatieniveau besproken.
- **Status**
 - Zijn dat statussen van een dataservice bv. Operationeel of niet? In beta? Productie?
 - In opbouw, compleet af, gearchiveerd vanuit de geo-wereld?
 - End of life (kan nog even gebruikt worden maar erna niet meer)
 - Test (TNI), productie, beta, waar kunnen we dit kwijt? Of zit dat in versie? De soorten protocollen waar zit dat?
 - Mogelijkheid om meerdere statussen te hebben voor een endpoint met een begin en einddatum bv. Productie die heeft geen data maar bv. Beta-endpoint wel.
 - *Dat is granulariteit. Een beta-endpoint vs een productie-endpoint, is dat een en dezelfde dataservice waarin dat we de verschillende varianten zetten en meerdere eigenschappen zitten. OF maken we verschillende dataservice instanties die we onderling gaan verbinden met een relatie.*
 - Ik zou ook werken met een einddatum en die is dan niet is ingevuld voor dingen die nu beschikbaar zijn en beschikbaar blijven en dat je dan zegt vanaf die datum is die niet meer beschikbaar en dan ga je misschien de-publiceren.
 - *Dat is dan een gebruikstatus? Of de verschillende distributies in een service?*
 - Is de service end-of-life of is de data end-of-life? En dan zit je eerder op distributie niveau dat je moet aangeven of die niet meer beschikbaar is en je kunt aangeven of er recente gegevens zijn of historische gegevens beschikbaar zijn.
 - Suggestie: status van de ontwikkeling vs status van de service
 - Status van de ontwikkeling: TNI, productie, beta
 - Status van de service/operationele status: is een service beschikbaar of is ze niet beschikbaar (online of offline) = de actuele operationele status
- **Request parameter**
 - Waar wordt dat bijgehouden?
 - *Op dit moment als onderdeel van de EndpointBeschrijving.*
 - *Bv. Bij SOAP de WSDL beschrijving*
 - Bij dataset staat er datamodel maar het kan zijn dat er zoveel inputparameters zijn dat je daar ook eventueel een model voor nodig hebt of een XSD?
 - *2 opties hier:*
 - *In de EndpointBeschrijving een verwijzing naar het gebruikte datamodel voor dat specifieke endpoint, complex*
 - *Ofwel onmiddellijk zichtbaar voor de gebruikers van de catalogus, het datamodel is onmiddellijk toegankelijk via deze concrete eigenschap.*
- **Temporele begrenzing**
 - Bij MAGDA beperkt dit zich tot actueel en historiek
 - *Kan dit ook in de beschrijving, endpointbeschrijving of in de landingspagina?*

- Het is vaak ook in de machtiging bepaald hoever je mag teruggaan in de tijd. Je mag bv. Gegevens tot zoveel jaar terugvragen maar dat hangt af van de toepassing.
- Zit deels in de naam zelf en is ook beschreven in de handleidingen via de landingspagina.
- Kan je het ook in de beschrijving kwijt welk soort gegevens kan opgevraagd worden door welk soort groepen?
 - *Willen we deze vraag kunnen stellen: "Ik wil alle services die data kunnen leveren tussen 1990 en 2000?" of "Ik wil alle services die historische gegevens bevatten?" Willen we dit nu ondersteunen of is dit eerder voor een later gebruik.*
- Niet zeker of hier ooit iemand een query over gaat doen.
 - *In de bron/dataset zelf kun je dit wel kwijt als temporal resolution.*
- **Ruimtelijke begrenzing**
 - Niet zeker of iemand ooit een query gaat doen om te kijken of dit binnen de ruimtelijke begrenzing valt.
 - *In de bron/dataset zelf kan je wel spatial resolution (in meters) kwijt.*
 - Iemand kan een bepaalde service aanbieden van een bepaalde gemeente, zei het dat de dataset over heel Vlaanderen gaat. Zal die dat dan opnemen in bv. De landingspagina of in de endpointbeschrijving?
- **Authenticatie en toegangsinformatie**
 - Belangrijk, maar hoort eerder thuis bij de landingspagina of bij een andere eigenschap
- **Suggestie: ServiceType**
 - Technische typologie, zit dit al ergens? Aangeven of het SOAP, REST, LOD, OGC-service etc is? Kan dit ook met keywords? Kunnen we hier een lijst van maken?
- **Wettelijke beperkingen**
 - Interesse
- **Extent van de dataservice**
 - Geen interesse
- **Quality of service**
 - Enkeling is geïnteresseerd
- **Te gebruiken tot/afneembaar tot**
 - Interesse
- **Bron**
 - Interesse
- **Data elementen**
 - Geen interesse
- **Distributieprotocollen**
 - Interesse
- **Gebruikte/extra WS afhankelijkheden**
 - Enkeling is geïnteresseerd
- **Actuele/historische gegevens**
 - Interesse
- **Doelpopulatie**
 - Geen interesse
- **Hydra-cg (hypermedia gebruik)**
 - Minder van toepassing in deze discussie

