

Gids naar kennis over Proactief Datadelen

**Een praktijkbeproeving van een
federatieve catalogus**

Vakgroep Architectuur Grenzeloos Datalandschap

November 2025



COLOFON

Titel	Gids naar kennis over Proactief Datadelen - een praktijkbeproeving van een federatieve catalogus
Versie	1.0
Datum	November 2025
Opdrachtgever	Stuurgroep Grenzeloos Datalandschap
Opdrachtnemer	Vakgroep Architectuur Grenzeloos Datalandschap (GDL)
Auteurs	Ronald Kok, Hugo Mostard, Sahin Özaltun, Sergio Richardson Nils Vissenberg
Contactpersoon	Info@grenzeloosdatalandschap.nl
Vertrouwelijkheid	Openbaar

Inhoudsopgave

Samenvatting	3
Grenzeloos Data Delen	5
Context van data publiceren en zoeken over meerdere databronnen	7
Resultaten Praktijkbeproeving.....	10
Bijlage Solution Architectuur Prototype	11



Samenvatting

Doel en context

Het *Grenzeloos Datalandschap* (GDL), een samenwerkingsverband tussen overheden, kennisinstellingen en bedrijven in de provincie Zuid-Holland, wil data delen makkelijker maken en samenwerking stimuleren. Met deze gids delen we kennis die we opgedaan hebben met een praktijkverkenning naar een **federatieve catalogus**. Wij willen hiermee aansluiten op de Europese en Nederlandse ontwikkelingen in het delen van data zoals data spaces en het federatief datastelsel. In deze ontwikkelingen staat de autonomie van de data leverancier centraal en kunnen data afnemers eenvoudiger en sneller data consumeren. Met de beproeving bouwen wij kennis op van de praktische uitdagingen die wij beschikbaar kunnen stellen aan de GDL – partners in de vorm van deze gids.

Wat is een federatieve catalogus?

Een federatieve catalogus is geen klassieke catalogus. Het is een verwijzingsindex naar catalogi van organisaties die de beschrijving van de data (metadata) zelfstandig beheren. Het federatieve element is dat iedere organisatie deze data volgens afspraken en standaarden beschikbaar stelt met behoud van autonomie. Dat betekent dat de data bij de bron blijft en dat men met een zoekdienst de data kan vinden.

De scope van de praktijkbeproeving is aldus:

- Realiseren van een werkend prototype van een federatieve catalogus
- Focus op datasets die gaan over *mobiliteit* in lijn met de Europese *High Value Datasets* (HVD)
- Ophalen van feedback bij deelnemers over nut en noodzaak.

De kern van de beproeving:

De federatieve zoekdienst werkt volgens drie stappen:

1. **Publiceren:** de aanbiedende partij publiceert de kenmerken van datasets in het afgesproken formaat.
2. **Vinden:** men kan zoeken op de kenmerken van deze datasets.
3. **Binden:** de gevonden datasets kan men integreren in eigen systemen.

Opgedane inzichten

- Toepassing van de open source OpenCatalogi ¹technologie kan gebruikt worden als basis voor de federatieve catalogus. Deze ondersteunt 'publiceren', 'vinden' en 'binden'.
- Door uniformeren van kenmerken van een dataset naar de DCAT-AP-NL 3.0 standaard is bevraging met natuurlijke taal met ondersteuning van AI/LLM eenvoudig toepasbaar.
- De gewenste autonomie van de publicerende deelnemer blijft gewaarborgd.
- De individuele catalogi met verwijzing naar datasets zijn goed vindbaar. De hoogwaardige datasets zijn als zodanig niet snel af te leiden uit deze catalogi. Dit vroeg om een extra inspanning.

¹ <https://github.com/OpenCatalogi>



- Processen voor het vrijgeven van open data zijn ondoorzichtig. Daarmee is lastig vast te stellen of er meer beschikbaar is dan opgenomen in de catalogi. Dat vraagt tevens om een extra inspanning.
- Samenvoeging van datasets over meerdere regio's is een expliciete wens van gebruikers.



Grenzeloos Data Delen

Introductie Grenzeloos Datalandschap

Het Grenzeloos Datalandschap (GDL) is een samenwerkingsverband tussen overheden, kennisinstellingen en bedrijven in de provincie Zuid-Holland. Deze samenwerking beoogt op hoofdlijnen twee belangrijke resultaten, zijnde:

- Het verlagen van drempels in het delen van data – vandaar de term Grenzeloos. Dit gaat over het delen van data over organisatiegrenzen heen.
- Het stimuleren van samenwerkingen tussen overheden en private partijen waarbij data ingezet wordt voor het ondersteunen van maatschappelijke vraagstukken.

