



Data Architectuur NS - Datafundament

Yorrick Tillemans

Lead Solution architect Data, Innovatie &
Analytics bij NS

www.linkedin.com/in/yorrick/

www.yorrick.consulting





Datareis NS



Historie <2016 tot Heden



Invloeden binnen en buiten NS



Nieuwe technologieën





Samen maken we Nederland bereikbaar

Voor iedereen

Nederland duurzaam
bereikbaar | Strategie |
Over NS | NS



"Sneller digitaliseren en data-gedreven werken" is een van de vijf tactische thema's binnen NS

De missie van NS op gebied van data is “Bewust beter keuzes maken met betrouwbare data”



Over NS / Nederland duurzaam bereikbaar /

Mensgericht en datagedreven

Nederland duurzaam bereikbaar maken voor iedereen, vraagt wat van de NS-organisatie en de medewerkers. We zijn wendbaar en innovatief, duurzaam en financieel gezond. Onze organisatie wordt steeds meer datagedreven, maar is en blijft mensgericht.



Vacatures Jouw interesse Leren & Werken Onze verhalen Events NS als werkgever

Q Digital > Data & Technology

Data & technology

Data oplossingen met impact

Van innovatieve toepassingen op basis van talloze datastromen tot stuurinformatie voor business intelligence en datagovernance: bij NS zetten wij vol in op datagedreven processen. Zo kom je als data professional bij de afdeling Data, Innovatie & Analyse (DIA) te werken aan modellen, algoritmes en simulaties voor uitdagende vraagstukken van de hele organisatie. Oplossingen die NS miljoenen kunnen besparen én die impact maken op heel reizend Nederland!

Vacatures Data & Technology

Help NS vooruit
Business Intelligence

Bij Business Intelligence put jij inzichten uit zo'n 150 bronssystemen: hoeveel treinen reden er op tijd? Is het juiste materieel ingezet? Zo niet, kunnen we een oorzaak aanwijzen? Jij zet data om in bruikbare stuurinformatie middels rapportages en dashboards die de business ondersteunen bij continue verbetering en het nemen van data-driven beslissingen.

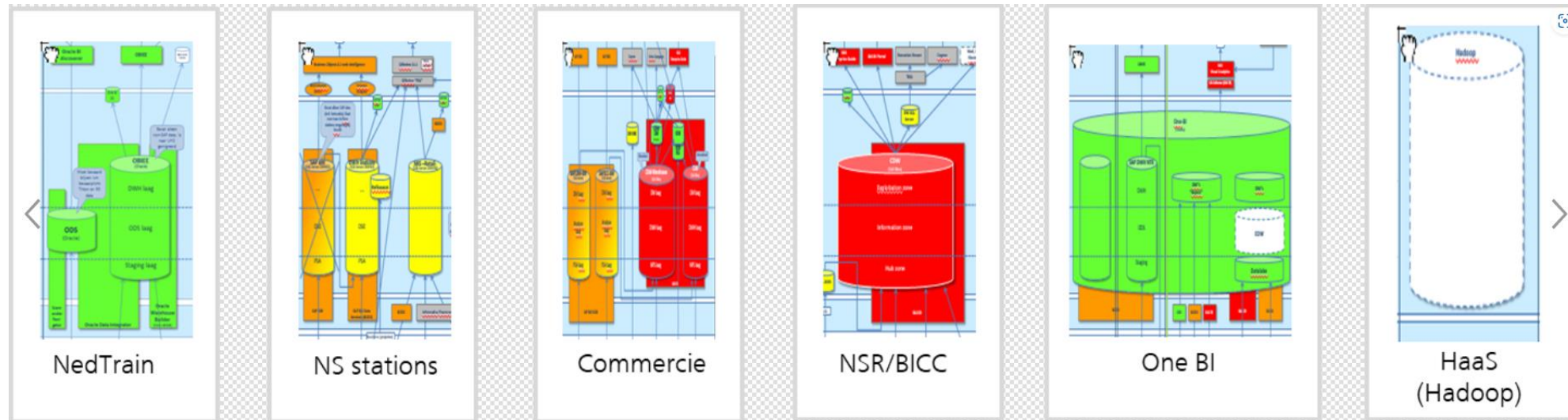
Direct impact
Advanced Analytics

Realtime worden gigantische hoeveelheden stromende data uit IoT-systemen geanalyseerd op het Hadoop-dataplatform. Bij Advanced Analytics heb jij direct impact met innovatieve oplossingen voor onder meer druktevoorspelling, een grotere beschikbaarheid van materieel en het nog punctueler rijden van onze machinisten. Valt er een trein uit? Incheckgegevens helpen bepalen hoeveel bussen nodig zijn.

