

Handreiking

# Een open-source dataplatform in de Europese cloud

Een voorbeelduitwerking van de gemeente Nijmegen

GGI-Cloud Expertisecentrum

**VNG Realisatie**

Nassaulaan 12  
2514 JS Den Haag

Auteur: Jan Brinkkemper (VNG), Arjen Verhulst (gemeente Nijmegen)

Datum: april 2026

Versie: 1.0

## Inhoud

|                                                                   |   |
|-------------------------------------------------------------------|---|
| Inleiding .....                                                   | 3 |
| 1. Hosting .....                                                  | 3 |
| 1.1. Selectiecriteria hostingpartij en implementatiepartner ..... | 4 |
| 2. Onderdelen dataplatform .....                                  | 5 |
| 2.1. Datawarehouse en objectstorage .....                         | 5 |
| 2.2. Datagovernance .....                                         | 6 |
| 2.3. Reporting .....                                              | 6 |
| 2.4. Security .....                                               | 7 |
| 3. Beheer en benodigde expertise .....                            | 7 |
| 4. Risico's en aandachtspunten .....                              | 7 |
| 5. Conclusie .....                                                | 8 |

## Inleiding

Veel gemeenten maken gebruik van een dataplatform of willen hier gebruik van gaan maken. Een dataplatform biedt gemeenten de mogelijkheid om data uit verschillende bronnen (applicaties binnen de gemeente, externe registraties, open-data etc.) op één plek beschikbaar te maken en deze data te bewerken, analyseren en beschikbaar te maken voor de organisatie. Voor gemeenten is een dataplatform vaak het startpunt om meer datagedreven te werken. De data uit het platform wordt bijvoorbeeld gebruikt voor het vullen van dashboards, managementrapportages of gebruikt als startpunt voor het opstellen van beleid. Ook kan de data uit een dataplatform gebruikt worden in AI-toepassingen.

Veel dataplatformoplossingen zijn gebaseerd op diensten van de grote techbedrijven, zoals Microsoft (Azure Fabric) en Amazon (AWS). Deze zijn vaak snel te implementeren en gebruiksvriendelijk. Tegelijkertijd roept het gebruik van deze diensten vragen op over soevereiniteit en de afhankelijkheid van Big-Tech. Deze overwegingen hebben gemeente Nijmegen aangezet om een volwaardig dataplatform te realiseren met minimale afhankelijkheid van grote (vaak) Amerikaanse Tech bedrijven (Big-Tech), op basis van open-source software en gehost vanuit een Europese cloud aanbieder. Deze notitie schetst de oplossing waar de gemeente Nijmegen op uitgekomen is en de overwegingen die zij daarbij hebben gemaakt.

## 1. Hosting

Voor de gemeente Nijmegen was soevereiniteit een belangrijk punt om mee te nemen in hun overweging voor een hostingpartij. Daardoor lag de focus op Europese aanbieders. Bij het selectieproces is gekeken naar security (voldoet aanbieder bijvoorbeeld aan ISO27001, kan omgeving gepentest worden), techniek (aangeboden services), kosten en de geboden ondersteuning.

Een open-source dataplatform kan op elk gewenst cloudplatform gehost worden. Dit kunnen zowel platformen van Amerikaanse als van Europese aanbieders zijn. Voor de gemeente Nijmegen was het belangrijk dat de hostingpartij een aantal technische zaken kon leveren namelijk:

- PostgreSQL
- S3-compatible object storage
- OpenSearch / Elasticsearch
- Logging services
- Private connections naar on-prem omgeving
- API resource benadering
- Identity and Access Management (IAM)
- Entra ID integratie
- Infrastructure as Code (IaC)

Om de oplossing onafhankelijk te houden van de hostingaanbieder, is het van belang dat geen gebruik wordt gemaakt van propriëtaire platformservices of ten minste wordt nagedacht hoe deze vervangen zouden kunnen worden wanneer gemigreerd wordt naar een andere hostingpartij. Propriëtaire services zijn services die sterk gerelateerd zijn aan een specifieke cloudplatform en vaak ook alleen werken binnen het eco-systeem van deze aanbieder. Denk bijvoorbeeld aan Azure Monitor of Azure Functions wat, zoals de naam al zegt, alleen goed functioneert binnen het Azure ecosysteem.