Het GDL kent een regionale focus. Het GDL sluit aan bij de Europese en landelijke ontwikkelingen die gericht zijn op het delen van data ('interoperabiliteit').

Binnen het GDL spannen we ons in voor het delen van kennis over de ontwikkelingen en het vertalen naar praktische handvatten voor het delen van data ('interoperabiliteit') tussen publieke en private partijen. Wij zien steeds meer de behoefte dat mensen sneller antwoord willen op hun vraag waarbij zij zoeken vanuit hun eigen beleving en context. Daarin zie je een stijgende toepassing van generatieve AI die een betrouwbare bron vereist. Met de federatieve catalogus voorzien wij in die bron door zoveel mogelijk te standaardiseren en uniformeren met behoud van autonomie.

Wat is een federatieve catalogus?

Een federatieve catalogus is geen klassieke catalogus. Het is een verwijzindex naar catalogi van organisaties die de beschrijving van de data (metadata) zelfstandig beheren. Het federatieve element is dat iedere organisatie deze data volgens afspraken en standaarden beschikbaar stelt met behoud van autonomie. Dat betekent dat de data bij de bron blijft en dat men met een zoekdienst de data kan vinden.

Wat is dan het federatieve eraan? Federatief zegt zoveel als; iedereen behoudt zijn eigenheid en autonomie en conformeert zich aan gemaakte afspraken. In de context van data delen; je bepaalt zelf welke data je onder welke condities beschikbaar stelt, maar je maakt wel afspraken over de uitwisselbaarheid (interoperabiliteit) van die data. Het zijn overkoepelende afspraken hetgeen een beeld oproept van een 'paraplu' zoals geschetst in onderstaand figuur.

Een federatieve catalogus is een katalysator voor duurzaam datadelen.



Reden, opzet en aanpak van de praktijkbeproeving naar een federatieve catalogus

De belangrijkste reden voor de beproeving is het ervaren van hoe je een federatieve catalogus toepast. Wat komt erbij kijken, zowel organisatorisch als technisch?

De opzet van de beproeving is tweeledig;

1. Aantonen wat de federatieve catalogusfunctie inhoudt door het demonstreren van een werkend prototype.
2. Het verkrijgen van waardevolle feedback van de deelnemers tijdens de demo sessies over nut en noodzaak van deze katalysator. Wie maken we er blij mee en waar zouden we accent op moeten leggen?

Qua aanpak hebben we gekozen voor werken in 6 sprints van 2 weken. In deze sprints realiseerden we het prototype met tussentijdse demo/feedback sessies.

In het prototype hebben we gekozen voor het delen van data van zogenaamde hoogwaardige datasets. Dit is de vertaling van de term 'high-value dataset' welke afkomstig is uit de Europese verordening voor het proactief delen van data door overheden. Wij hebben willekeurig het thema 'mobiliteit' gekozen in deze praktijkbeproeving en de federatieve catalogus opgezet op bestaande open data van de deelnemers.



Context van data publiceren en zoeken over meerdere databronnen

Het beschikbaar stellen van (open) data in een catalogus bestaat uit vijf generieke fases²:

1. Data verkrijgen
2. Datakwaliteit verbeteren
3. Semantische harmonisatie en data integratie
4. Data publiceren
5. Data ontsluiting en visualisatie



De fase ‘Data verkrijgen’ is het proces van het systematisch verzamelen van informatie uit verschillende bronnen.

De fase ‘Data Kwaliteit verbeteren’ omvat technieken en procedures om de nauwkeurigheid, consistentie en betrouwbaarheid van de verzamelde gegevens te verbeteren en ervoor te zorgen dat deze voldoen aan specifieke normen en criteria.

In de fase ‘Semantische harmonisatie & Data integratie’ worden gegevens uit verschillende bronnen op één lijn gebracht met consistente formaten en betekenissen, waardoor naadloze integratie en interoperabiliteit tussen uiteenlopende datasets mogelijk wordt.

De fase ‘Data Publiceren’ omvat alle activiteiten en functionaliteiten die erop gericht zijn om verzamelde en verwerkte gegevens toegankelijk te maken via verschillende kanalen, zoals open API's, open dataportalen en -platformen, en om transparantie en gemakkelijke toegang te garanderen.

De fase ‘Data ontsluiting & Visualisatie’ omvat de mogelijkheden voor het aanbieden van aansluitpunten voor toegang tot en bevragen van open datasets. Tevens bevat dit functionaliteiten voor het visualiseren van deze open data.