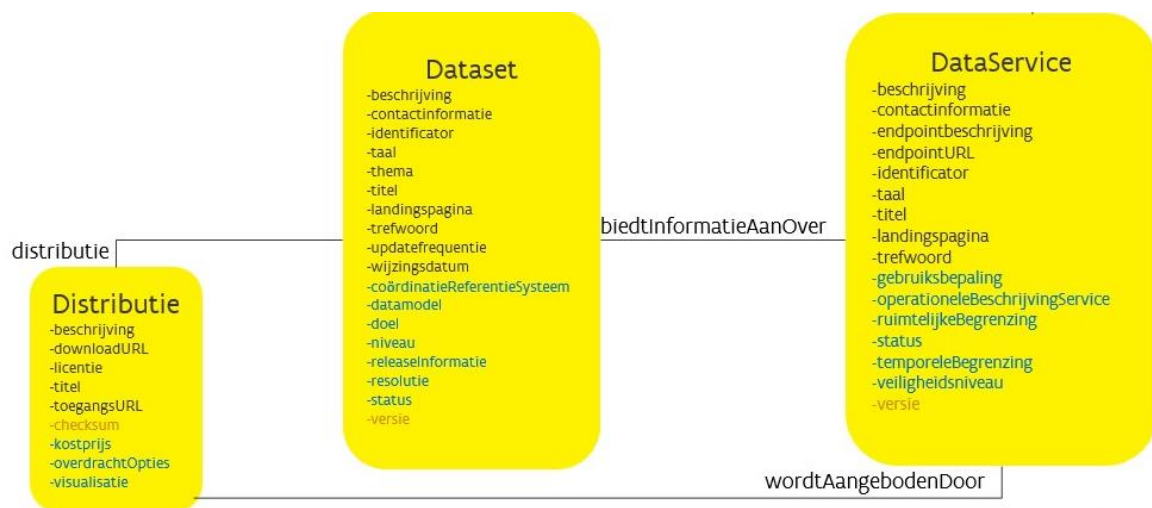
- Geen interesse vanuit de deelnemers

Wat nemen we mee: operationele status vs status van de service, begrenzungen onduidelijk, veiligheidsniveau niet opnemen, gebruiksbepaling zeker interesse

Voorstel relaties:

Vanuit DCAT is nu aanwezig:

- 1) Dataset-Distributie relatie: een dataset wordt gematerialiseerd en toegankelijk in verschillende fysieke vormen, in bepaalde distributie.
- 2) Distributie-Dataservice: Een bepaalde distributie wordt aangeboden door een DataService.
- 3) DataService-Dataset: Een DataService gaat informatie gaan aanbieden over een Dataset.



- Waarom staat er bij Distributie downloadURL en bij DataService endpointURL, is dit niet hetzelfde?
 - *Dataset is iets conceptueel, een verzameling van informatiepunten dat wordt gematerialiseerd in een distributie, in een volledig neergeslagen vorm die je kan downloaden. Met de downloadURL wordt bedoeld dat je de volledige dataset zoals beschreven op het conceptuele niveau volledig kunt downloaden of binnentrekken.*
 - *Een DataService is intelligenter en waarbij die niet zomaar downloadbaar is maar waarbij die data op een intelligente manier kan aangeboden worden (in stukjes, in brokjes) eventueel ook met een ander protocol etc. Dit is complexer dan een URL die je kunt downloaden.*
- Maar je kunt een volledige dataset via een dataservice ook downloaden?
 - *Een Dataset is een afgesproken beschreven set van informatie en de Distributie daarvan is de concrete vorm. Die DataService is een intelligente manier om informatie te delen.*
- Vraag omtrent de service op/van de catalogus: beschrijf je de 'endpointURL' in 'Catalogus' of in 'DataService'?

