



Visie op gegevensplatform

Versie 0.4

voor het structureren van het zoeken, vinden, begrijpen, vergelijken, kiezen en gebruiken van gegevens

Jaap van den Berg / Jan Willem van Veen
CIO Office 2021©



Omgaan met gegevens als een **strategisch bedrijfsmiddel** impliceert een **modern gegevensplatform** voor Gegevensmanagement, zodat belanghebbenden rijke analytische inzichten kunnen verkrijgen via Artificial Intelligence (AI) en Machine Learning.

Microsoft



Gegevensplatform

Een **modern** gegevensplatform is een (cloud)omgeving waarin samenwerking en overleg tussen verschillende (keten)partijen over het beheer en gebruik van gegevens tot stand gebracht wordt.

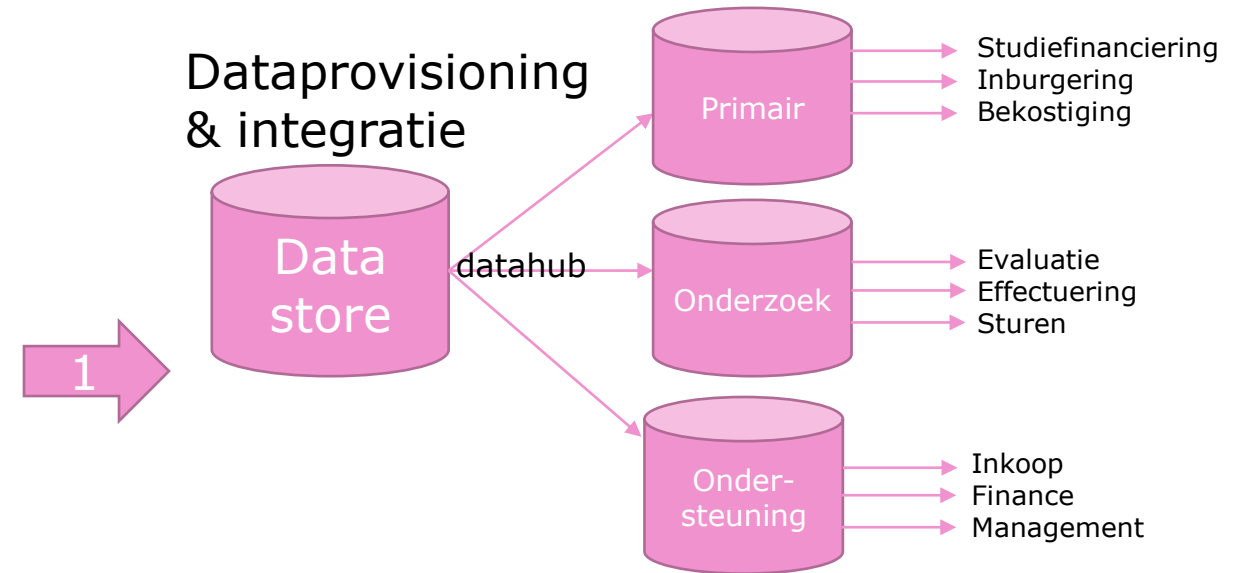
De volgende aspecten spelen hierbij een belangrijke rol :

- › **Sturing en besluitvorming (datagovernance)**
- › **Lifecycle Management (LCM), Artificial Intelligence (AI) en Machine Learning**
- › Monitoring, datakwaliteit en metadata
- › Beveiliging, schaalbaarheid en performance (cloud-first strategie)
- › Kostenmanagement



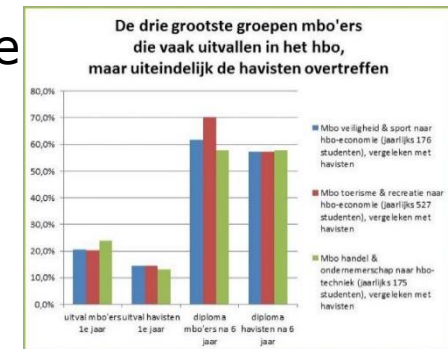
Uitdagingen gegevensplatform DUO

- Angela : interne en externe gegevensuitwisselingen niet compliant
 - 100 structureel
 - 300 incidenteel
- Nynke / Anton : wil gebruik maken van persoonsgegevens om wettelijke taak uit te voeren, weet alleen niet waar mandaat vastligt
- Willem : wil onderzoek doen (evalueren), maar onduidelijk precies waarover (zicht op cohorten, ander leenstelsel, etc.)
- Rudy : wil context / complexiteit tussen wettelijke en private taken halen uit de keten → WRO artikel 23
- Aangewezen organisaties : Hoe de belangen van de raden te behartigen in samenhang met de DUO belangen?
- DUO : Met welke technieken doen we dit?



2

Moderne Analyse Omgeving (BI)



3

AI en Machine Learning ready !





Advies Gartner : dataplatform patronen gecombineerd inzetten

Architectural Patterns Serve Different Purposes, But Are Best Used in Combination

	Data Lake	Data Warehouse	Data Hub
Repository where data persists long term	✓	✓	✗
Foundation for analytic workloads	✓	✓	✗
Primary focal point of governance	✗	✗	✓
Enables data flow between diverse endpoints	✗	✗	✓
Architectural pattern (not a product/tool)	✓	✓	✓
Best used in combination with other patterns	✓	✓	✓