Hoe ziet de slimme zoekdienst voor mens en machines eruit?

De slimme zoekdienst voor mens en machine is gebaseerd op het eerdergenoemde mechanisme van publiceren-vinden-gebruiken³. Dit mechanisme is een patroon dat veel wordt gebruikt in data- en service gebaseerde systemen, zoals metadata-catalogi en webservices. Het bestaat uit drie stappen:

1. Publiceren
Een data- of serviceprovider registreert zijn dataset of dienst in een centrale catalogus of register. Dit gebeurt met behulp van metadata die beschrijft wat de dienst of dataset inhoudt, hoe je die kunt gebruiken en waar die zich bevindt.

² <https://beopen-dep.eu/wp-content/uploads/D2.1-BeOpen-technical-framework-design.pdf>

³ In technische termen wordt dit het publish-find-bind principe genoemd.



2. Vinden

Een gebruiker of applicatie zoekt in de catalogus naar relevante datasets of diensten. De zoekresultaten zijn gebaseerd op de gepubliceerde metadata. In de praktijkbeproeving is open data gebruikt. Bij gesloten data zal er een autorisatie mechanisme nodig zijn.

3. Binden

Zodra de juiste dataset of dienst is gevonden, verbindt de gebruiker of applicatie zich ermee. Dit betekent dat de data of dienst direct wordt geïntegreerd in de eigen omgeving, bijvoorbeeld via een API-koppeling.

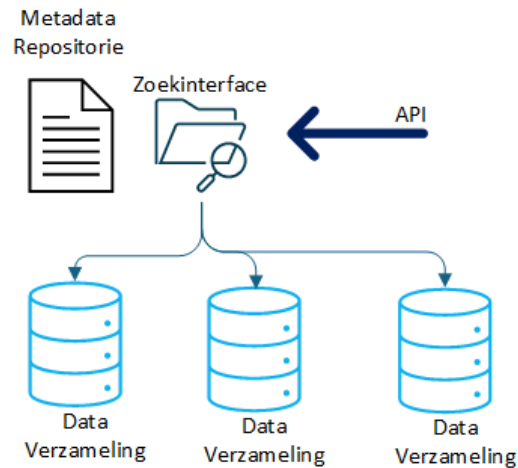
Het is duidelijk dat goede metadatering van de dataset en dienst de basis vormt voor het slimme zoeken. De metadata moet voor zowel mens als machine interpreteerbaar zijn.

Wat zijn de theoretische vereisten voor een federatieve catalogus?

Een federatieve data catalogus biedt een uniforme portal voor datavoorziening over meerdere bronnen, terwijl elke bron controle behoudt over zijn eigen data en beleid. Dit helpt om datasilo's te doorbreken en samenwerking te verbeteren zonder alle data in één repository te dwingen.

Een federatieve data catalogus bestaat uit verschillende kerncomponenten, waaronder een metadata repository en opslag, dataverbindingen, een gebruiksvriendelijke zoekinterface, en integratie-API's. De metadata repository moet diverse datatypes kunnen beheren en grote volumes metadata kunnen verwerken.

Component	Beschrijving
Metadata Repository	Centraal metadata-opslag voor datasets
Dataverbindingen	Integratie met verschillende dataverzamelingen
Zoekinterface	Gebruiksvriendelijke interface voor datavoorziening
API's en Integratie	API's voor integratie met bestaande tools



Essentiële elementen voor de federatieve catalogus zijn aldus⁴:

Afspraken en standaarden: Er moeten afspraken worden gemaakt over hoe metadata worden beschreven, gedeeld en geïnterpreteerd. Dit omvat het gebruik van gemeenschappelijke semantiek en technische standaarden.

Betekenisvolle verbindingen: De catalogus moet verschillende metadata bronnen kunnen verbinden op een manier die semantisch correct en technisch uitvoerbaar is.

Voordelen van een federatieve data catalogus

Geïntegreerde Data Ontdekking: Gebruikers kunnen data snel vinden door een enkele catalogus te doorzoeken in plaats van door losse silo's met eigen kenmerken te navigeren.

Behouden Domein Autonomie: Data aanbieders behouden controle over hun data, wat innovatie bevordert.

Verbeterde Data Kwaliteit en Vertrouwen: Transparantie over data herkomst en metadata verhoogt het vertrouwen in de data.