De basis voor data-innovatie
Plan & Platform

Jij zorgt dat teams worden ondersteund, processen op orde zijn en er regie en samenhang is. Van ondersteuning bij de transformatie tot dataplatform diensten, datamanagement en datagovernance NS-breed. Ook zorg je dat data efficiënt en veilig bewaard wordt en vanuit diverse bronnen beschikbaar komt voor analyse en efficiënt bewerkt kan worden.

Vóór 2016



- Definities van een soortgelijk gegeven verschilden per bedrijfsonderdelen
- Gebrek aan NS-brede metadata
- Verwerking van data onvoldoende transparant, traceerbaar en reproduceerbaar
- Technische issues om data tussen omgevingen uit te wisselen
- Hoge kosten aan verschillende omgevingen en personeel
- Omgeving onvoldoende schaalbaar voor Big Data



ACE-programma (2017-2021)

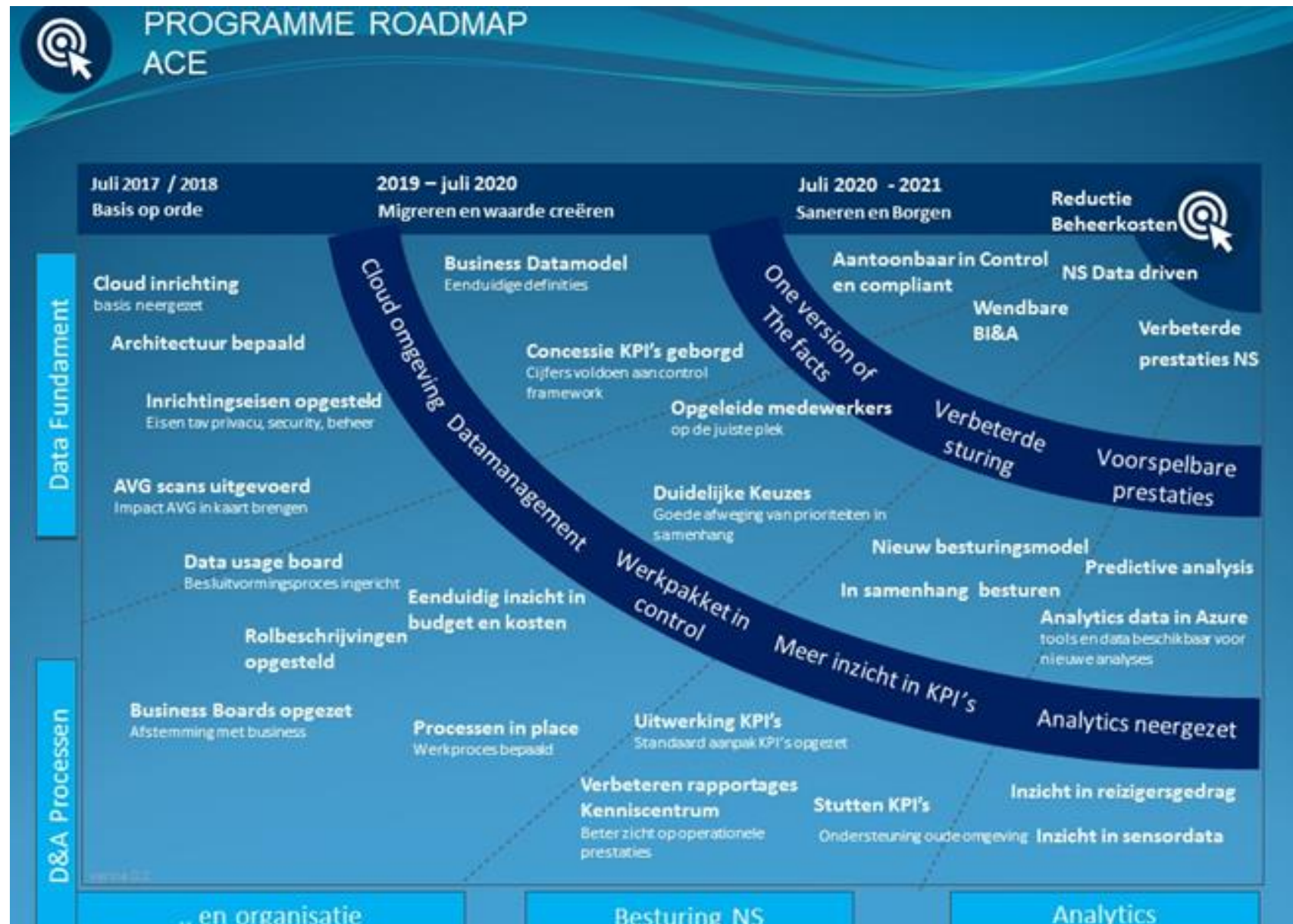
- ‘Adoptie Cloud Ecosysteem’
- Doel: Data-gedreven besturing eenvoudiger te maken voor héél NS
 - NS-breed proces- en bedrijfsgegevensmodel om samenhang te duiden tussen processen, data en (business) stakeholders.
 - afgestemde, uniforme definities NS-breed
 - alle data centraal beschikbaar op een nieuw NS-datafundament in de "cloud"
 - AVG compliant