Source: Gartner (February 2020)
ID: 465401_C



Metadata op basis van NORA vijflaagsmodel

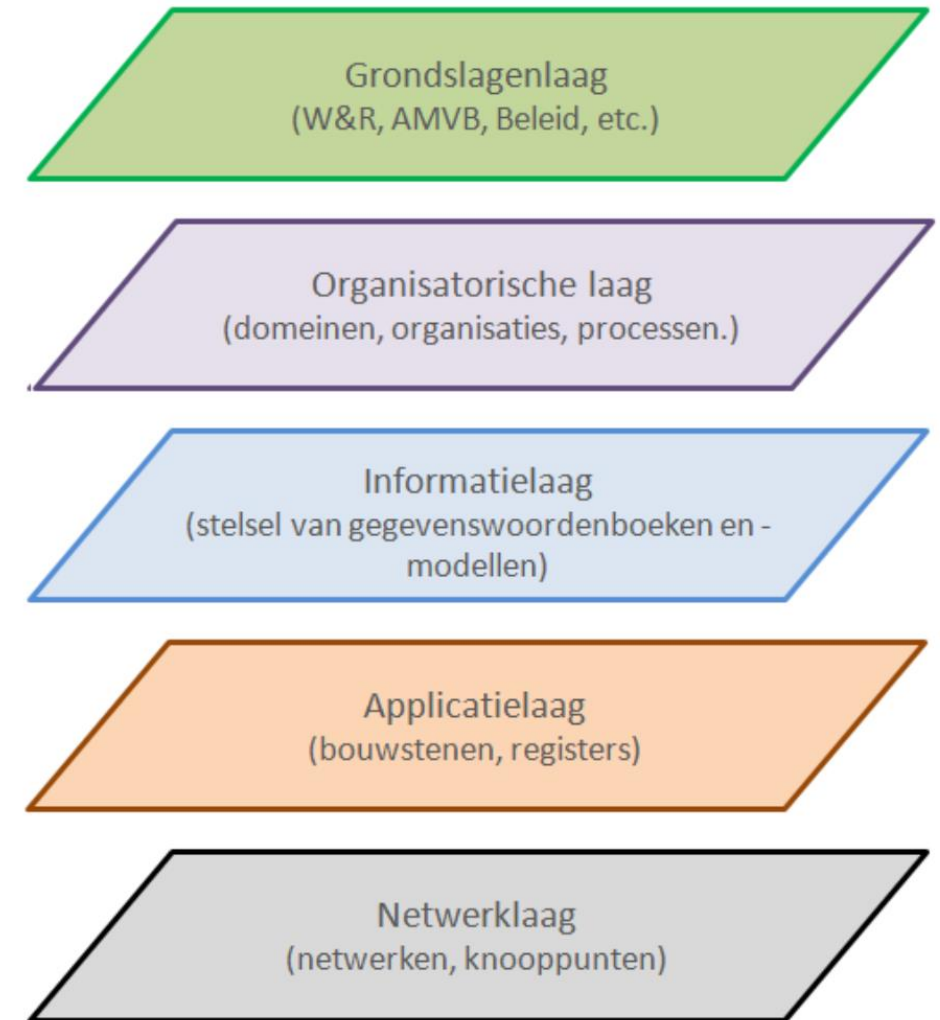
Metadata

Een belangrijk onderdeel van gegevensmanagement is metadata. In dit model worden de onderdelen van metadata in samenhang en met voorbeelden getoond.

Vijflaags model NORA

- **Laag 1** : wetten
- **Laag 2** : organisatie
- **Laag 3** : informatie
- **Laag 4** : applicatie /data
- **Laag 5** : techniek

<https://www.noraonline.nl/wiki/Vijflaagsmodel>



Voorbeelden

AVG, Onderwijsnummer, Wet op het PO, VO, etc.

Onderwijs specifiek, NORA, ROSA, HORA, FORA

DUO, Kennisnet, Onderwijsbegrippen

RWS OTL, CB-NL, NEN2660

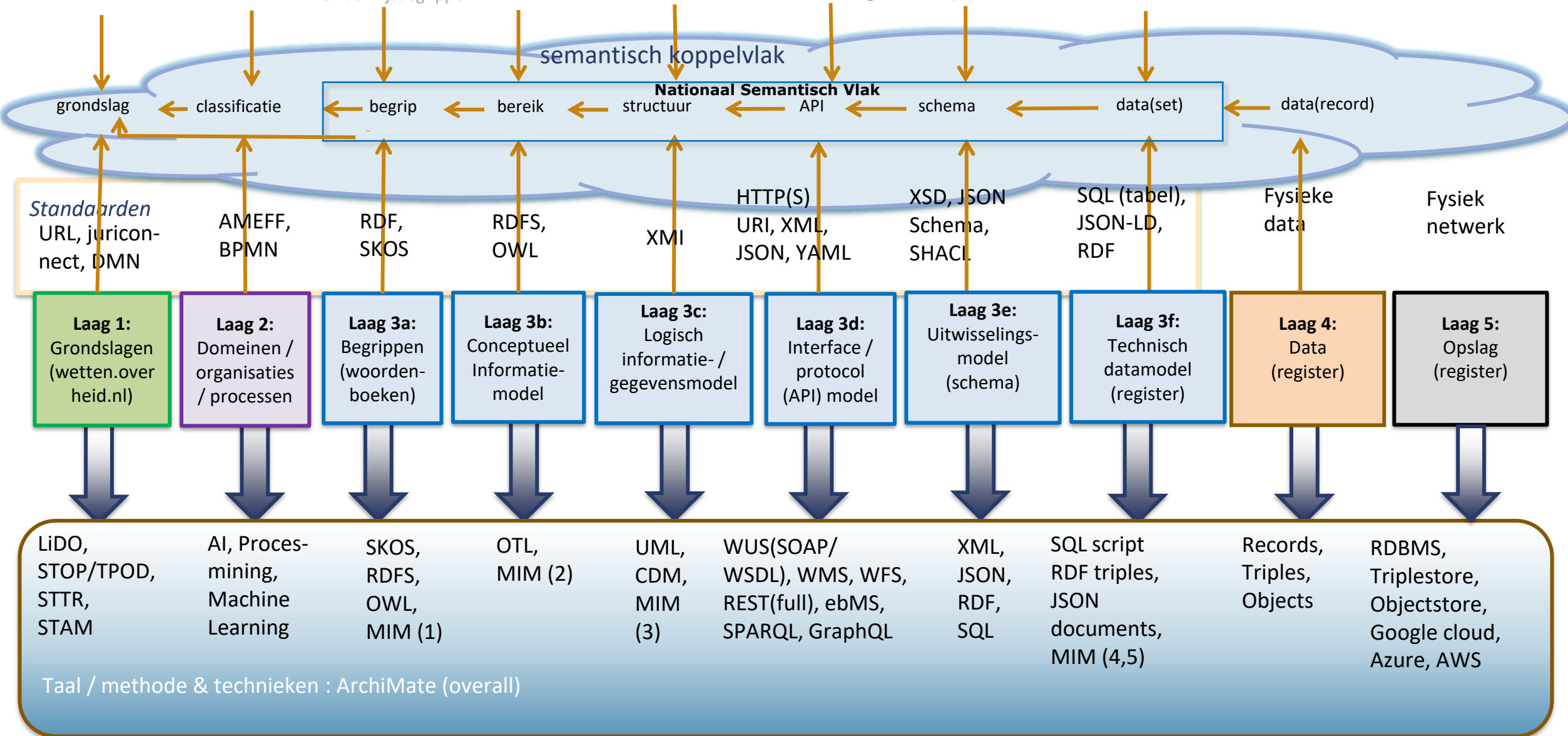
IMGEO, NEN3610
Onderwijs informatie-
modellen



