



Factsheet Data-as-a-product

Inhoud

Doel van factsheet.....	3
Leeswijzer	4
Datamesh topology	7
Dataproducten recursiviteit	8
Dataproduct canvas en specificaties	9
Vindbaar	13
Begrijpelijk	13
Betrouwbaar en waarheidsgetrouw	14
Adresseerbaar	14
Interoperabiliteit	14
Veilig toegankelijk	15
Datasoevereiniteit	15
Datakwaliteit	16
Referenties	17
Auteurs	17



Stichting DAMA-NL, The Netherlands, 2021. This work is licensed under the Creative Commons Attribution 4.0. To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>.

Versie 0.1
10 July 2023

1. Introductie

De socio-technische enterprise data architectuur 'Data Mesh' wint terrein bij data intensieve organisaties over de hele wereld. Data Mesh concept is gestoeld op vier principes:

1. Domein gericht decentraal data-eigenaarschap (Domain-oriented decentralized data ownership & architecture)
2. Dataproduct (Data as a Product)
3. Self service dataplatform (Self-Serve data platform)
4. Federatieve governance (Federated computational governance)

De vier principes zijn in hoofdstuk 2 verder uitgewerkt.

Deze factsheet beschrijft het principe "Data-as-a-product" van het data mesh concept. Data wordt aan gebruikers en afnemende applicaties beschikbaar gesteld in de vorm van dataproducten. Volgens de bedenker van Data Mesh concept Zhamak Dehghani van Thoughtworks¹ is een dataproduct "een autonome, voor gebruik geoptimaliseerde, gestandaardiseerde data-eenheid die ten minste één dataset (Domain Dataset) bevat, gemaakt om aan de behoeften van de gebruiker te voldoen". Organisaties ontwikkelen op basis van data waarover zij beschikken voor gebruikers beschikbare dataproducten, die tegen bepaalde condities (gebruikersvoorwaarden) afgenomen kunnen worden.

Een dataproduct wordt ontwikkeld zodat het vindbaar, begrijpelijk, betrouwbaar, interoperabel, toegankelijk, beveiligd, adresseerbaar is en business dan wel publieke waarde heeft. Het dataproduct biedt gebruikers en afnemende applicaties ook informatie over onder meer de relevante context, bruikbaarheid, kwaliteitsaspecten, houdbaarheid en de herkomst van de geleverde data. Al deze informatie is bij het dataproduct zelf aanwezig. De FAIR principes zijn richtinggevend in het maken van dataproducten en er zijn in het Data Mesh concept aspecten toegevoegd voor het dataproduct-eigenaarschap. Sinds de publicatie in 2016 zijn de FAIR² (Findable, Accesible, Interoperable, Reusable) principes in steeds grotere mate richtinggevend in datamanagement. Het doel van de principes is data en metadata machine leesbaar maken. Hierdoor wordt de toenemende kracht en invloed van computers beschikbaar voor data domein en binnen data management.

FAIR staat voor:

- Findable (Vindbaar): data vindbaar maken voor zowel mensen als machines.
- Accesible (Toegankelijk): data toegankelijk maken voor de gebruiker, door het verschaffen van informatie over de toegang.
- Interoperable (Uitwisselbaar): data wordt vaak gecombineerd met andere data en gebruikt in applicaties of informatiestromen.
- Resuable (Herbruikbaar): data hergebruik wordt mogelijk gemaakt met een goede beschrijving van de data en meta data.

Vanaf 2018 is de GO-FAIR community voor kennis, begeleiding en implementatie van de FAIR principes door de Europese Commissie opgezet, zie hiervoor de website [GO FAIR initiative: Make your data & services FAIR \(go-fair.org\)](https://go-fair.org/).

Doel van factsheet

Deze factsheet biedt organisaties een leidraad voor kunnen toepassen van één van de 4 principes van het datamesh concept: "Data-as-a-product". Het toepassen kan zijn in de interne organisatie en/of in een digitaal stelsel met ketenpartners. Tevens is de factsheet een hulpmiddel om uitleg te geven aan medewerkers in de organisatie over het principe: "Data-as-a-product".

¹ Data Mesh Thoughtworks [Data Mesh | Thoughtworks](https://www.thoughtworks.com/data-mesh)

² [Turning FAIR into reality - Publications Office of the EU \(europa.eu\)](https://ec.europa.eu/eip/eip-ai/eip-ai-fair/)

Het doel is tweedelig. Enerzijds scheppen we helderheid in de specificaties van een dataproduct. Ten tweede stellen we de voorwaarden waaraan dataproducten behoren te voldoen om aan te sluiten op het netwerk van dataproducten of te wel een digitaal datastelsel. De twee doelstellingen zijn:

Doel 1: Specificaties van een dataproduct

Doel 2: Voorwaarden aan dataproduct om aan te sluiten op een digitaal datastelsel in de organisatie en/of in een digitaal stelsel van de organisatie met ketenpartners

Leeswijzer

De factsheet bevat 3 inhoudelijke hoofdstukken:

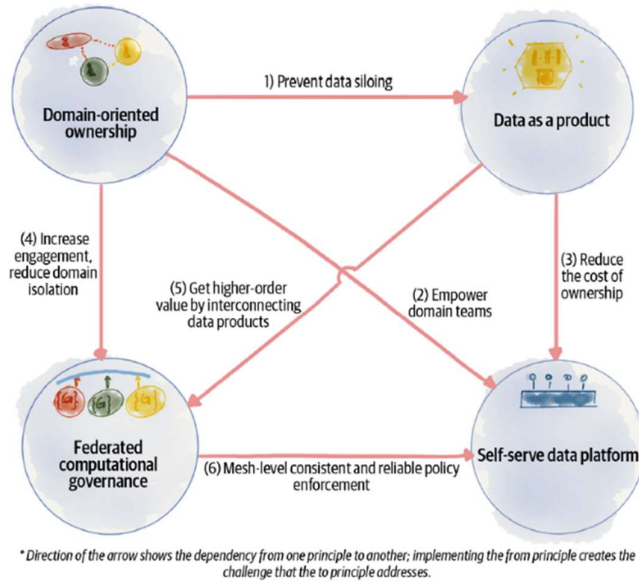
- Hoofdstuk 2 beschrijft de bredere context van het data mesh concept en de vier principes op hoofdlijnen.
- Hoofdstuk 3 beschrijft meer in detail één van de vier principes: “data as a product” en de specificaties van een dataproduct.
- Hoofdstuk 4 beschrijft de voorwaarden voor het aansluiten van een dataproduct in een digitaal datastelsel in de organisatie en/of in een digitaal stelsel van de organisatie met ketenpartners.