Er zijn in Europa inmiddels een aantal volwassen partijen waarnaar gekeken kan worden. Gemeente Nijmegen heeft onderstaande partijen meegenomen in hun vergelijking:

- Scaleway (Frankrijk)
- OVHCloud (Frankrijk)
- T Cloud Public (Duitsland)
- STACKIT (Duitsland)
- UPCloud (Finland)

### 1.1. Selectiecriteria hostingpartij en implementatiepartner

Op basis van de architectuurschets vooraf zijn de volgende requirements gehanteerd om de hostingpartijen te beoordelen:

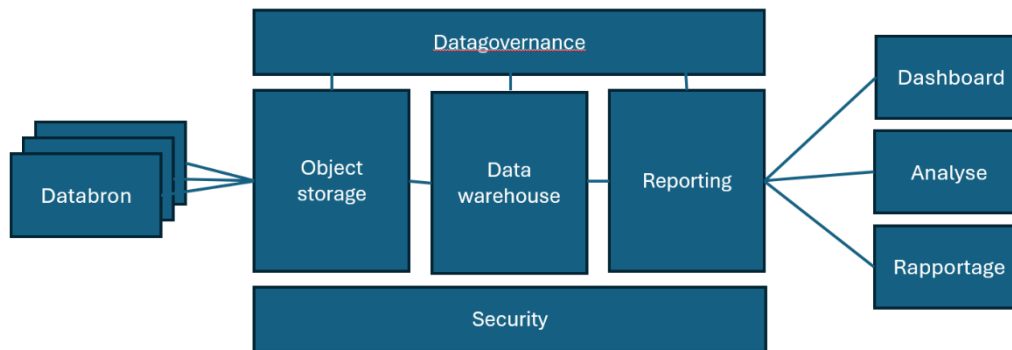
|                                  | Hostingpartij 1 | Hostingpartij 2...n |
|----------------------------------|-----------------|---------------------|
| Soevereiniteit                   |                 |                     |
| Availability zones               |                 |                     |
| Milieu                           |                 |                     |
| CSR                              |                 |                     |
| Roadmap                          |                 |                     |
| ISO27001                         |                 |                     |
| Pentesten                        |                 |                     |
| Beschikbaarheid managed services |                 |                     |
| PostgresSQL                      |                 |                     |
| S3-compatible object storage     |                 |                     |
| Opensearch/ Elasticsearch        |                 |                     |
| Logging                          |                 |                     |
| Private connections              |                 |                     |
| API resource benadering          |                 |                     |
| IAM en accountmanagement         |                 |                     |
| Entra ID integratie              |                 |                     |
| Infrastructure as code           |                 |                     |
| Egress                           |                 |                     |
| Ingress                          |                 |                     |
| Billing direct mogelijk          |                 |                     |
| Support-levels                   |                 |                     |
| Inhoudelijke kennis              |                 |                     |
| Taal                             |                 |                     |

Gemeente Nijmegen heeft gekozen voor T Cloud Public. Deze aanbieder kon het beste invulling geven aan een aantal technische vereisten zoals een direct connect met de on-prem omgeving en PostGIS. Meer informatie over de afweging wordt beschreven in het [“Whitepaper “Vergelijking van Europese Cloudproviders”](#).

Bij de selectie van de implementatiepartner is gekeken naar kennis, ervaring, continuïteit en prijs. Doorslaggevend was voor Nijmegen dat de partner de implementatie gezamenlijk met Nijmegen wilde oppakken en ontdekken, in plaats van een kant-en-klare oplossing te bieden. Dat laatste is voor andere gemeenten wellicht juist wel wenselijk dus dit vraagt om een bewuste keuze.

## 2. Onderdelen dataplatform

De open-source dataplatform oplossing van gemeente Nijmegen bestaat uit verschillende onderdelen die gezamenlijk het dataplatform vormen. Op elk van deze onderdelen zijn verschillende open-source producten en standaarden relevant. De volgende alinea's beschrijven op hoofdlijnen de verschillende onderdelen en de keuzes die gemeente Nijmegen heeft gemaakt.



*Figuur 1 Hoofdonderdelen dataplatform*

### 2.1. Datawarehouse en objectstorage

De data uit verschillende bronnen moet verwerkt en vastgelegd worden. Voor gestructureerde opslag en analyse maakt Nijmegen gebruik van PostgreSQL. Dit is een veelgebruikt database management systeem en wordt door vrijwel alle cloudhostingpartijen als managed service aangeboden. Door PostgreSQL uit te breiden met de PostGIS plugin kan de database naast administratieve data ook goed omgaan met GEO data. De data wordt via REST API's beschikbaar gesteld. Hiervoor wordt gebruik gemaakt van PostgREST.

Om ruwe data op te slaan wordt vaak gebruik gemaakt van object storage. S3 comptabiliteit wordt voor dit type storage gezien als marktstandaard en wordt wereldwijd gebruikt. Nijmegen heeft voor object storage gekozen voor een door T Cloud Public geleverde managed service. Hierbij is een afweging gemaakt tussen onafhankelijkheid, beheerlast en gebruiksgemak. Omdat het S3-compatible is, is de object store (relatief) makkelijk te verhuizen naar een ander S3-compatible oplossing, mocht die behoefte ontstaan. Er is gekozen voor gemak en ontzorging nu, met een relatief eenvoudige exitmogelijkheid.