OAS 3, RAML, API strategie

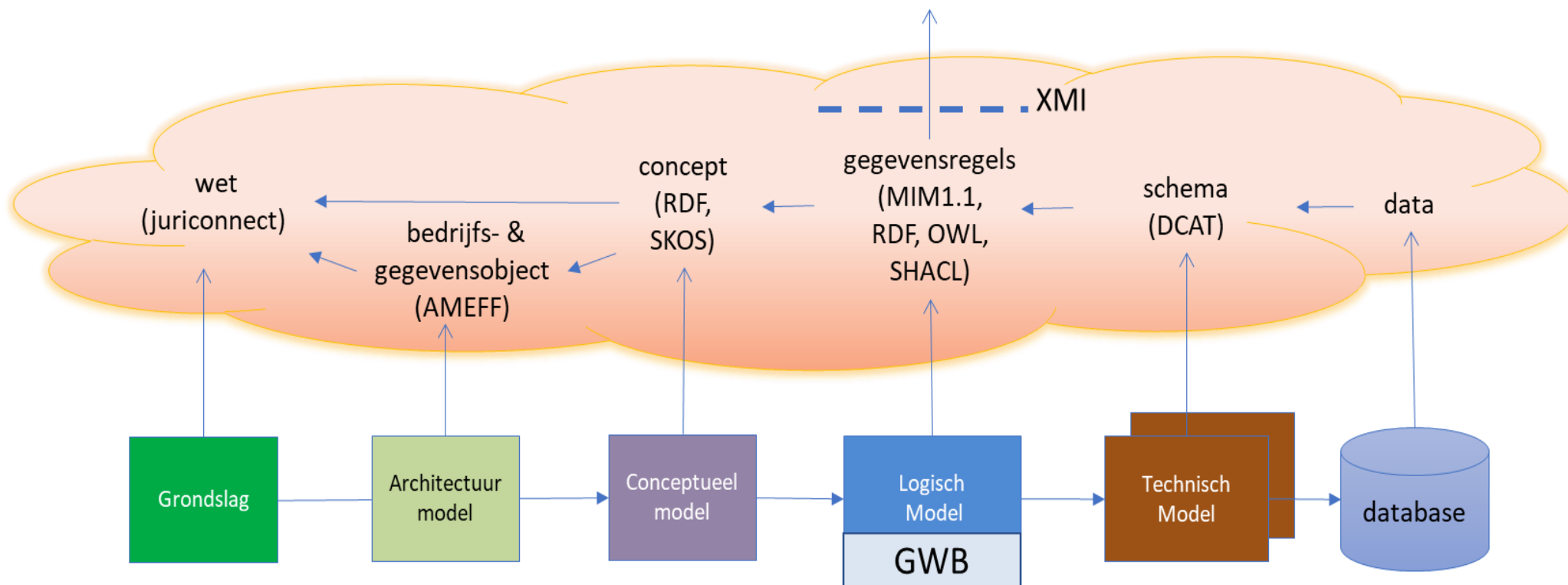
SHACL, XSD, DCAT

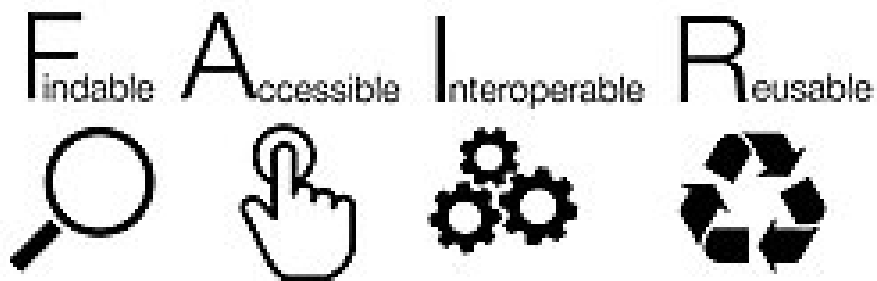
Data.overheid.nl, fysieke modellen



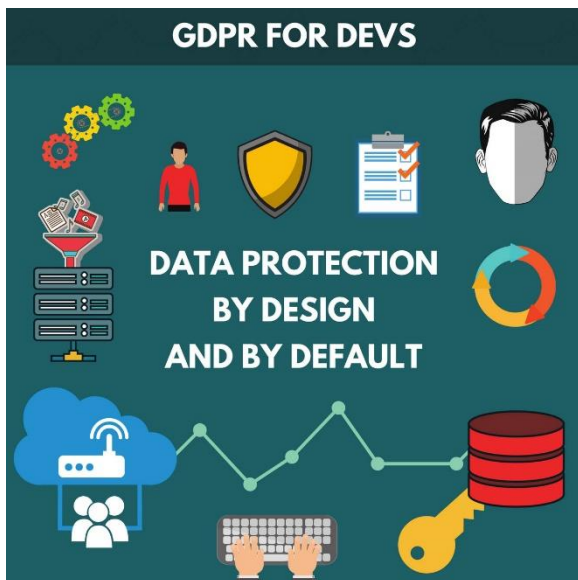


Linked Data metamodel (Gerald Groot-Roessink)