2. Data Mesh

Het Data Mesh³ concept is een socio-technische enterprise data architectuur met vier hoofdprincipes:

1. Domein gericht decentraal data-eigenaarschap (Domain-oriented decentralized data ownership & architecture)
2. Dataproduct (Data as a Product)
3. Self service dataplatform (Self-Serve data platform)
4. Federatieve governance (Federated computational governance)

In onderstaand figuur 2 weergegeven:



Figuur 2 Vier principes van data mesh concept (2)

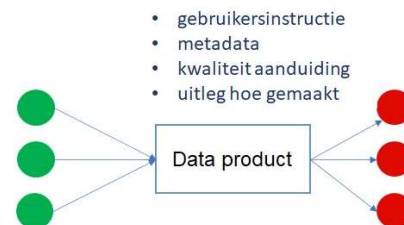
Hier volgt een toelichting van de vier principes van data mesh concept.

1. Domein gericht decentraal data-eigenaarschap

Het concept gaat ervan uit dat organisaties, afdelingen of teams in organisaties – gegeven hun businessfunctie – verantwoordelijk zijn voor het beheeren en beschikbaar stellen van bepaalde data. Data mesh respecteert deze decentraal belegde verantwoordelijkheden en het daaraan verbonden data-eigenaarschap. Data mesh gaat uit van een federatie van samenwerkende partijen.

2. Data als een product

Het dataproduct is een op zichzelf staande eenheid (atomair) en bevat materiedata (operationele data, analytische data of modellen) inclusief metadata, semantiek, ontstaansgeschiedenis, kwaliteit en 'houdbaarheid'. Het dataproduct kan een technische interface hebben, en we onderscheiden daarbij 3 patronen: batch, API of event driven. Data kunnen betrekking hebben op het verleden, het heden of een mogelijke toekomst.



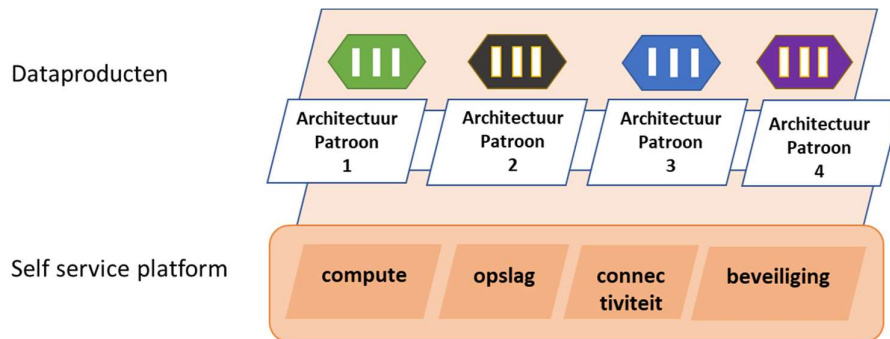
Figuur 3 dataproduct principe van data mesh concept

³ Data Mesh Thoughtworks [Data Mesh | Thoughtworks](https://www.thoughtworks.com/data-mesh)

3. Self service platform

Data mesh gaat uit van een federatieve infrastructuur in een digitaal (data)stelsel van een organisatie of datastelsel(s) van organisatie met ketenpartners. Het self service platform heeft betrekking op de functionaliteit, capabilities om dataproducten door autonome teams te ontwikkelen én te beheren.

In onderstaand figuur is een globale schets weergegeven van het self service platform met functies, capabilities in architectuurpatronen voor dataproduct ontwikkeling en beheer. ²



Figuur 4 self service platform met dataproducten

Capabilities die onderdeel zijn van een self service platform zijn en in architectuur patronen beschikbaar zijn bijvoorbeeld:

- Opslag: analytische gegevensopslag in datalake, datawarehouse, datalakehouse
- Compute: kunnen ontwikkelen datapipelines om gegevens in batch- en streaming te verwerken met rekenkracht
- Compute: Kunnen bevragen van gegevens, in datastreams, datalake of SQL-omgevingen
- Connectiviteit: kunnen verbinden voor data uitwisseling met andere dataproducten
- Beveiliging: kunnen beveiligen van gegevens, code van de datapipelines met beveiligingsmaatregelen zoals toegangsbeheer, anonimiseren, pseudonimiseren, encryptie technologie, Trusted Third Parties, etc. op basis van kaders en richtlijnen voor gegevensverwerking

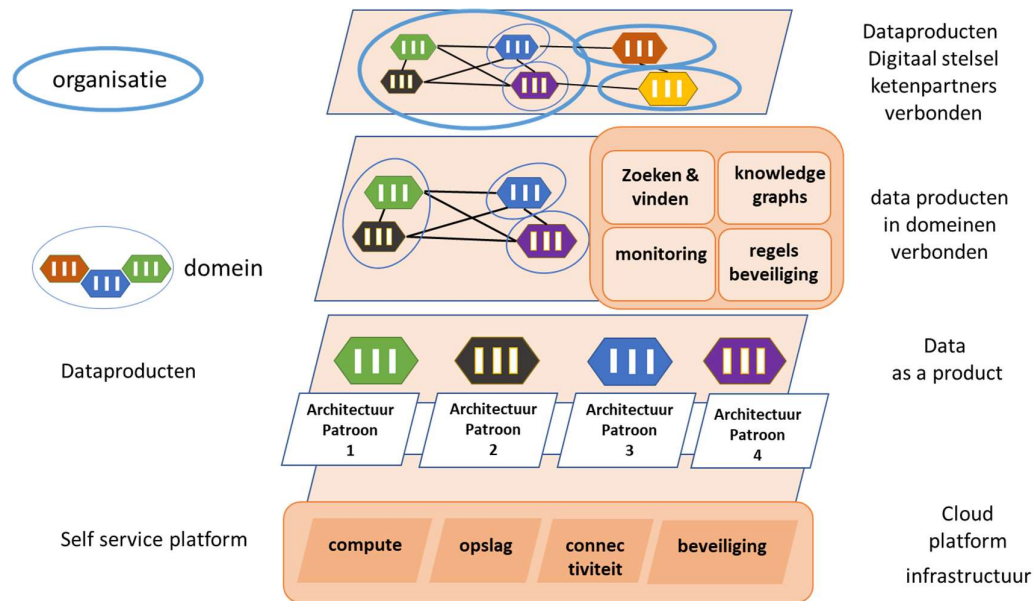
Afhankelijk van de benodigde heden om een dataproduct te ontwikkelen wordt een architectuurpatroon gebruikt om bijvoorbeeld de datapipeline met compute (rekenkracht), beveiliging maatregelen en connectiviteit met bronsystemen te realiseren en te beheren.

In het digitaal datastelsel van een organisatie of in een digitaal datastelsel tussen organisaties is ook functionaliteit, capabilities⁴ zijn nodig voor bijvoorbeeld het vinden van de gewenste dataproducten, data van dataproducten veilig en betrouwbaar te kunnen uit te wisselen en het reguleren van de toegang tot dataproducten.