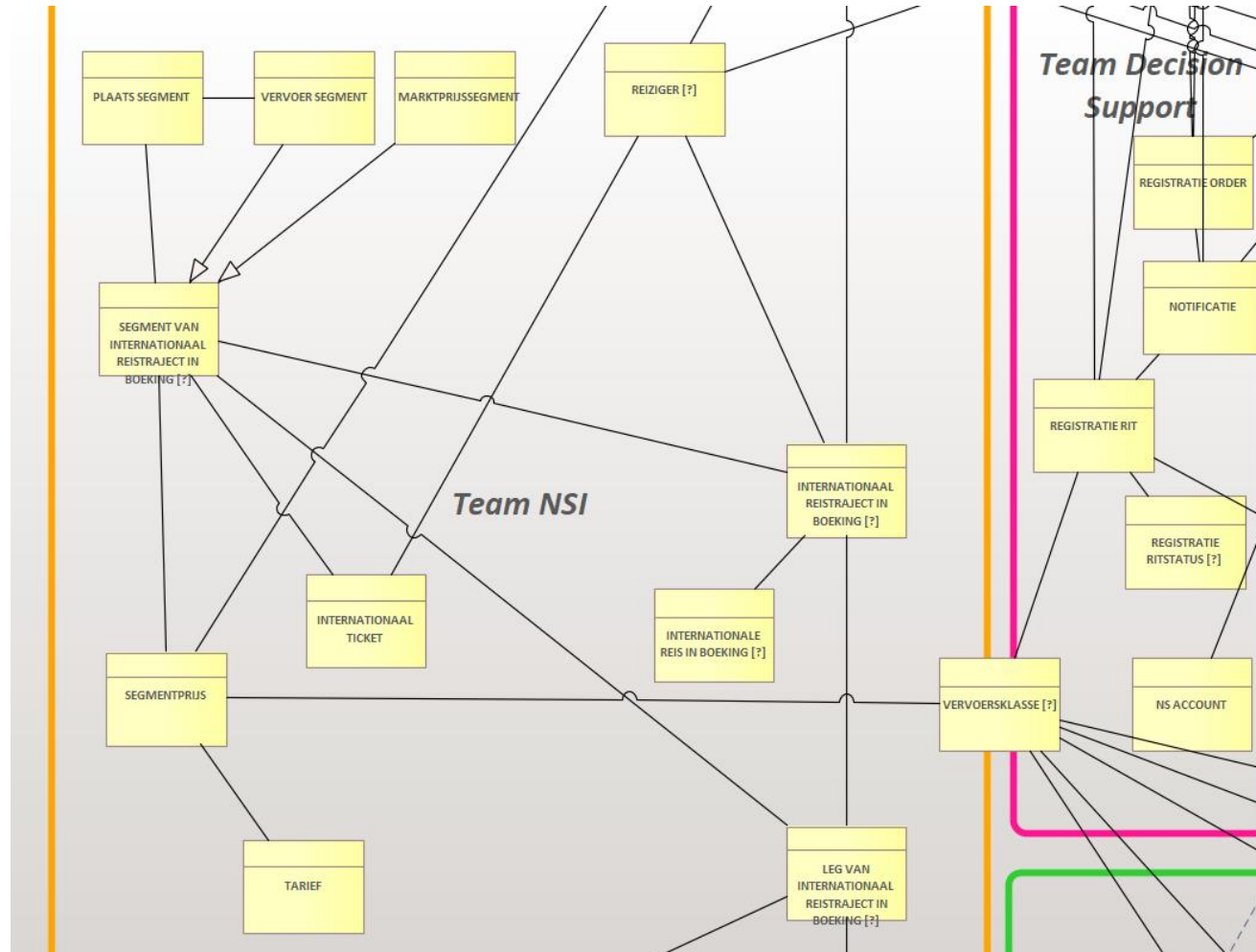


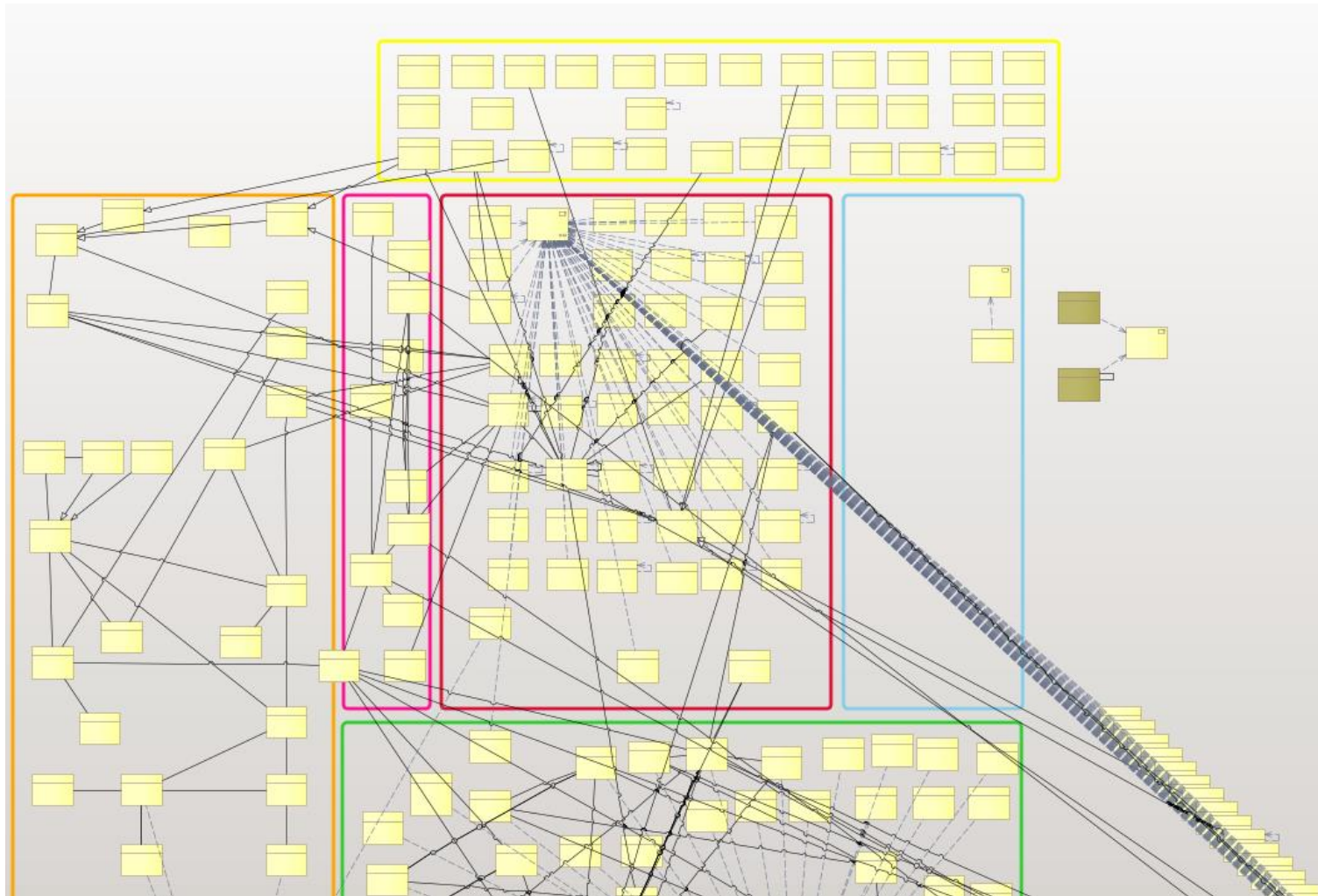
ACE Programma:

- Centrale afdeling 'Data & Analyse'
- NS breed Conceptueel, Logisch én Fysiek datamodel
- Één NS datafundament in Azure cloud
- Centrale data governance, incl. een Data Usage Board (DUB)