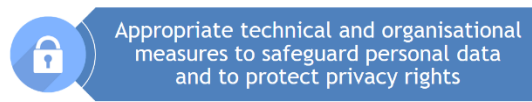




Wet
Registers
Onderwijsdeelnemers
(WRO) artikel 23 –
Verstrekking aan aangewezen
instellingen



Dataminimalisatie by design



Appropriate technical and organisational
measures to safeguard personal data
and to protect privacy rights

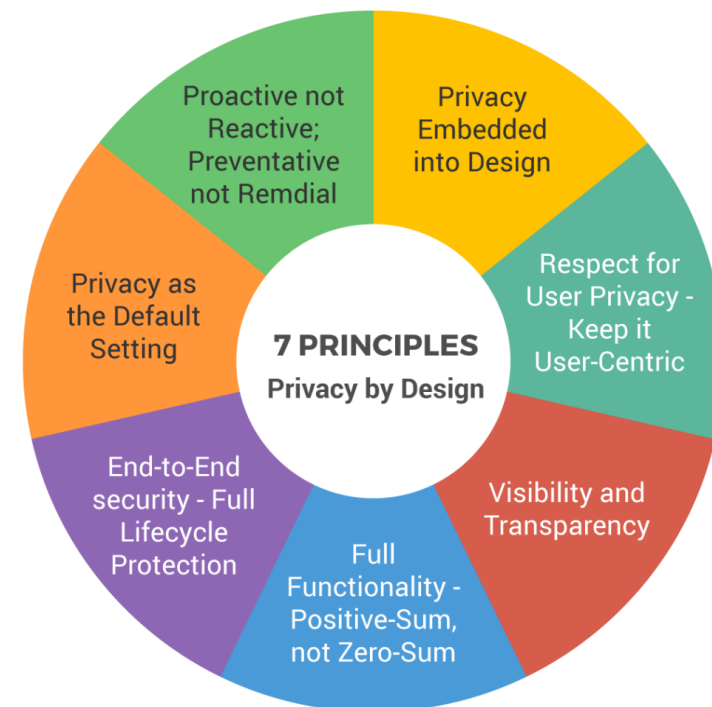
In practice: Controllers shall develop systems and processes with data protection
by design and by default

Privacy by design

- Consider privacy at the initial design stages and throughout the complete development process when personal data is processed

Privacy by default

- When choices regarding data privacy are present, the most privacy friendly one is selected by default