Capabilities die onderdeel zijn van een digitaal (data)stelsel zijn bijvoorbeeld:

- Federatieve catalogus (*zoeken & vinden*)
- Proxy model interoperabel data uitwisseling
- Identiteit, authenticatie en autorisatie (*regels beveiliging*)
- Monitoren beschikbaarheid en datakwaliteit (*monitoring*)
- Vertrouwde data uitwisseling (*regels beveiliging*)
- masterdata verrijken, begrippen en definities (*knowledge graph*)
- Kennisgraaf en datamodelling domein overstijgend (*knowledge graph*)
- Compliance aan wet- en regelgeving en intern beleid

⁴ Data Sharing Coalition; Data Sharing Canvas [Data Sharing Canvas - Data Sharing Coalition](#)



Figuur 5 Samenhang self service platform en generieke capabilities in digitaal stelsel

Toelichting op bovenstaand figuur 5. In een organisatie zijn diverse dataproducten verbonden in een digitaal stelsel (datamesh) en de organisatie heeft met ketenpartners in een ander digitaal stelsel dataproducten verbonden voor data uitwisseling. In een organisatie zijn een of meerdere domeinen met dataproducten verbonden. Een digitaal stelsel heeft diverse functionaliteit ('zoeken & vinden', 'verbonden knowledge graphs', 'monitoring', 'regels beveiliging' zodat data van dataproducten uitgewisseld kan worden. De dataproducten zijn ontwikkeld en worden beheerd op een self service platform volgens architectuur patroon(nen).

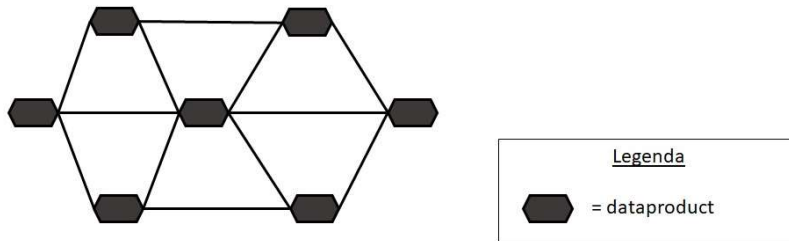
4. Federatieve governance

Om het gebruik van de dataproducten door gebruikers goed te kunnen laten werken zijn een aantal (stelsel)afspraken en is een beheerorganisatie voor de federatieve componenten nodig. De (stelsel)afspraken wordt gemaakt in een organisatie, maar ook tussen organisaties in het geval organisaties samenwerken in een digitaal stelsel. Zo zijn onder meer (stelsel)afspraken nodig over de toe te passen informatiemodellen, kennismodellen (knowledge graphs), de semantiek, de open standaarden en dataformaten van dataproducten en datacommunicatie. Daarnaast zijn afspraken en eigenaarschap nodig over de werking en het beheer van de federatieve componenten in de organisatie.

Datamesh topology

Het toepassen van data mesh is impliciet gebruik maken van een mesh topology. Dit is een vermaasd netwerk⁵ waarin de data van dataproducten uitgewisseld worden tussen aanbieder en afnemers van data. Conceptueel is een mesh een graph, een netwerk, bestaande uit knooppunten en verbindingen, elk knooppunt is een dataproduct. Het digitaal stelsel van een organisatie, maar ook een digitaal stelsel tussen organisaties maakt gebruik van een mesh topology. In onderstaand figuur weergegeven:

⁵ Vermaasd netwerk [Vermaasd netwerk - Wikipedia](#)



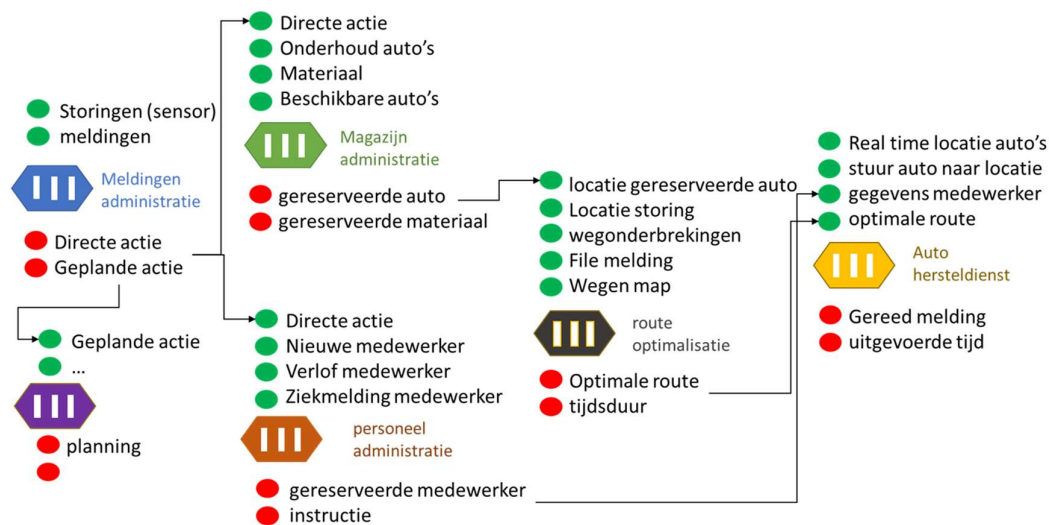
Figuur 6 Data Mesh

Dataprodukten recursiviteit

Dataprodukten kunnen recursief toegepast worden. Met recursiviteit van dataprodukten bedoelen we dat een dataprodukt gebruikt maakt van de data output van een of meerdere andere dataprodukten, het is een opeenvolging (aaneenschakeling) van dataprodukten waarvan elk afzonderlijk gebaseerd is op een of meer data output(s) van voorgaande dataprodukten.

Wat is een recursief systeem?

Recursie⁶ (Latijn: recurrere, 'teruglopen') is het optreden van een opeenvolging van constructies waarvan elk afzonderlijk gebaseerd is op een of meer soortgelijke voorgaande constructies. Doorgaans verschilt de volgende constructie in waarde van de voorgaande en is er een beginpunt. In onderstaand figuur een voorbeeld:



Figuur 7 Recursiviteit van dataprodukten

⁶ Recursie: [Recursie - Wikipedia](https://nl.wikipedia.org/wiki/Recursie)