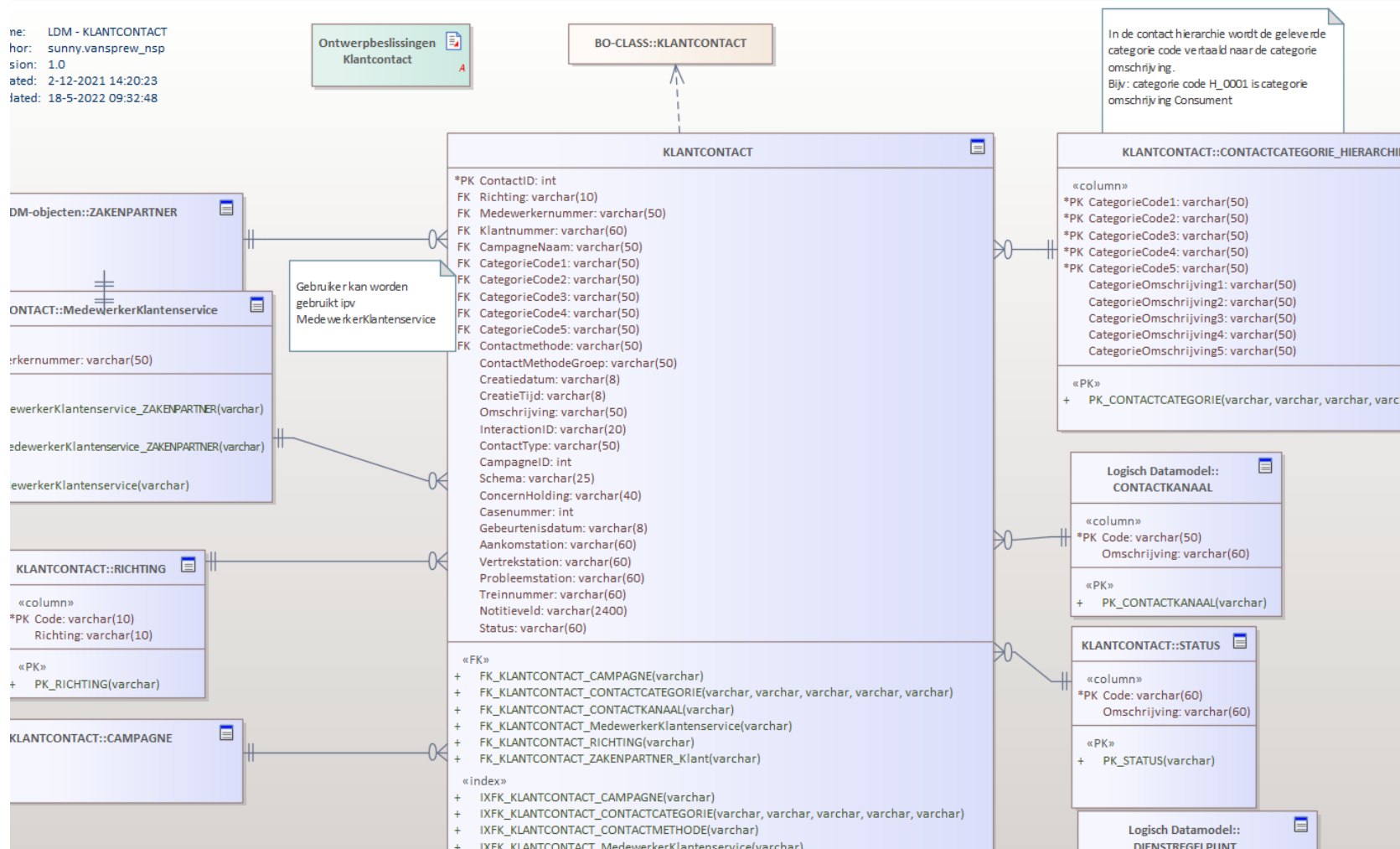


Voorbeeld van een onderdeel van NS BOM