Data masking

Data fading

Anonimiseren

Open Data

Redigeren

Polymorfe

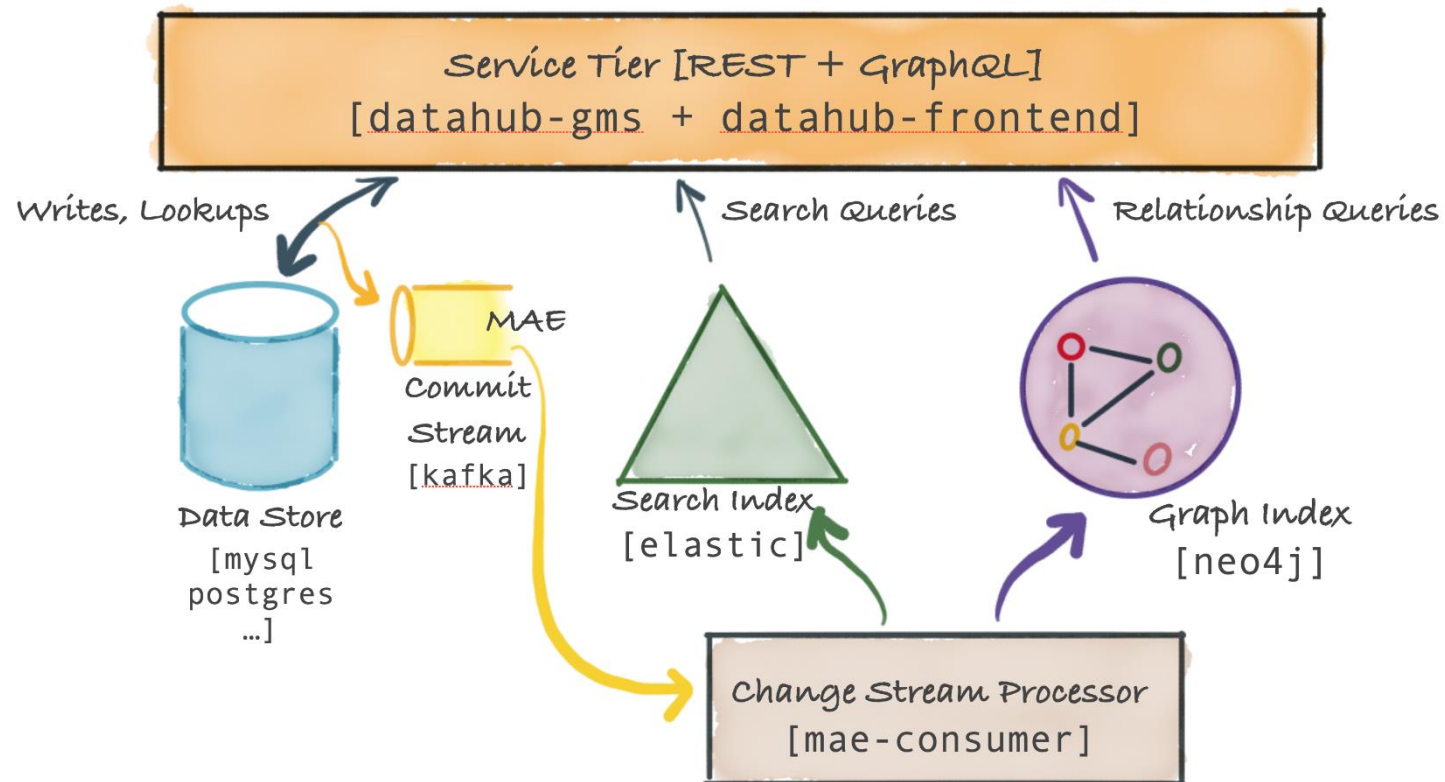
Big Data

Pseudosynonimiseren

Linked Data

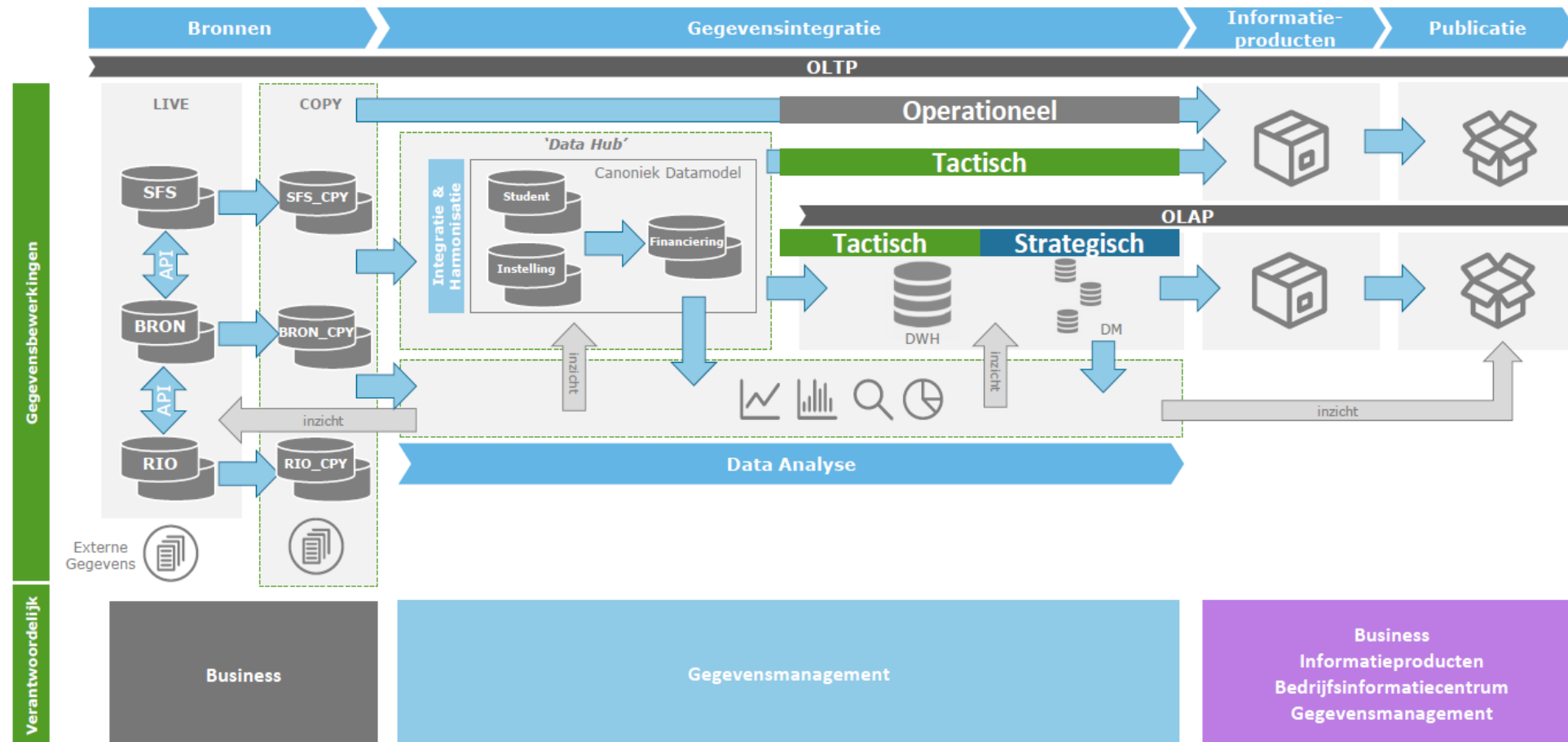


Datahub patroon (ingevuld met technologie)





Huidiaa architectuur



Nb. De bovenstaande figuur dient ter illustratie met de genoemde databronnen als voorbeeld.