3. Specificaties Dataproduct

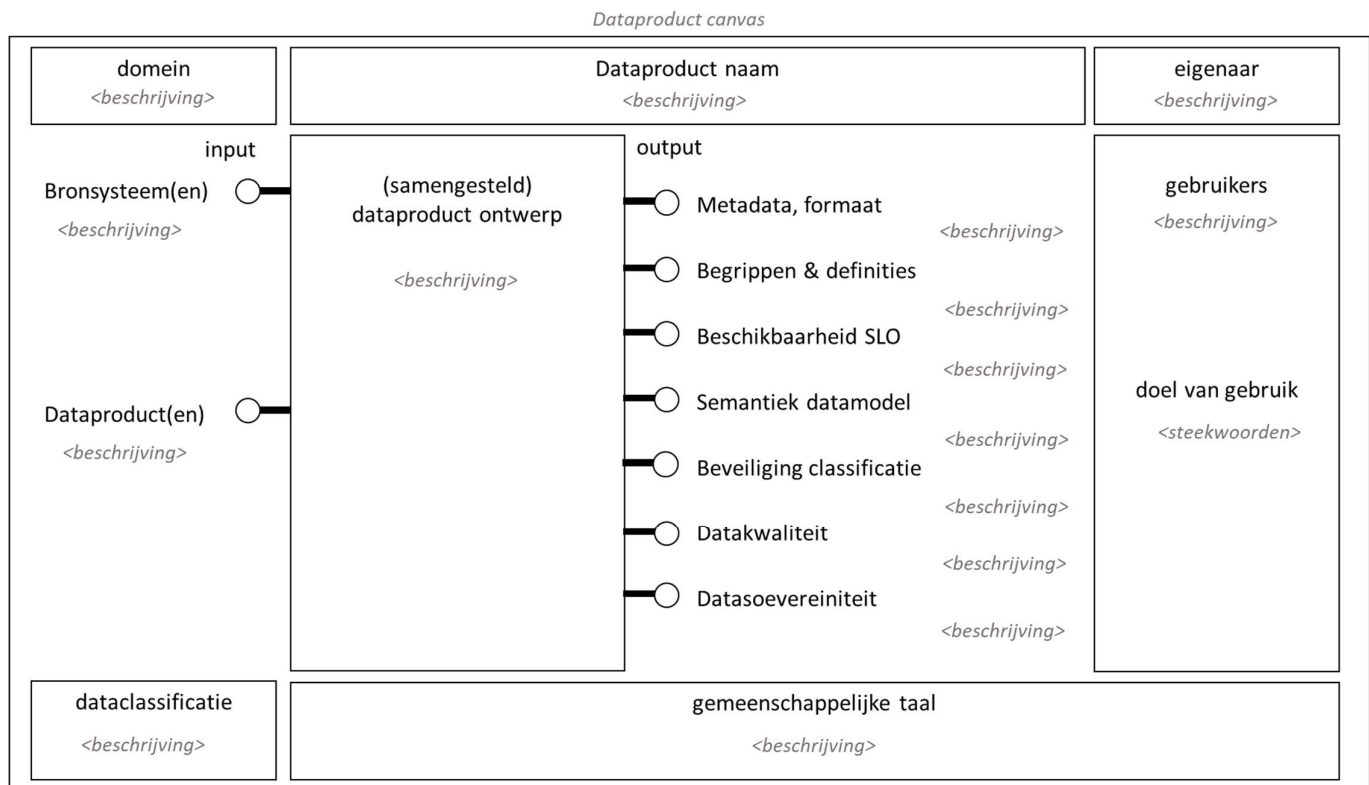
Zoals in de inleiding is beschreven is een dataproduct "een autonome, voor gebruik geoptimaliseerde, gestandaardiseerde data-eenheid die ten minste één dataset (Domain Dataset) bevat, gemaakt om aan de behoeften van de gebruiker te voldoen". Elk dataproduct bestaat binnen een begrensde context en een dataproduct kan een of meerdere databronnen maar ook een of meerdere dataproducten als bron gebruiken (recursiviteit). Recursiviteit van een dataproduct is een dataproduct dat gebruikt maakt van de output van een of meerdere dataproducten. Een dataproduct wordt opgenomen in een digitaal stelsel (met datamesh topology) van de interne organisatie en/of een of meerdere digitaal stelsel(s) tussen organisaties, zodra het aan de aansluitvoorwaarden voldoet.

Dataproduct canvas en specificaties

Voor implementatie van dataproducten en kunnen aansluiten op een digitaal stelsel, hebben we een canvas-sjabloon⁷ ontwikkeld, het is het raadzaam om gebruikt te maken van dit canvas met daarin de specificaties van een dataproduct.

Canvas-sjabloon voor dataproducten

Het onderstaand canvas bevat de specificaties van een dataproduct waarmee het ontwerp van dataproducten gemaakt kan worden. De output-kenmerken van een dataproduct zijn tevens de aansluitvoorwaarden om dataproducten aan te sluiten op het digitaal stelsel en verbonden worden in een data mesh.



Figuur 8 Dataproduct canvas

⁷ Bron: [Data Mesh Architecture: Designing Data Products \(datamesh-architecture.com\)](https://datamesh-architecture.com/)

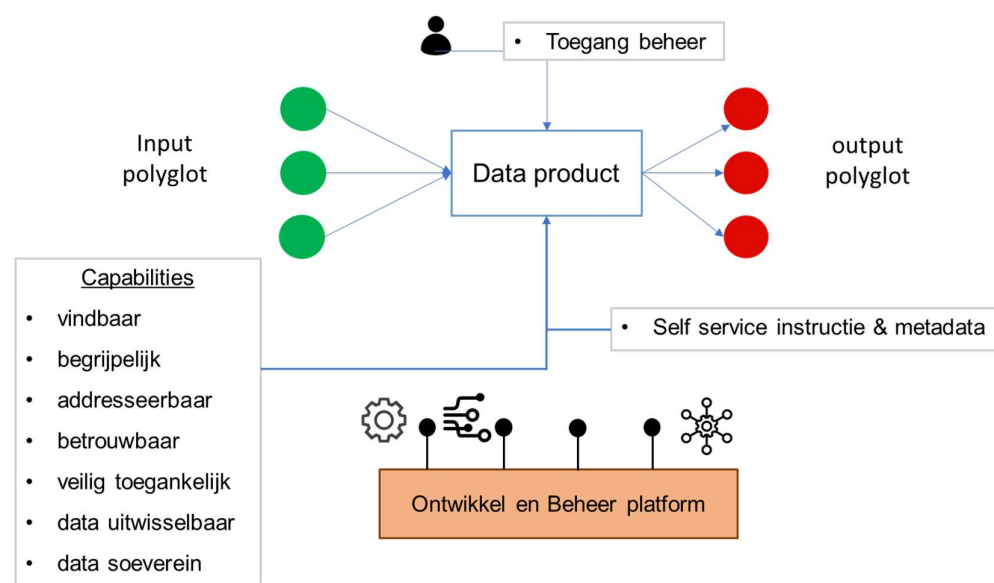
Het data mesh concept heeft een socio-technisch perspectief, naast de technologie zijn juist ook de gebruikers en de businessprocessen belangrijk. De gebruikers, de dataproduct consumenten, verwachten een bepaalde kwaliteit van de dataproducten. Data van dataproducten moet toegankelijk zijn en gebruikers moeten dataproducten eenvoudig kunnen vinden, data kunnen interpreteren en gebruiken. De leveranciers bepalen de kwaliteit van dataproducten en houden rekening daar waar het mogelijk is en het bekend is, met de behoeften van de gebruikers. Het socio-technische perspectief is de filosofie van het 'product denken' voor data - dataproducten.

Wat is een dataproduct?

Het Data-as-a-product principe introduceert een nieuw concept: dataproduct. Twee belangrijke aspecten van dataproducten zijn de bruikbaarheid en de waarde die ze hebben voor de gebruikers van de dataproducten.