Voorbeeld NS LDM

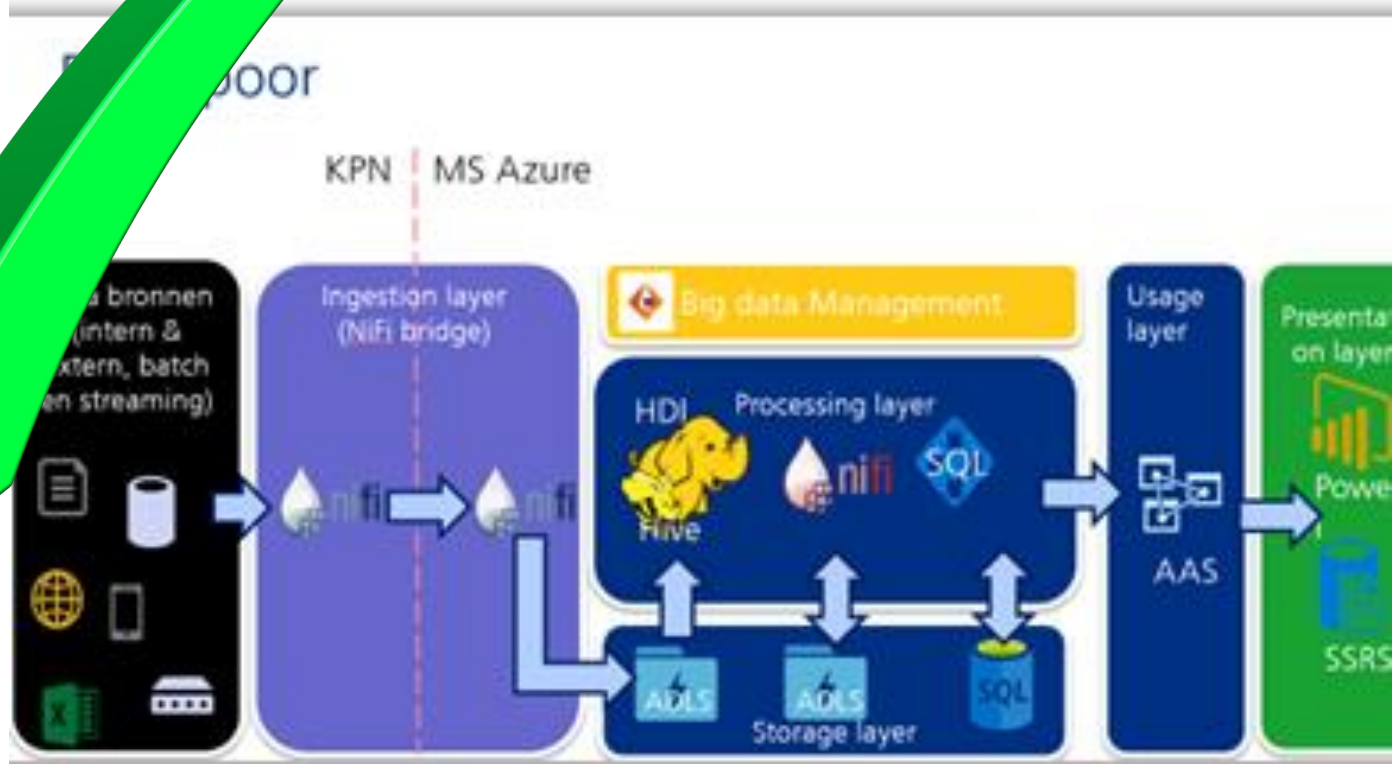
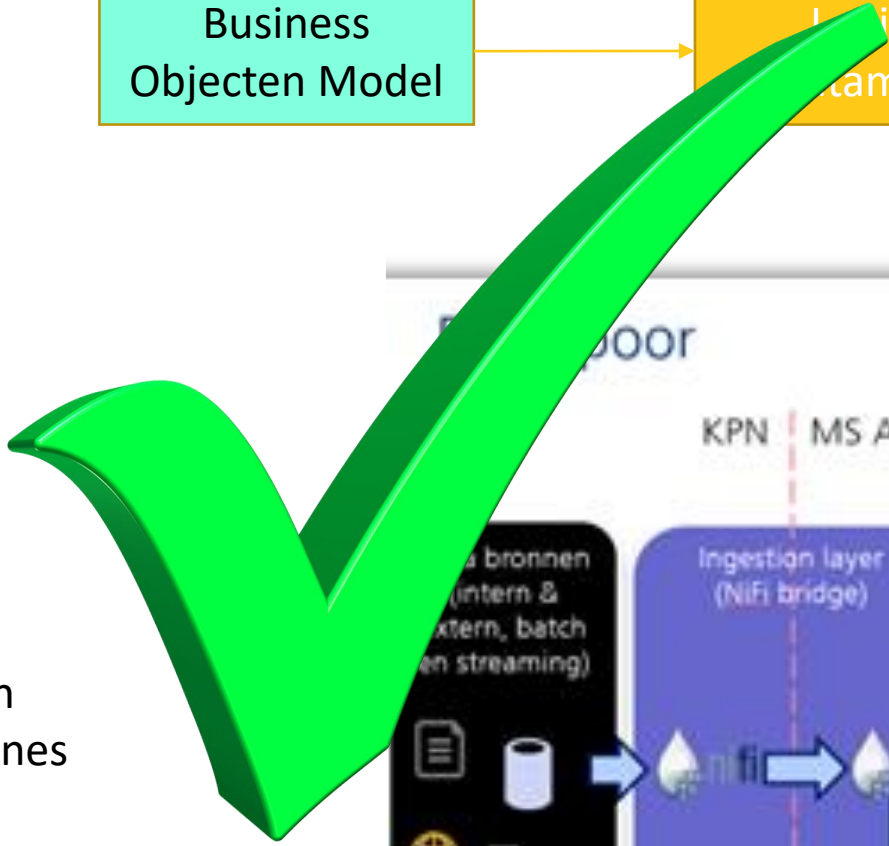