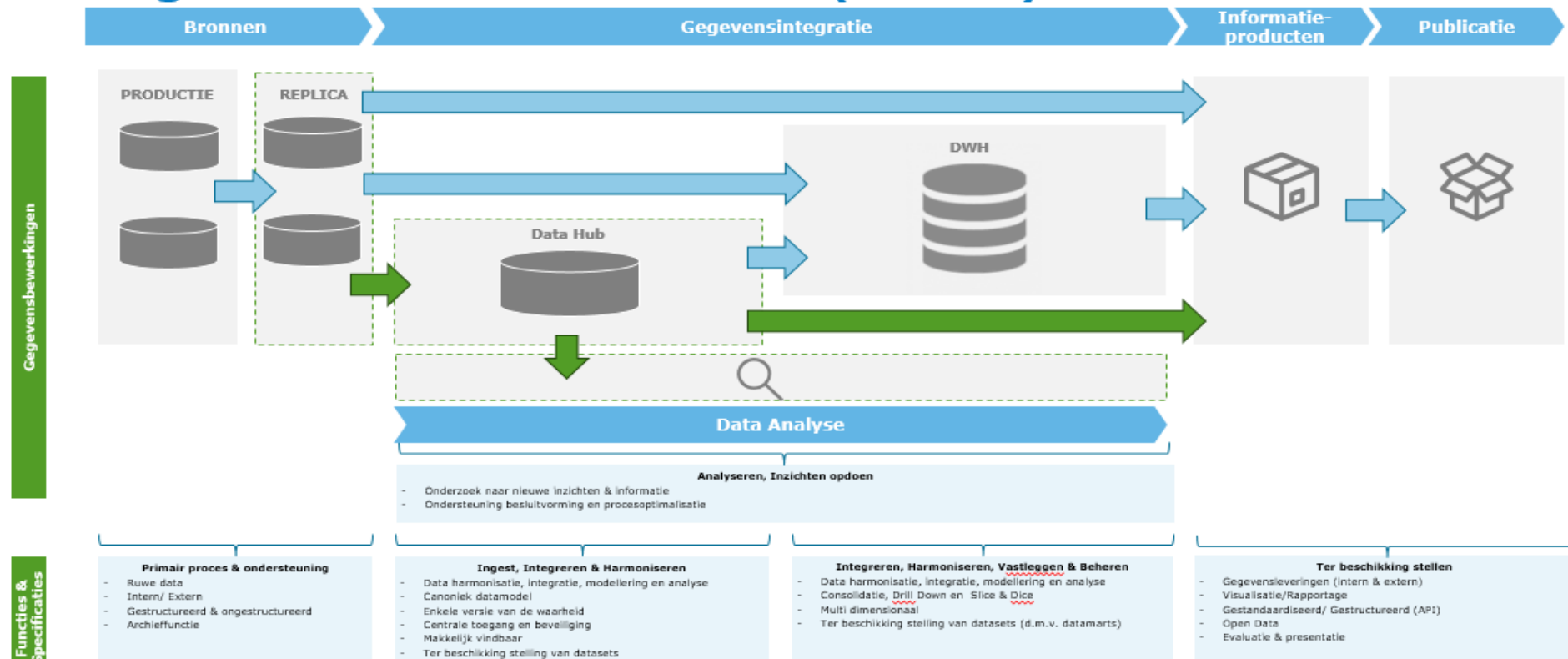


To-be architecture






➔ Scope pilot

High level architecture (to-be)

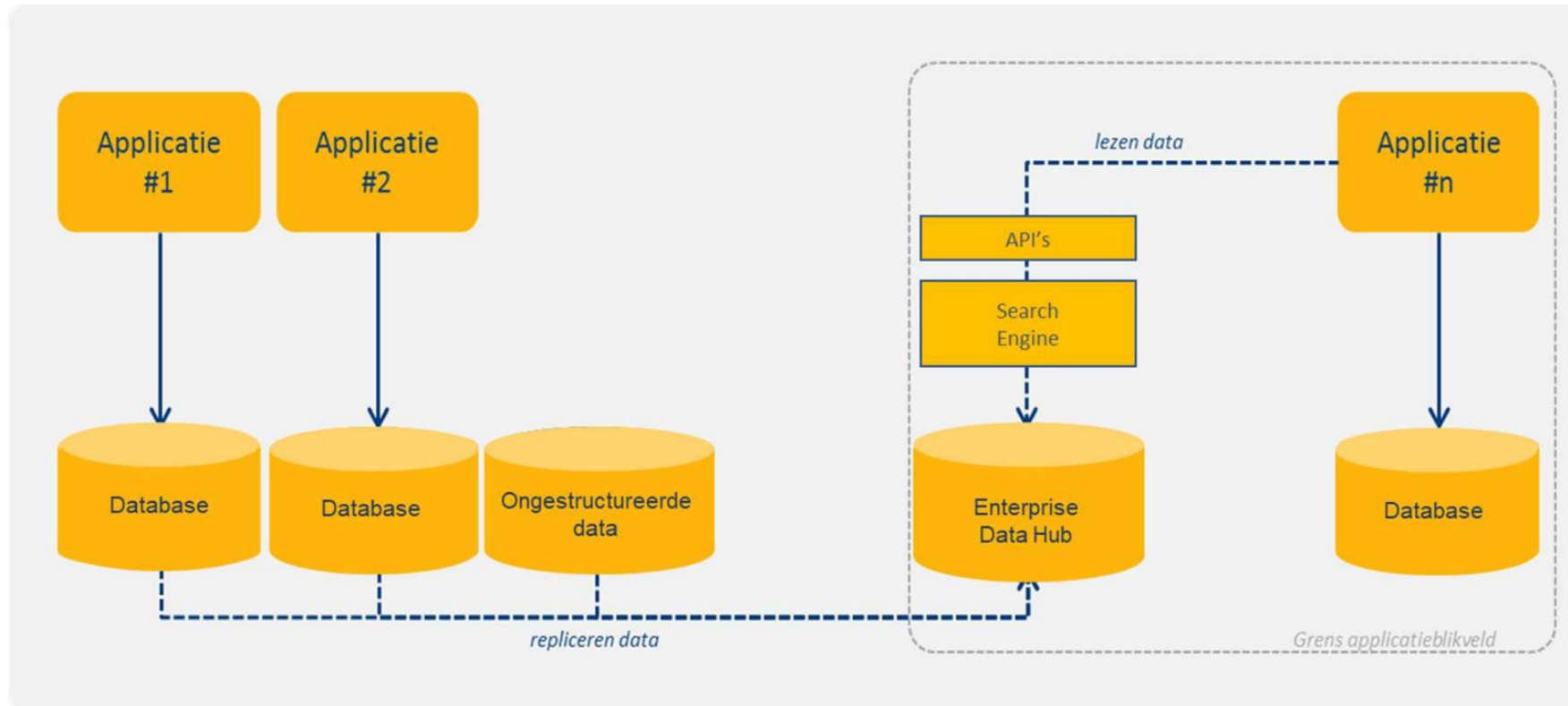


Nb. De bovenstaande figuur dient ter illustratie met de genoemde databronnen als voorbeeld.



DATA WAREHOUSE	DATA LAKE	DATA HUB
		
<ul style="list-style-type: none">• STRUCTURED FOR ANALYTICS• CONSUMED BY PEOPLE AS A SELF-SERVICE• FOCUSED ON DECISION MAKING	<ul style="list-style-type: none">• (UN)STRUCTURED FOR DISCOVERY• CONSUMED BY DATA PROFESSIONALS AND ALGORITHMS• FOCUSED ON DEEP LEARNING, AI	<ul style="list-style-type: none">• STRUCTURED FOR DATA PORTABILITY• CONSUMED BY PEOPLE AND APPS• FOCUSED ON DATA INTEGRITY AND SPEED FOR SHARING

Enterprise Data Hub



Deze vorm van applicatie-integratie is een doorontwikkeling van de operational data store.