Een dataproduct bevat onder andere materiedata (operationele data, analytische data of modellen) en metadata, de semantiek, de ontstaansgeschiedenis, kwaliteit- en 'houdbaarheid' kenmerken.

Een dataproduct kan door gebruikers – onder bepaalde voorwaarden – eenmalig of meerdere keren worden afgenomen.



Figuur 9 Data-as-a-product principe

Dataprodukten kunnen op allerlei manieren of te wel architectuurpatronen, worden gerealiseerd waarbij ze inputdata verwerken naar output data. In het algemeen kan elke gegevensverwerking die waarde heeft voor de gebruikers een goede kandidaat zijn. Hieronder een lijst met voorbeelden van mogelijke dataproducten:

- Bewerkte of onbewerkte ongestructureerde data zoals bijvoorbeeld afbeeldingen of video's voorzien van metadata
- Gegevensstroom van data entiteiten uit een transactiesysteem
- Gegevensstroom van gebeurtenissen die betrekking hebben op wijzigingen die zijn aangebracht in een transactiesysteem

- Gegevensstroom waarbij data bewerkt wordt door rekenregels, algoritmen, machine learning-modellen

Het bouwen en onderhouden van dataproducten is een financiële investering en neemt de menscapaciteit van het domeinteam in beslag. Daarom is het belangrijk dat de publieke of business waarde samen met de kosten vooraf worden geëvalueerd in een businesscase met besluitvorming door producteigenaar voor het wel of niet realiseren.

Een dataproductcanvas is een middel om een dataproduct specificatie te maken door het ontwikkelteam. Hieronder beschrijven we de specificaties van een dataproduct waaruit het bestaat.

SPECIFICATIES VAN EEN DATAPRODUCT

Onderstaand de specificatie van een dataproduct

Domein

Elk dataproduct mag slechts door één domeinteam worden geïmplementeerd, ontwikkeld en onderhouden. Elk dataproduct behoort tot één domein.

Dataproduetnaam

- Elk dataproduct heeft een unieke naam en de naamgevingsstrategie van de organisatie volgen.

Dataproduet eigenaar in een domein

- Elk dataproduct heeft een eigenaar die moet worden geïdentificeerd.

In deze specificatie zijn de volgende vragen relevant:

- Wie kunnen invloed uitoefenen op de richting van het dataproduct?
- Door wie wordt het dataproduct gefinancierd?
- Wie specificereert de eisen?
- Wie beantwoordt vragen over het dataproduct?
- Wie repareert het dataproduct als de data levering stagneert?
- Wie is verantwoordelijk voor datakwaliteit?
- Is er een duidelijke technisch verantwoordelijke aangesteld?
- Wie is verantwoordelijk voor wettelijke vereiste archivering, privacy, en retentie van data?

Gebruiker en use case(s)

Deze specificatie beschrijft de bestaansreden van het dataproduct. De data eigenaar maakt het data product makkelijk bruikbaar, maar hoeft niet de use cases te kennen, die zijn namelijk in andere domeinen. De data eigenaar moet wel bepalen of de data gebruikt mag worden voor bepaalde use case (soverein) De gebruiker van het toekomstige dataproduct kan in hetzelfde domeinteam zijn of afkomstig van een ander domein of buiten de organisatie.

Input dataproduet

Deze specificatie beschrijft de invoerdata voor het toekomstige dataproduct. De input is een ontvangstmecanisme voor gegevens dat het dataproduct zal bewerken. De input specificaties definiëren het formaat en protocol waarin gegevens kunnen worden ingelezen. De input kan afkomstig zijn van een (of meerdere) operationele bronsystemen en/of een of meerdere andere dataproducten, die intern kunnen zijn of uit andere domeinen kunnen komen.

Dataproductontwerp

Dit is de specificatie waar we de binnenkant van het dataproduct ontwerpen. Dit is de plaats om alles tussen de invoer- en uitvoerpoorten te specificeren. Het ontwerp omvat alles wat nodig is om een dataproduct te realiseren, bijvoorbeeld gegevensopname uit bronsysteem(en), opslag, transport, wrangling, opschoning, datatransformatie(s), dataverrijking, berekening, aggregatie, etc.

Gemeenschappelijke taal, alomtegenwoordige taal

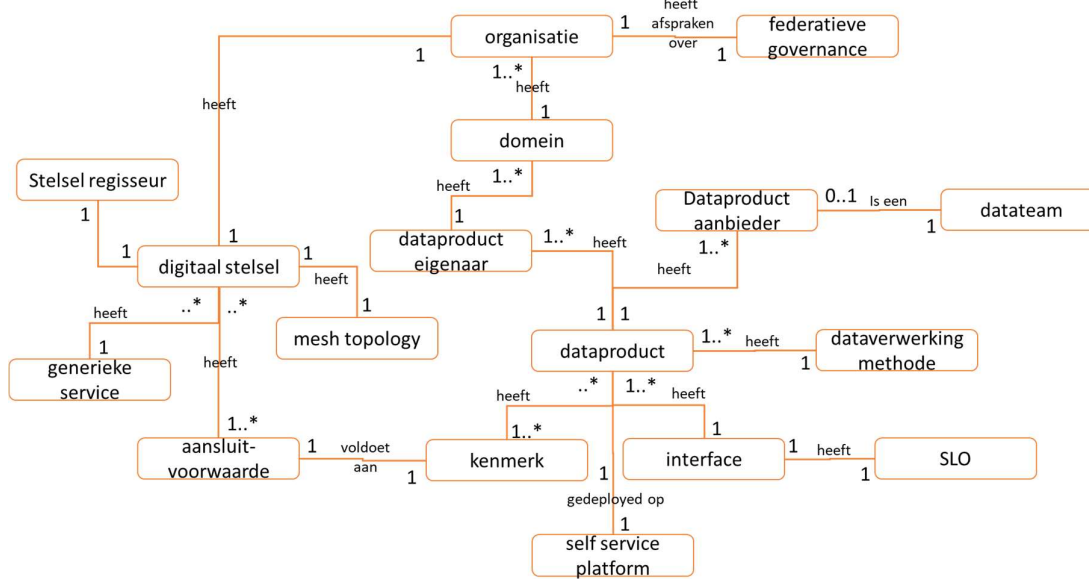
Beschrijf hier een gemeenschappelijke taal die wordt toegepast in het dataproduct. Dit is een context specifieke domeinterminologie die relevant is voor het dataproduct.

Dataclassificatie

De dataclassificatie van de data die verwerkt wordt in het dataproduct en de beveiligingseisen en -maatregelen die genomen moeten worden, wordt bepaald in een DPIA Detail Privacy Impact Assessment. De DPIA wordt eerst uitgevoerd voordat de verwerking plaatsvindt. Tevens wordt ook vermeld of het publiekelijk open dan wel gesloten data betreft. In een aantal gevallen moet eerst een ethisch assesment uitgevoerd te worden en/of een IAMA Impact Assessment Mensenrechten, bijvoorbeeld in het geval een dataproduct gebruik maakt van een algoritme voor gegevensverwerking.