⁴ Volgens de kennisbankpagina over de FDS-catalogus



Resultaten Praktijkbeproeving

Kernconstateringen

Tijdens de beproeving bleek dat organisaties verschillend omgaan met het publiceren van open data, en dat processen en datakwaliteit vaak niet transparant zijn. Tegelijkertijd zijn hoogwaardige datasets goed beschikbaar via standaarden zoals DCAT, en is de inzet van een federatieve catalogus veelbelovend, zeker in combinatie met AI.

Wel is er behoefte aan betere metadata, meer kennis over de gebruikte infrastructuur, en duidelijkere publicatieprocessen. Tot slot is het belangrijk om herbruikbaarheid, interoperabiliteit en combinatiemogelijkheden van datasets (bijv. over organisatiegrenzen heen) mee te nemen in toekomstig beleid en inkoop.

Gedetailleerde toelichting:

Tijdens de praktijkbeproeving zijn we tegen een aantal punten aangelopen.

1. Alhoewel 'High Value Datasets' door de EU zijn voorgeschreven in de 'uitvoeringsverordening voor High Value Datasets' zijn er toch nog verschillen tussen de verschillende pilot participanten.
2. Iedere organisatie (gemeente, provincie) heeft zijn eigen processen om data openbaar te maken.
3. Datakwaliteit is niet altijd duidelijk
4. Processen om open data te publiceren zijn zelf niet open en daarom lastig om te achterhalen.
5. Voor de praktijkbeproeving hebben we gebruik gemaakt van het HVD-thema mobiliteit welke vaak vertaald werd in wegen en verkeer of transport en infrastructuur.

Successen en resultaten

1. Er zijn meerdere hoogwaardige datasets beschikbaar om te gebruiken als data bij de bron.
2. Hoogwaardige datasets zijn goed gepubliceerd via een DCAT koppeling (standaard).
3. Open Catalogi kan goed gebruikt worden als federatieve catalogus.
4. Extra eigenschap: de federatieve catalogus kan aangesproken worden via API interface voor natuurlijke bevragingen.
5. Een federatieve catalogus is een goede eerste stap als katalysator voor een data uitwisseling binnen het Grenzeloos Data Landschap.

Lessons learned

1. Proces van vrijgeven open data dient verduidelijkt en openbaar te worden.
2. Verdiepingsslag van kennis over de gebruikte componenten is wenselijk, zoals OpenCatalogi en de AI-technologie.
3. Neem de vereisten voor het beschikbaar stellen van datasets conform DCAT (dumps/exports) bij de aanschaf van nieuwe applicaties mee in de inkoopvoorwaarden.
4. Uit de praktijkbeproeving is gebleken dat zoeken over elkaars organisatie grenzen wenselijk is (bijvoorbeeld 'fietsroutes' over Den Haag en Rotterdam als 1 geheel).
5. Er is extra aandacht nodig voor het automatisch verkrijgen van metadata (automatiseren waar mogelijk).
6. Positie van OpenCatalogi binnen common-ground (Haven) valideren.



Bijlage Solution Architectuur Prototype

Beschrijving van de gebruikte applicaties en voorzieningen

Het prototype ontvangt de metadata van de high-value datasets die nodig is voor zoeken en vinden van de diverse catalogi die we als bron benut hebben. In ons geval waren dat de open data catalogi van de gemeente Rotterdam, Den Haag en de Provincie Zuid-Holland.

De metadata wordt eerst omgezet naar het DCAT- AP-NL 3.0 profiel en gecontroleerd op verplichte waarden. Als dat klopt dan leest OpenCatalogi de metadata in met als doel deze te indexeren en vervolgens beschikbaar te stellen met een Zoek API. Deze Zoek API kan resultaten vinden op basis van trefwoorden maar ook als bron dienen voor het voeden van andere applicaties.

Bij een catalogus denk je al snel aan een (web)applicatie waarin je kan zoeken. Dat kan het zeker zijn, maar dat is niet alles. De applicatie is niet meer dan een voorkant voor een slimme zoekservice. Achter de (web)applicatie zit een soort van 'orakel' wat je kan bevragen voor het vinden van het antwoord op de vraag, waar je relevante en betrouwbare data kan vinden. De slimmigheid daarvan is dat je die vraag kan stellen in je eigen taal met je eigen context.

Het slimme zoeken is mogelijk door de Zoek API aan te roepen en de resultaten aan te bieden aan een commercieel platform waarin je een prompt kan ingeven, vergelijkbaar met ChatGPT.

Demo – zoek en gij zult vinden en ... binden



Een meer technische uitwerking is onderstaand weergegeven.