Definities

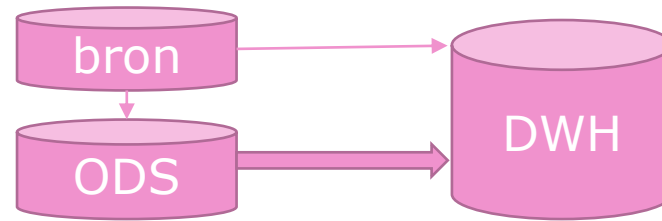
- › Een **Data Warehouse(DWH)** bestaat uit het geïntegreerd opslaan van informatie met als doel om zakelijke beslissingen en analyses te kunnen voeren. Inhoudelijk bevat een DWH **gestructureerde** gegevens.
- › Een **data lake** is een opslaglocatie waarin grote hoeveelheden ruwe data in zijn oorspronkelijke structuur zijn opgeslagen. Inhoudelijk kan een Data Lake **gestructureerde** en **ongestructureerde data** bevatten.
- › Een **data hub** slaat data zelf niet op, maar **verzorgt de flow** van data tussen bronsystemen en doelsystemen en gebruikers.
- › Een **data platform**, ook wel bekend als **data management platform**, is een geïntegreerde oplossing die de functionaliteiten van data lake, data warehouse, data hub en elementen van een Business Intelligence(BI) Platform combineert.