Output dataproduct

De output van een dataproduct definiëren het formaat en het verbruiksprotocol waarin gegevens kunnen worden weergegeven. De output kan bijvoorbeeld een databasetabel, bestand, API of visualisaties zijn. De output omvat ook het data model. Dus: het formaat, het logisch informatiemodel en het semantisch model. In onderstaand figuur is een metamodel gemaakt waarin de relaties zijn weergegeven (zie bijlagen met de begrippen en definities van dit model):



Data entiteit	relatie	Data entiteit	Data entiteit	relatie	Data entiteit
organisatie	heeft (1 of meerdere)	domein	dataproduct	heeft 1	self service platform
domein	heeft (1 of meerdere)	dataproduct eigenaar	organisatie	heeft 1	digitaal stelsel
dataproduct eigenaar	heeft (1 of meerdere)	dataproduct	digitaal stelsel	heeft 1 of meerdere	organisatie
dataproduct	heeft (1 of meerdere)	verwerking methode	digitaal stelsel	heeft meerdere	generieke service
dataproduct	heeft meerdere	kenmerk	digitaal stelsel	heeft meerdere	aansluitvoorwaarde
dataproduct	heeft (1 of meerdere)	interface	kenmerk	voldoet aan 1	aansluitvoorwaarde
interface	heeft 1	SLO	organisatie	heeft afspraken	fed. governance
dataproduct	Heeft 1	Dataproduct aanbieder	Dataproduct aanbieder	Is geen of 1	datateam

4. Aansluitvoorwaarden dataproduct

In deze paragraaf zijn de dataproduct **kenmerken** of te wel **aansluitvoorwaarden** uitgewerkt met als doel inzicht te bieden aan welke voorschriften moet dataproduct voldoen om opgenomen te kunnen worden in een digitaal stelsel van een organisatie en/of tussen organisaties (en opgenomen worden in de eventueel aanwezige datamarktplaats of datacatalogus). De aansluitvoorwaarden zijn bedoeld zodat een dataproduct: 'vindbaar', 'begrijpelijk', 'betrouwbaar', 'adresseerbaar', 'interoperabel' (uitwisselbaar), 'veilig toegankelijk' en 'data soeverein' is in een digitaal stelsel.

Vindbaar

Vindbaarheidsinformatie wordt verstrekt door het dataproduct in de vorm van metadata. De Dataproduct aanbieder moeten de vindbaarheidsinformatie op een standaard manier delen. Een standaard tav 'hoe de data beschreven moet worden' of te wel de structuur van de metadata, is heel belangrijk. Als ieder zijn eigen model hanteert, dan is het niet vindbaar en interpreteerbaar. Een gebruiker uit een ander domein kan dan deze metadata per keer opvragen, cachen in een eigen 'domein catalogus', dan wel in een 'enterprise catalogus' vastleggen. Elk dataproduct deelt continue metadata zoals zijn oorsprong, eigenaren, runtime-informatie zoals tijdigheid, kwaliteitsstatistieken, voorbeeldgegevenssets en eventueel ook informatie die is bijgedragen door de datagebruikers, zoals de belangrijkste gebruiksscenario's en toepassingen die door de gegevens mogelijk worden gemaakt. De vindbaarheidsinformatie van een dataproduct is opgenomen in een federatieve catalogus en index en die kan bevraagd worden door datagebruikers (afnemers).

Aansluitvoorwaarden

- internationale metadata standaard Dublin Core Metadata Initiative (DCMI). [Externe link: Dublin Core Metadata Initiative](#)
- Gestandaardiseerde metadata OWMS Overheid.nl Web Metadata Standaard; Dublin Core; bron: [standaarden.overheid.nl](#)
- Metadata Standards Catalog ISO19115; [ISO 19115 – Metadata Standards Catalog \(bath.ac.uk\)](#)
- DCAT; [DCAT | Data overheid](#) Vanuit de Europese Unie is er een toepassing profiel van DCAT opgesteld. Het Nederlandse toepassingsprofiel van DCAT is gebaseerd op versie 1.1 van het EU-profiel (meer over DCAT-AP-EU 1.1 van de EU ([Externe link](#))). Er wordt continu gewerkt aan een update van het DCAT-AP-EU.⁸
- GeoDCAT geospatial extensie van DCAT [GeoDCAT-AP - Version 2.0.0 \(semiceu.github.io\)](#)

Begrijpelijk

Het dataproduct is zelfbeschrijvend. De semantiek van dataproduct is inzichtelijk, zoals de begrippen, definities, datamodellen met relaties tussen data entiteiten, waardenlijsten, etc. Naast begrip van de semantiek, moeten datagebruikers begrijpen hoe de gegevens aan hen worden gepresenteerd. Inzichtelijk is hoe datagebruikers toegang hebben, hoe ze de data syntactisch kunnen opvragen en hoe ze de data-objecten kunnen lezen. Dit gaat idealiter vergezeld van voorbeelddatasets en voorbeeldcodes. Het moet eenvoudig zijn om als datagebruiker ervaringen te delen die weer zorgen voor betere ondersteuning op bruikbaarheid. Dit vraagt ook inzicht in business rules en de fysieke implementatie daarvan. Begrijpelijk voor de mens maar ook voor de machine.

“Begrijpelijkheid” is gekoppeld aan het logisch informatiemodel van het dataproduct, hierin is beschreven welke dataentiteiten met definities en eigenschappen gemodelleerd zijn. Het logisch informatiemodel is leesbaar moeten voor diverse soort gebruikers van de data. Éen manier om dit te doen is de entiteiten van het logisch informatiemodel te koppelen aan de begrippen uit een Business glossary of thesaurus. Het logisch informatiemodel beschrijft de structuur van de data (zoals bijvoorbeeld primary/foreign keys, relationships/constraints, etc). Het logisch informatie model is gerelateerd aan een business data model dat de semantiek beschrijft.

⁸ DCAT standard W3C: [Data Catalog Vocabulary \(DCAT\) - Version 2 \(w3.org\)](#)

Aansluitvoorwaarden

- Beschrijving van begrippen en definities in Business Glossary of Thesaurus, SKOS standaard
- Duiding van de samenhang data entiteiten met datamodel:
 - Gebruik van landelijke datamodellen conform landelijke conventies
 - Standaard voor datamodelleren: Metamodel informatiemodellering (MIM) voor informatiemodellen)⁹
 - Europese en Nederlandse standaard informatie modellen (bijvoorbeeld model: "SAREF" SAREF Portal (etsi.org) in Nederland bijvoorbeeld; IMBOR, GWSW, etc.)
- Duiding van de samenhang data entiteiten met ontologie (linked data):
 - Gebruik van landelijke ontologiën conform landelijke conventies¹⁰
 - Standaarden: JSON-LD, SparQL op basis van RDF of OWL
- NEN-EN-ISO- 19110; ISO 19103; ISO 19109
- De route bewerkingsgang (datapipeline) als metadata bij dataproduct en is reproduceerbaar middels uitleg welke keuzes zijn gemaakt in realiseren dataproduct, bijvoorbeeld data aggregatie
- Beschrijving in menselijk leesbaar formaat en in computer leesbaar formaat beschikbaar
- Functionele uitleg met bijvoorbeeld via gebruikersscenario's