#### 2.1.1. Dataverwerking (ETL/ELT)

Dataverwerking wordt door Nijmegen gerealiseerd door Feature Manipulation Engine (FME). Dit is een low-code ETL tool en is binnen het gemeentelijk domein redelijk bekend. Hoewel FME oorspronkelijk gericht is op het verwerken van GEO data, kan het ook goed worden ingezet voor andere datatypes. De keuze voor FME is vooral ingegeven door het feit dat binnen gemeente Nijmegen al kennis en ervaring aanwezig was met deze tooling en is dus vooral praktisch.

Er wordt nog onderzocht of naast FME ook een alternatief neergezet kan worden op basis van zogenaamde 'high code' oplossingen zoals Python of dbt. Belangrijk uitgangspunt is wel om bij het realiseren van de dataverwerkingsregels bewust te zijn van het gebruik van clouddienst-specifieke orkestratie of pipeline-diensten. Deze kunnen ervoor zorgen dat de verwerkingsregels niet of minder makkelijk gemigreerd kunnen worden naar een andere hostingaanbieder.

## 2.2. Datagovernance

Door het inrichten van goede datagovernance zorg je ervoor dat de data in het dataplatform betrouwbaar, begrijpelijk en herleidbaar is. Wat is bijvoorbeeld de bron van bepaalde data, welke partij is eigenaar en/of verantwoordelijk voor de kwaliteit? Zijn er duidelijke definities? Het opzetten en inrichten van een datacatalogus helpt om antwoorden te krijgen op deze vragen.

Nijmegen onderzocht meerdere open source datacatalogi, waarna de keuze viel op OpenMetadata. Datahub viel af, omdat een groot deel van de functionaliteit achter een paywall zit. Bij OpenMetadata speelt dat ook, maar in aanzienlijk mindere mate. Bij andere alternatieven ontbraken (voor Nijmegen) belangrijke functionaliteiten.

### 2.2.1. OpenMetadata

OpenMetadata is een open-source datacatalogus die metadata verzamelt en beheert. Metadata is de beschrijvende data over de data-assets, en die wordt middels automatische koppelingen uit verschillende applicaties gehaald. Uiteindelijk zal OpenMetadata de metadata van alle Nijmeegsedata-assets bevatten, zoals tabellen/views in het datawarehouse, ETL-processen in FME; dashboards in Tableau en kaartlagen in Kaartviewer. Het beschrijft onder andere wat de herkomst is van een object, wie de eigenaar is, een omschrijving kolommen in een tabel etc. Op basis van deze metadata wordt data beter doorzoekbaar, worden definities makkelijker deelbaar en is de impact van wijzigingen sneller in te zien. Hiervoor kan gebruik worden gemaakt van een UI of van APIs.

Onderliggend wordt gebruik gemaakt van PostgreSQL, ElasticSearch en Airflow.

### 2.2.2. Gemeentelijk Gegevens Model (GGM)

Het [GGM](#) is een logisch gegevensmodel met daarin vertegenwoordigd alle beleidsterreinen van de gemeente. Gemeente Nijmegen heeft ervoor gekozen om het datamodel van het dataplatform te baseren op het GGM. De definities uit het GGM worden ook door veel andere gemeenten gebruikt waardoor uitwisseling van gegevens binnen gemeentelijke ketens eenvoudiger wordt.

Door het datamodel te baseren op het GGM wordt het voor Nijmegen ook veel makkelijker en sneller om data uit hun softwaresystemen te importeren naar het dataplatform. Andere gemeenten die werken met het GGM hebben namelijk vertalingen (zogenoeten 'mappings') beschikbaar gesteld tussen data in veelgebruikte softwaresystemen en definities in het GGM. Door deze mappings te hergebruiken is het mogelijk om snel de relevante data uit hun softwaresystemen te selecteren en deze te mappen naar de juiste tabellen en kolommen in hun dataplatform.

## 2.3. Reporting

Reporting zorgt ervoor dat gebruikers de data kunnen gebruiken zoals het raadplegen, analyseren en visualiseren. Ook hiervoor zijn verschillende open-source tools beschikbaar. Nijmegen maakt hier nog gebruik van Tableau. Dit is geen open-source tool maar was al in gebruik bij de gemeente. Hier is dus een praktische keuze gemaakt omdat kennis en ervaring met deze tooling al aanwezig was. Er wordt wel onderzoek gedaan of Tableau vervangen kan worden door open-source tooling. Veelgebruikt hiervoor is bijvoorbeeld Apache Superset, Metabase of Grafana. Voor de analyse van Geo informatie wordt gebruik gemaakt van GeoServer en Kaartviewer.