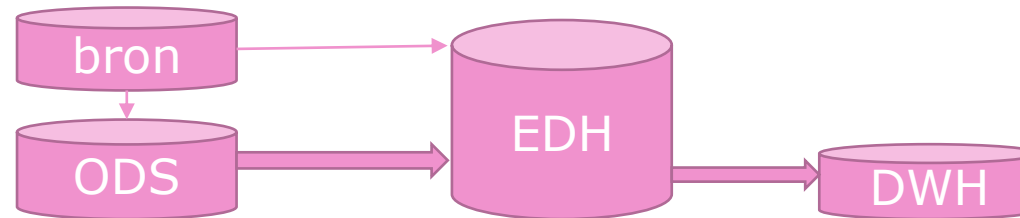


Fasering in migratieplateaus : van store via hub naar Gegevensplatform

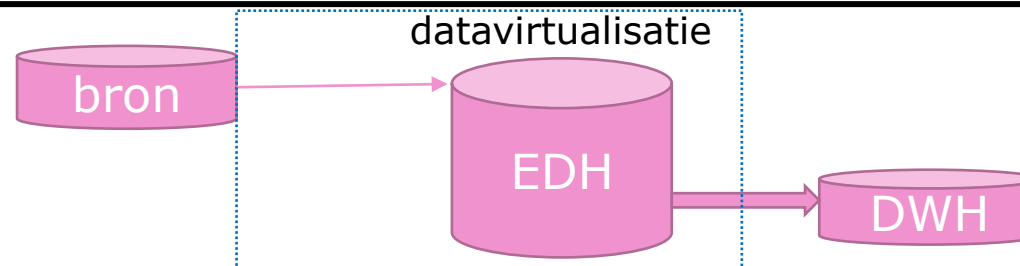
Fase 1 (=nu) : store



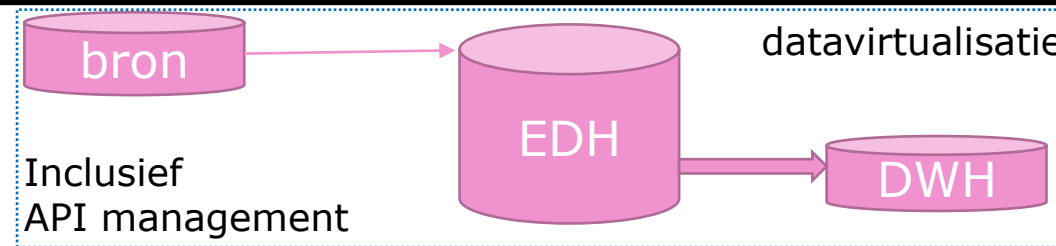
Fase 2 : ODS + hub



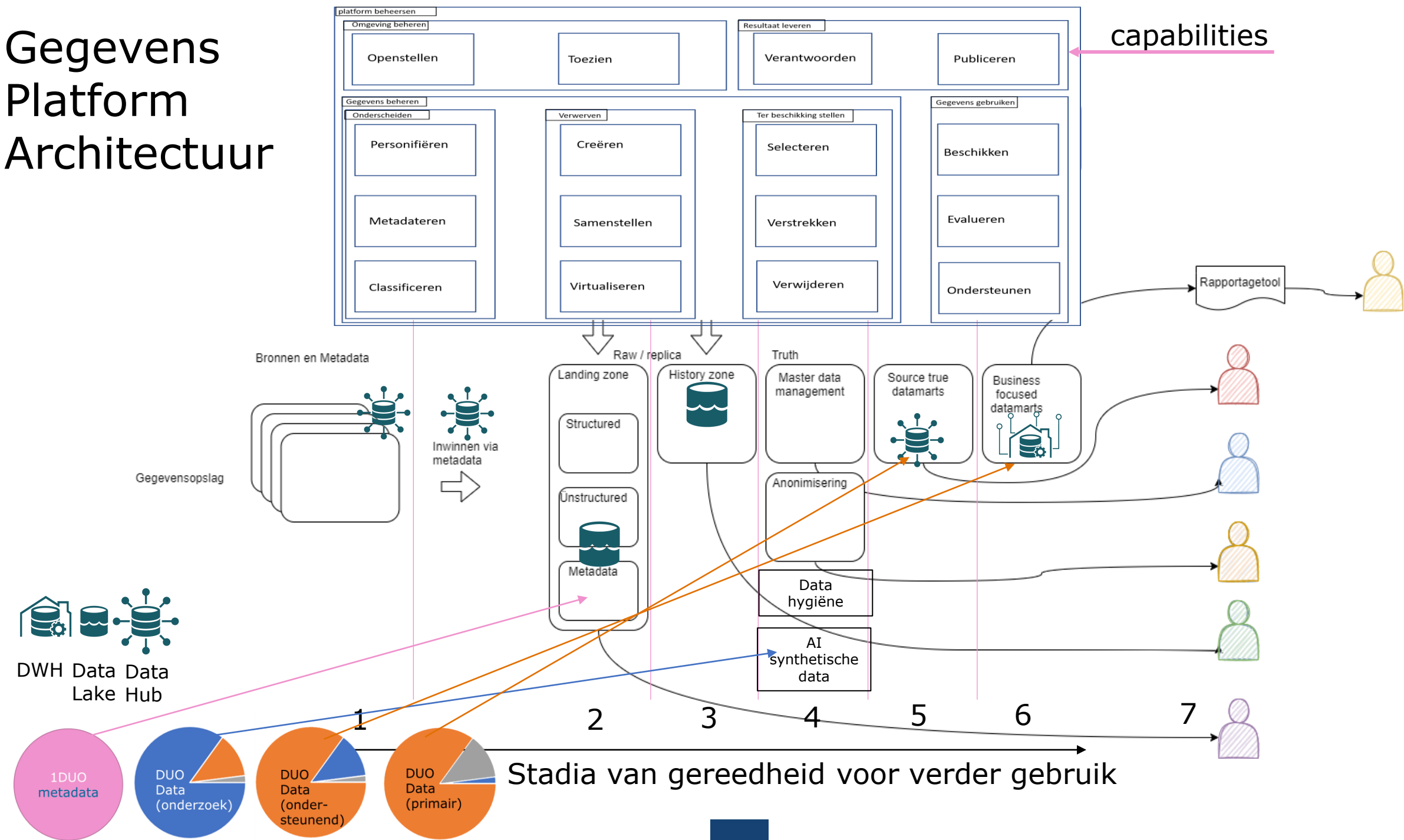
Fase 3 : Data hub



Fase 4 : Gegevensplatform



Gegevens Platform Architectuur





Veranderplateau's Onderwijsgegevensplatform

Plateau 1 : Algoritme + doel op orde
(verwerkingsregister)

Grondslagenlaag
(W&R, AMVB, Beleid, etc.)

Plateau 2 : Governance op orde
(ketenregie, beheer en gebruik)

Organisatorische laag
(domeinen, organisaties, processen.)

Plateau 3 : Metadata op orde
(verticale + horizontale datalineage)

Informatielaag
(stelsel van gegevenswoordenboeken en -modellen)

Plateau 4 : Data op orde (GRIP)
(Data + Integratie architectuur,
inrichting Datahub / proxy)

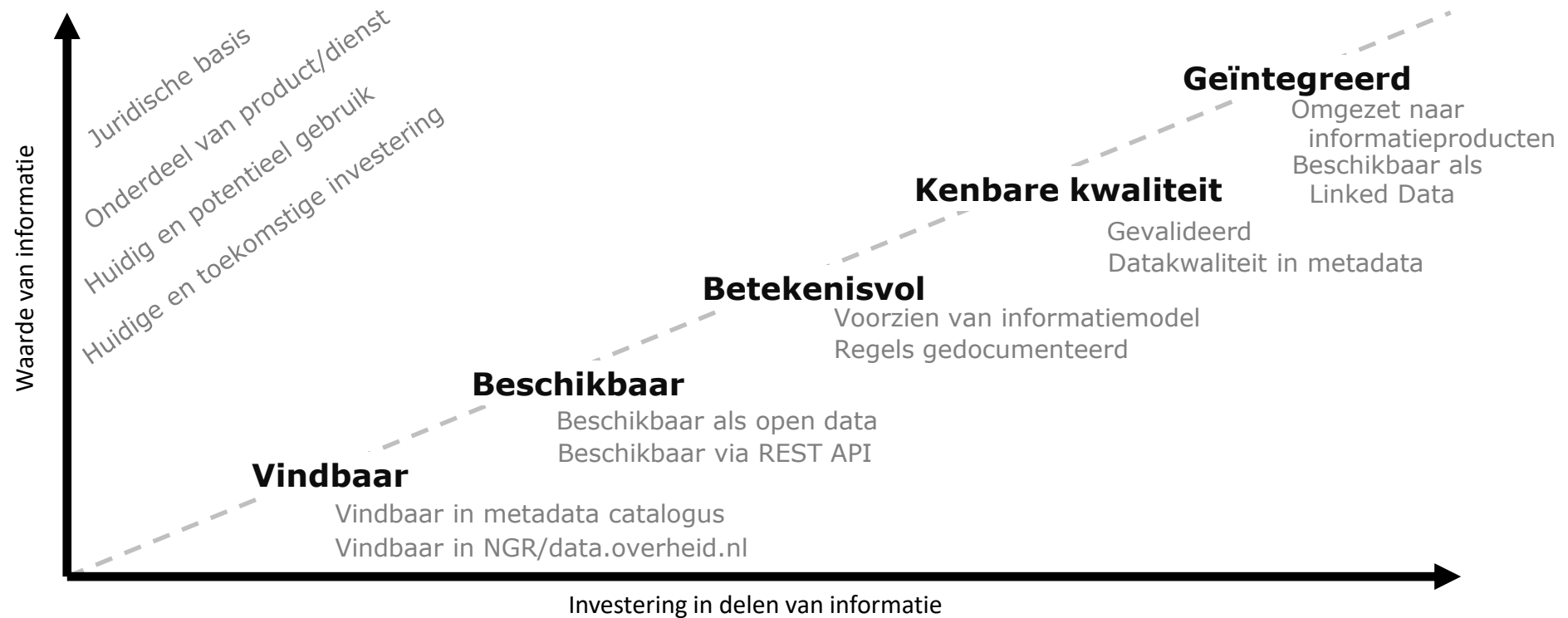
Applicatielaag
(bouwstenen, registers)

Plateau 5 : Omgeving op orde
(data platform)

Netwerklaag
(netwerken, knooppunten)



Data op orde



Proces

Stroomschema werken met CBS-microdata