Betrouwbaar en waarheidsgetrouw

Elk dataproduct heeft Service Level Objectives (SLO's) die een acceptabel kwaliteitsniveau garanderen en worden nageleefd door de dataproduct eigenaar en producent. Een dataproduct kan meerdere interfaces hebben en per interface kan er een andere SLO zijn. Een SLO omvat onder meer: interval van verandering; actualiteit; volledigheid; statistische vorm van gegevens; ontstaansgeschiedenis, de datareis; precisie en nauwkeurigheid in de loop van de tijd; operationele kwaliteiten. Hoe beter de dataproducteigenaar deze gegevens bij zijn dataproduct vastlegt, bijhoudt en beheert, hoe meer vertrouwen de datagebruikers zullen hebben in het dataproduct zelf. De SLO is op termijn zodanig gestandaardiseerd dat de tooling van een afnemende partij de SLO als code kan lezen.

Aansluitvoorwaarden:

- Service Level Objectives (SLO's)
 - Verantwoordelijke rollen en taken zijn helder beschreven, eigenaar dataproduct, technisch beheer code, hosting eigenaar code, verantwoordelijke voor de databeschrijving
 - Beschikbaarheid en continuïteit van dataproducten
 - Functionele uitleg wordt gegeven: bijvoorbeeld via gebruikersscenario's
 - Beschreven is hoe de data bevraagd kan worden
 - De actualiteit, volledigheid, versiebeheer, kwaliteit van het product is beschreven
 - De eventuele licenties voor het gebruikt van het product is beschreven

Adresseerbaar

Een dataproduct moet een vaste en unieke identificatie hebben voor de datagebruiker om het handmatig of vanuit software te kunnen openen. De unieke identificatie is tevens als metadata opgenomen bij het dataproduct in de federatieve catalogus ten behoeve van de vindbaarheid.

Aansluitvoorwaarden:

- gebruik IRI (International resource identifier) conform de URI strategie¹¹

Interoperabiliteit

De data van dataproducten moet voor diverse toepassingen gebruikt kunnen worden door gebruikers (consumenten). Daarnaast moet het mogelijk zijn dat dataproducten met elkaar inhoudelijk combineerbaar kunnen worden. De mogelijkheid om data te correleren en in samenhang over domeinen heen in een gedistribueerde data-architectuur te gebruiken. Het volgen van standaarden en

⁹ MIM standaard: [Metamodel Informatiemodellering \(MIM\) | Geonovum](#)

¹⁰ Linked data conventies: [Linked Data - Design Issues \(w3.org\)](#)

¹¹ URI strategie: [URI-strategie | Geonovum](#)

harmonisatieregels, die het mogelijk maken om gegevens over domeinen heen zinvol in samenhang te bevragen.

Aansluitvoorwaarden:

Een dataproduct heeft een interface conform meerdere architectuurpatronen

- API-mogelijkheid: REST, JSON, WFS, SOAP, algorithm push, CityGML en CityJSON, etc.
 - Richtlijnen: API strategie en design rules: API design rules en API strategie en Forum Standaardisatie [Home](#) | [Forum Standaardisatie](#)
- Met open standaard formats wordt data aangeboden: JSON, XML, etc.
- Data technisch uitwisselbaar met dataconnecties: internet, intranet, VPN's, etc.
- Indien autorisatie verplicht: open standaard autorisatie mechanisme zoals OAuth en OpenDoc.
- Data bevindt zich in authentieke bron
- De oorsprong van het dataproduct is beschreven
- De beschrijving is in menselijk leesbaar formaat en in computer leesbaar formaat beschikbaar

Veilig toegankelijk

De dataproducten moeten zoveel als mogelijk met elkaar inhoudelijk combineerbaar kunnen worden. Aan dataproducten worden eisen gesteld ten aanzien van de beveiliging van persoonsgegevens. De Dataproduct aanbieder is transparant ten aanzien van de verwerking (verzamelen, vastleggen, ordenen, bewaren, bijwerken, wijzigen, opvragen, raadplegen, gebruiken, verstrekken en verspreiden) van persoonsgegevens teneinde met voldoende zekerheid vast te kunnen stellen of wordt voldaan aan de wettelijk gestelde eisen ten aanzien van beschikbaarheid, integriteit en vertrouwelijkheid. Dit geldt ook voor algoritme, voorspelmodellen die toegepast worden in dataproduct. Vaste en zelflerende algoritmen worden uitgelegd.

Aansluitvoorwaarden:

- DPIA Detail Privacy Impact Assessment dataproduct (Wetgeving Awb en AVG/BIO) is uitgevoerd voor dataproduct en goedgekeurd
- Ethisch assessment is uitgevoerd voor dataproduct en goedgekeurd
- Indien het dataproduct gebruik maakt van algoritmes: richtlijnen voor het toepassen van algoritmen door overheden en publieksvoorlichting over data-analyses, zoals Impact Assessment Mensenrechten en Algoritmes en de Europese verordening Risico analyse van AI-systemen hoog en registratie in algoritme register. Sinds 2023 is de AI-act van toepassing
- Indien een dataproduct een bewerking maakt van sensordata: AI-act van toepassing met verplichting van registratie in sensor register .
- Self-Sovereign Identity (SSI) ecosysteem met een Decentralised Identifier (DID) Concept en Verifiable Credentials als basis voor Self-Description. ¹²
- Standaard voor communicatie protocol: DIDComm (Decentralised Identity Communication) en OIDC (OpenID Connect)
- Decentralised Identity Framework ESSIF (European Self Sovereign Identity Framework) voor publieke- en private sector¹³

Datasoevereiniteit

Datasoevereiniteit ('data sovereignty') is het vermogen van een natuurlijke persoon of organisatie om volledig zelfbepalend te zijn met betrekking tot zijn data, d.w.z. de mogelijkheid voor een rechtspersoon om exclusief te beslissen over het gebruik van zijn data als een economisch bezit. Het vereist dat mensen en organisaties controle hebben over de omstandigheden over hoe hun data wordt gedeeld en hoe deze door andere partijen mogen worden verwerkt.

Datasoevereiniteit heeft ook als doel dat de dataproduct eigenaar is in staat om volledig zelf te bepalen wat er met zijn gegevens gebeurt in de dataproducten, of te wel de eigenaar heeft regie op de eigen gegevens. Vanuit Europa is de ontwikkeling van International Data Spaces en is er een referentiearchitectuurmodel ontwikkeld voor data soevereiniteit met specifieke benodigde functionaliteit en capaciteit en gerelateerde aspecten, inclusief vereisten voor veilige en vertrouwde gegevensuitwisseling in organisatie-ecosystemen en digitale stelsel waaraan organisaties deelnemen.