Vanuit reporting wordt ook informatie gedeeld met andere gemeenten en bijvoorbeeld [opendata.nijmegen.nl](https://opendata.nijmegen.nl).

## 2.4. Security

Om gebruik te maken van het dataplatform moeten gebruikers inloggen met de SSO oplossing van Nijmegen. Hiervoor wordt gebruik gemaakt van Microsoft Entra ID. De rechten zijn gekoppeld aan specifieke rollen in Entra ID. De keuze voor Entra ID is vooral praktisch en zorgt ervoor dat er nog steeds een afhankelijkheid is van 'big tech'. Nijmegen wil gaan onderzoeken of hier een open-source alternatief voor kan worden ingezet.

## 3. Beheer en benodigde expertise

Een commerciële dataplatformoplossing van de grote techbedrijven biedt vaak een goed geïntegreerd ecosysteem van diensten. Dit is vaak de reden waarom veel partijen kiezen voor een dergelijke oplossing. Vrijwel alle functionaliteit is beschikbaar als managed service en werkt out-of-the-box samen. Daardoor wordt de beheerlast beperkt en kan met beperkte technische kennis een dataplatform ingericht worden.

Een dataplatform op basis van open-source is, zoals ook uit voorgaande alinea's blijkt, opgebouwd uit verschillende producten die in samenhang fungeren als dataplatform. Om een dergelijk platform goed te kunnen beheren is kennis nodig van deze producten. Daarnaast is expertise nodig over hoe de verschillende producten moeten samenwerken; de architectuur van het dataplatform. Tot slot is het van belang om de producten in samenhang te kunnen beheren. De releasecyclus van de individuele onderdelen van het dataplatform is verschillend en nieuwe versies van onderdelen zullen, voordat ze gebruikt kunnen worden, gecontroleerd moeten worden op een correcte werking in combinatie met de andere onderdelen van het dataplatform. Nijmegen heeft de beheerlast beperkt door waar mogelijk te kiezen voor managed services bijvoorbeeld voor PostgreSQL en S3-compatibele object storage. Een managed service zorgt ervoor dat de clouddaanbieder zorg draagt voor een juiste werking en onderhoud (je neemt het onderdeel af als een dienst). Daarnaast heeft Nijmegen ervoor gekozen om een specialistische partij te contracteren voor de inrichting en het beheer gedurende de eerste exploitatiejaren van het platform. Het beheer zal langzamerhand worden overgedragen aan de IT-organisatie van Nijmegen.

## 4. Risico's en aandachtspunten

Bij gebruik van open-source software om een dataplatform in te richten kunnen onder andere de volgende risico's en aandachtspunten meegenomen worden:

- Open-source software wordt onderhouden door 'de community'. Het kan gebeuren dat software na verloop van tijd niet meer wordt onderhouden en daarmee op termijn onbruikbaar wordt voor gebruik in productie. Dit risico wordt verkleind door gebruik te maken van veelgebruikte open-source software met een grote installed base/community.
- De documentatie van open-source software is van wisselende kwaliteit.
- Support, zeker van minder gangbare open-source producten, is beperkt omdat je afhankelijk bent van de community. Als er niemand is met een vergelijkbare vraag, kan je nergens op terugvallen.
- Gebruik Infrastructure as Code (IaC) om regels uitwisselbaar en onderhoudbaar te maken.

## 5. Conclusie

Deze handreiking beschrijft hoe de gemeente Nijmegen een dataplatform heeft gerealiseerd op basis van open-source en gehost bij een Europese cloudbaanbieder. Het dataplatform bestaat uit verschillende onderdelen die gezamenlijk het dataplatform vormen. Door de modulaire opbouw is het mogelijk om voor specifieke onderdelen te kiezen voor producten die het best aansluiten bij de behoefte van Nijmegen. Dit hoeft niet noodzakelijkerwijs open-source software te zijn. Zo wordt door Nijmegen gebruik gemaakt van Entra ID voor SSO en Tableau voor reporting. Dit was praktisch omdat deze tooling al door de gemeente werd gebruikt. Op een later moment kunnen deze alsnog vervangen worden voor open-source producten.

Door gebruik te maken van OpenMetadata wordt metadata over de data in het platform verzameld. Dit maakt de data beter doorzoekbaar en biedt inzichten in bijvoorbeeld de herkomst van bepaalde data of in het eigenaarschap.

Het datamodel van het dataplatform is gebaseerd op het GGM. Data uit het dataplatform kan hierdoor eenvoudiger uitgewisseld worden met andere partijen in de gemeentelijke keten. Ook zorgt het ervoor dat nieuwe bronnen sneller aangesloten kunnen worden op het dataplatform. Gebruik van een open-source dataplatform vraagt meer (technische) kennis en expertise. Naast kennis van de individuele producten moet ook getest worden zodat deze goed (blijven) samenwerken.