- *Dit is op een hoger niveau. Is een Catalogus al dan niet een Dataset. Wanneer we Catalogus zien als een Dataset dan zou je dat als een Distributie kunnen ontsluiten. Wanneer je dat zou aanbieden op een REST-API waarbij je Catalogi informatie kan ondervragen dan zal je dat waarschijnlijk als een DataService modelleren.*
- Distributie is de dataset in een bepaalde vorm. Wat bedoel je hiermee? De representatie van de Dataset?
 - *Bv. De Dataset Rijksregister, dit is iets conceptueel. De Distributie is dan de vorm dat het RijksRegister wordt ontsloten en dat kan zijn in zijn geheel als een .xml bestand volgens dit xsd schema. Maar tegelijkertijd kan je het ook downloaden via een .csv volgens een ander schema. De Distributie legt alle fysieke vormen op en eigenschappen van die Dataset.*
- Kunnen we niet spreken van representatie van Dataset ipv Distributie?
 - *Dat zou kunnen maar we hebben hier uit DCAT de benaming overgenomen.*
 - *Zijn er nog onderlinge relaties tussen DataServices?*
 - *Onderling bv. Is het een bèta, een test, productie? We zouden die ook aan elkaar kunnen relateren door bv. de relatie "versionOf", de statuseigenschap wordt dan een relatie. Een ander voorbeeld is voor de versie, versie 1, versie 2*
 - *Zijn er nog eigenschappen of relaties of use cases waarbij we services gaan koppelen buiten deze eigenschappen die zinvol zijn?*
 - *Chaining van services: services die jij gebruikt om deze services aan te bieden (een soort van bronverhaal, source)? Dit kan wel interessant zijn.*
- *Optie: we capteren dat services gerelateerd zijn aan elkaar maar we benoemen de relaties nog niet, we laten dit nog open voor de toekomst.*
- *Zijn er nog relaties met Agenten die belangrijk zijn?*
 - *Iemand met definitieve eindverantwoordelijkheid van de data en service*
 - *Iemand die de service zelf heeft opgezet*
 - *In de data ook nog de bewerker rol*
 - *Creator, publisher (=uitgever, distributeur, verdeler), contactpunt, rightsholder (=eigenaar), bewerker (van de data; zit bij de dataset)*
 - *Check met de ISO definities of we deze kunnen mappen => Geraldine deelt deze*

De concepten Dataset, Distributie, CatalogusResource, CatalogusRecord en Catalogus werden niet meer behandeld tijdens de werkgroep. We vragen de deelnemers om hier zelf nog eens door te gaan. De basis hiervoor is te vinden in [het verwerkingsdocument](#).

4. HUISWERK: DEMO REFERENTIE-IMPLEMENTATIE IN GEONETWORK

Bekijk in [de slides](#) hoe in te loggen en de demo zelf uit te proberen.

5. VOLGENDE STAPPEN

Verdere feedback die tijdens de werkgroep werd gegeven en die volgt op dit verslag wordt verwerkt tegen de voorlopig laatste thematische werkgroep.

Hoe geef ik feedback?

Er zijn twee mogelijkheden om feedback te geven:

- Via mail naar informatie.vlaanderen@vlaanderen.be
→ Vermeld het OSLO-thema: metadataVoorServices

of

- Via [GitHub](https://github.com/Informatievlaanderen/OSLO-Standaarden/tree/master/standaarden-in-ontwikkeling/metadata-voor-services) <https://github.com/Informatievlaanderen/OSLO-Standaarden/tree/master/standaarden-in-ontwikkeling/metadata-voor-services>

6. AANKOMENDE WERKGROEPEN

Indien u (of een collega) graag zou willen deelnemen aan een van de aankomende werkgroepen, kan u via de onderstaande link een overzicht van de workshops terugvinden en u ook zo inschrijven. De derde en voorlopig laatste thematische werkgroep zal plaatsvinden op dinsdag 10 november 2020 om 9u via Microsoft Teams. De link zal na inschrijving worden doorgestuurd naar de deelnemers.

<https://overheid.vlaanderen.be/opleiding/publieke-werkgroepen-oslo-metadataastandaard-voor-services>