¹² SSI standaard: [Self-Sovereign Identity \(SSI\) - DBC - English \(dutchblockchaincoalition.org\)](#)

¹³ ESSIF: [European Blockchain Services Infrastructure \(EBSI\) and the eSSIF | Decentralized Identity Web Directory \(decentralized-id.com\)](#)

Datasoevereiniteit gaat over het vinden van een balans tussen de noodzaak om je gegevens te beschermen en de noodzaak om iemands gegevens met anderen te delen.¹⁴

In het geval dat in dataproducten data gebruikt wordt van eigenaren van data (burgers en consumenten) dan gaat datasoevereiniteit óók over het doel dat gebruikers van dataproducten om meer beschikking over hun eigen data te krijgen. Wettelijk is het begrip van datasoevereiniteit in Europa met de GDPR vastgelegd. Belangrijk is dat de technologische middelen en menselijke maatregelen aanwezig zijn in dataproducten zodat de eigenaren van data deze rechten ook in de praktijk kunnen opeisen. Met de ontwikkelingen van zogenaamde identiteitsplatformen zijn er mogelijkheden.

Een interessante ontwikkeling hierin zijn zogenaamde dataplatformen, die burgers helpen om inzicht en controle over hun data te behouden. Een Nederlands voorbeeld hiervan is Medmij. Bij Medmij kunnen gebruikers een account aanmaken en hun patiëntgegevens eenvoudig en veilig delen met aangesloten zorgaanbieders. Het grote voordeel voor gebruikers is dat zij op één site (Medmij) overzicht kunnen behouden van de gegevens die zij met een veelvoud aan partijen gedeeld hebben.¹⁵

Aansluitvoorwaarden:¹⁶

- Op moment van schrijven zijn er nog geen ervaringen met praktijk richtlijnen voor datasoevereiniteit aansluitvoorwaarden. In Europese International Data Spaces wordt op hoofdlijnen datasoevereiniteit beschreven.

Datakwaliteit

De kwaliteit van de data die beschikbaar wordt gesteld vanuit een dataproduct wordt door de producent of autonoom datateam actief gemonitord en continu verbeterd in op dracht van de product eigenaar. De norm van datakwaliteit is in overeenstemming met de eisen die vanuit de datagebruikers (consumenten) en wetgeving worden gesteld. Datakwaliteit sluit daarmee aan op het proces waarbinnen het gebruik gegenereerd wordt en waarbinnen het gebruikt wordt.

Voor de vastlegging van datakwaliteit zijn verschillende informatiemodellen beschreven. De W3C¹⁷ organisatie ontwierp de Data Quality Vocabulary (DQV) voor gebruik op het web. De Europese Unie publiceerde de Data Quality Management Vocabulary (DQM) als structuur om datakwaliteit in op te slaan. Beide zijn modellen zijn uitgedrukt in RDF Resource Description Framework en zijn uit te breiden naar de behoefte van de gebruiker.

Aansluitvoorwaarden:

- Toepassen van een datakwaliteit framework voor beheersing van datakwaliteit en inzichtelijk maken van de datakwaliteit(label) gerelateerd aan de afgesproken norm
- De syntactische correctheid en de integriteit van data worden bewaakt
- De kwaliteitsnormen worden afgesproken en vastgelegd de SLO (Service Level Objectives of te wel: Service Levering Overeenkomst)
- Datakwaliteit wordt door de dataproduct eigenaar regulier gerapporteerd en datakwaliteit inzichtelijk aangeboden in federatieve datacatalogus, aangeduid met een kwaliteitslabel

Opmerking: kwaliteitseisen van data kunnen worden aangegeven door datagebruikers, maar de dataproduct eigenaar bepaalt (uiteindelijk) welke data kwaliteit geleverd wordt. Bovendien is het niet altijd inzichtelijk wat de kwaliteitseisen van gebruikers zijn, zeker in het geval het grote aantal betreft (bijvoorbeeld de inwoners van een stad).

¹⁴ [Data Sovereignty and the European Data Governance Act - International Data Spaces](#)

¹⁵ Medmij: [Homepage van Stichting MedMij](#)

¹⁶ Data soevereiniteit standaard [Data Sovereignty in the Digital Decade – For a stronger digital Europe in the world | Futurium \(europa.eu\)](#)

¹⁷ W3C: [Data on the Web Best Practices: Data Quality Vocabulary \(w3.org\)](#)

Begrippen & definities

nr	begrip	Definities
1	Dataproduct eigenaar	data eigenaar een natuurlijk persoon NP of organisatie, publieke- of private instelling, kan zijn. De dataproduct eigenaar is de partij die de verantwoordelijkheid draagt voor het definiëren, aanbieden en het juist leveren van de output van een dataproduct dat door Dataproduct aanbieder (dit kan een autonoom datateam zijn) ontwikkeld wordt.
2	Dataproduct aanbieder	Een partij die een of meerdere dataproducten ontwikkeld en beheerd. Dit kan een autonoom datateams in een organisatie zijn, maar ook een leverancier
3	Rechthebbende	De rechthebbende is de partij die rechten heeft over data en toegang kan verlenen aan derden
4	Dataproduct	Een dataproduct bevat onder andere materiedata (operationele data, analytische data of modellen) en metadata, de semantiek, de ontstaansgeschiedenis, kwaliteit- en 'houdbaarheid' kenmerken.
5	Dataproduct gebruiker	De dataproduct gebruiker is de partij die de verantwoordelijkheid draagt voor het gebruiken van een aangeboden dataproduct omdat dit waarde oplevert (bv. dienstverlening verbeteren of mogelijk maken)
6	Autonoom datateam	Het zelfstandig opererend team bestaande multidisciplinaire rollen in een organisatie die een of meerdere dataproducten ontwikkeld en beheerd
7	Domein	Een domein is een organisatiedomein is een verzamelterm voor de producten en diensten die een organisatie levert aan de klanten en belanghebbenden door een of meerdere organisatie onderdelen en afdelingen. Een van de diensten kunnen dataproducten zijn.
8	Self service platform	Het self service platform heeft betrekking op de functionaliteit, capabilities om dataproducten door autonome teams te ontwikkelen én te beheren
9	Digitaal datastelsel	Omgeving waarin partijen in verschillende rollen (zoals Dataproduct aanbieder, afnemers, stelselregisseur) samenwerken en datadelen en waar ze op elkaar vertrouwen terwijl ze hun eigen zelfstandigheid houden. Digitaal stelsel heeft capabilities datadelen (geautomatiseerd) ondersteunen.
10	Interoperabiliteit	De functionaliteit om data uit te wisselen tussen partijen die elkaar vertrouwen met hun ICT-systemen.

Referenties

- DAMA International (2017). *DAMA-DMBOK Data Management Body of Knowledge, Second Edition (DMBOK2)*. Technics Publications 2017
- Data Mesh Z. Dehghani

Auteurs

Medewerkers van werkgroep Data Mesh van DAMA-NL.