ACE Architectuur

Datamodelleren



1 Generieke solution voor alle Data pipelines



'De knel' : Theorie vs. Praktijk

- **Hoge kosten en lange doorlooptijd.**

- Datamodelleren heeft lange doorlooptijd
- Kennis van de data is nodig, niet up-front aanwezig
- Door applicatierationalisatie geen optimaal gebruik van cloud mogelijkheden

- **Niet alle data is NS breed:**

- Veel data is niet domein overstijgend. Het opnemen in een NS-breed datamodel met alle stappen is dan 'overkill'.

- **Schaalvergroting:** met honderden aan 1 datamodel werken... werkt niet. Weinig integratie en niet 1 signatuur.

- **Meer kennis van Cloud mogelijkheden:** datateams willen zelf binnen kaders geschikte tools en methodes kunnen kiezen

- **Niet alles hoeft 100% geborgd**

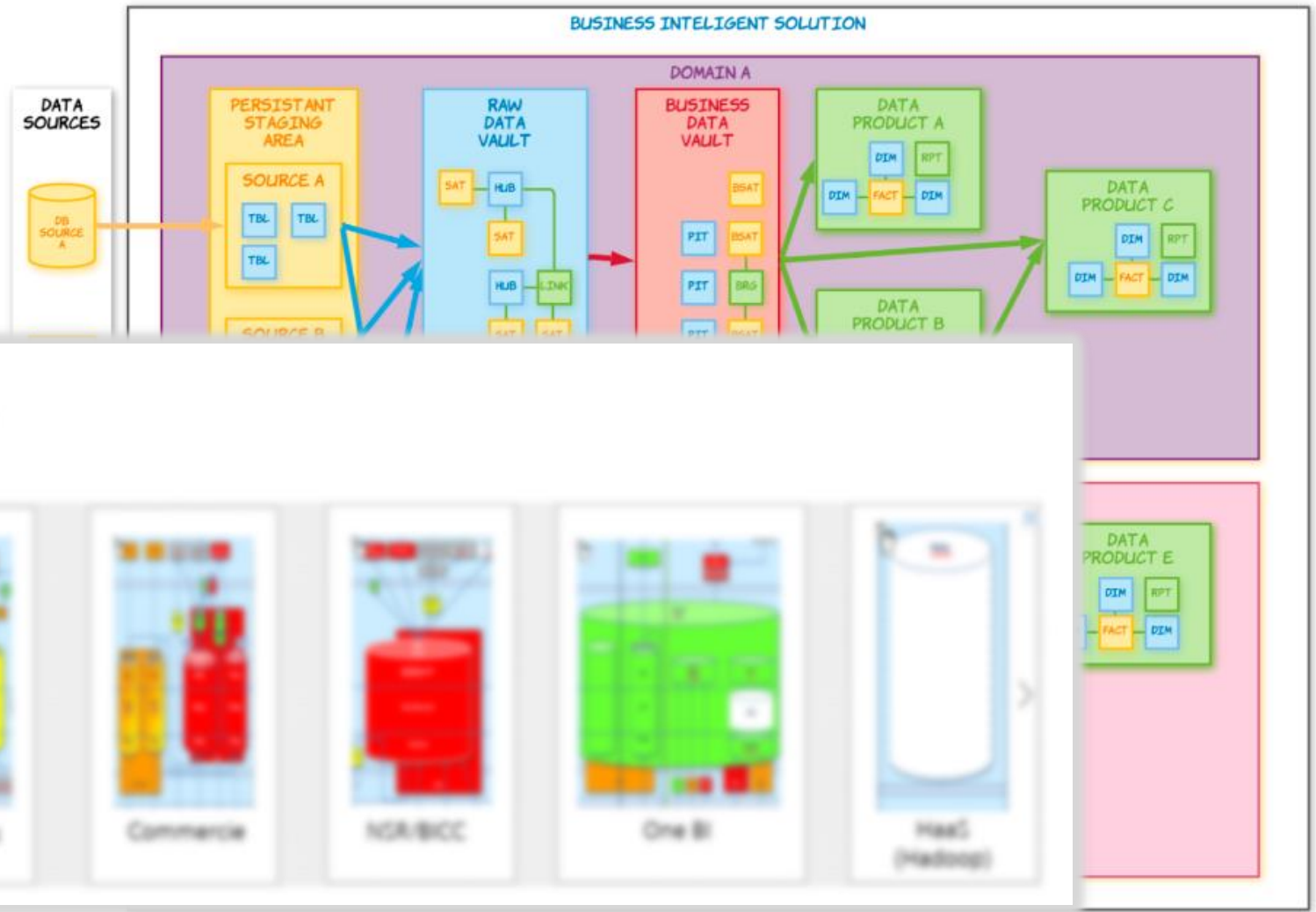


**Data
modelleren
is dood**

**GEEN BETUTTELING,
maar Waarde #Agile**

**DATA
MESH !!**
.....





Vóór 2016

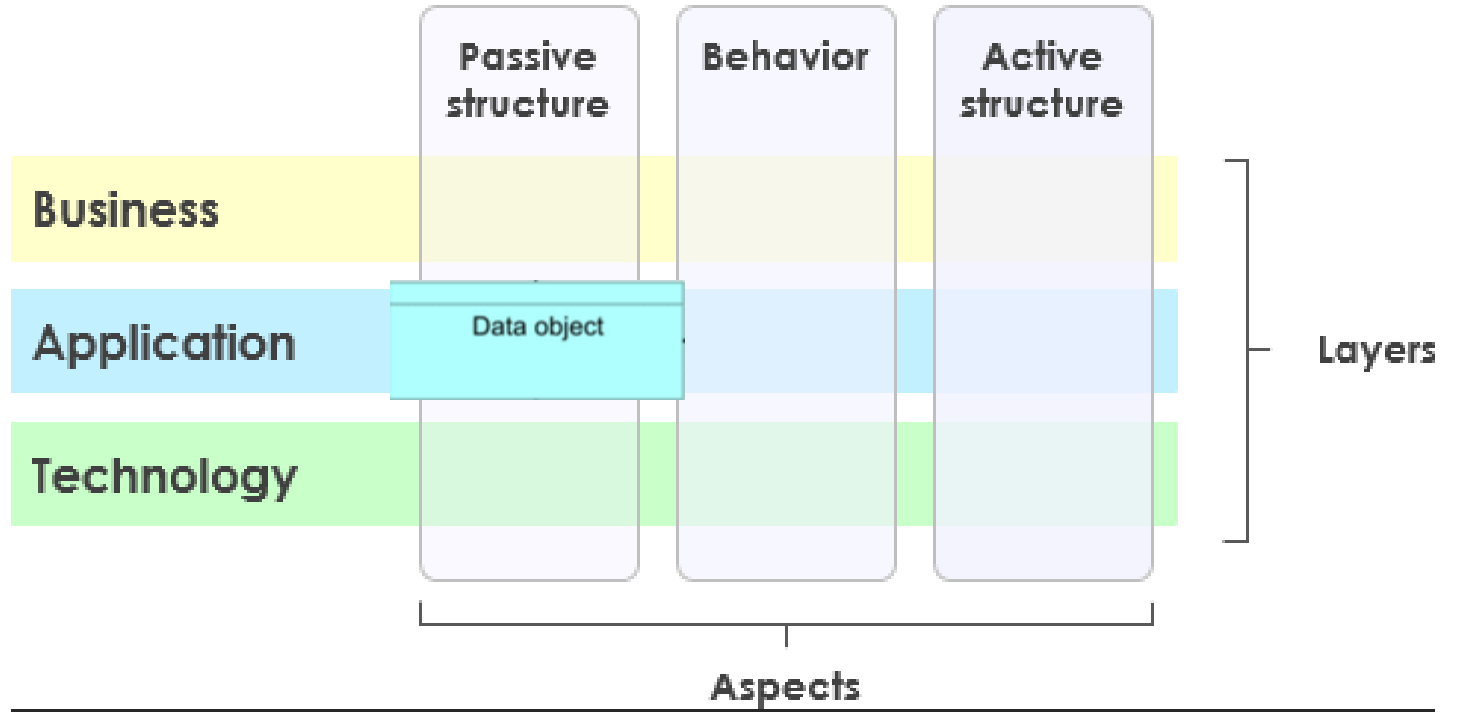
- NetFlix
- NS stations
- Commerce
- NCR/BICC
- One BI
- MapG (Maplog)

Opdracht

Creëer een data architectuur die voorziet in:

- Uniforme stuurinformatie (geen semantische verschillen) – **eis RvB**
- Meer wendbaarheid en innovatie in het NS datafundament
- Kaders voor de NS Resultaatgebieden m.b.t. NS datafundament



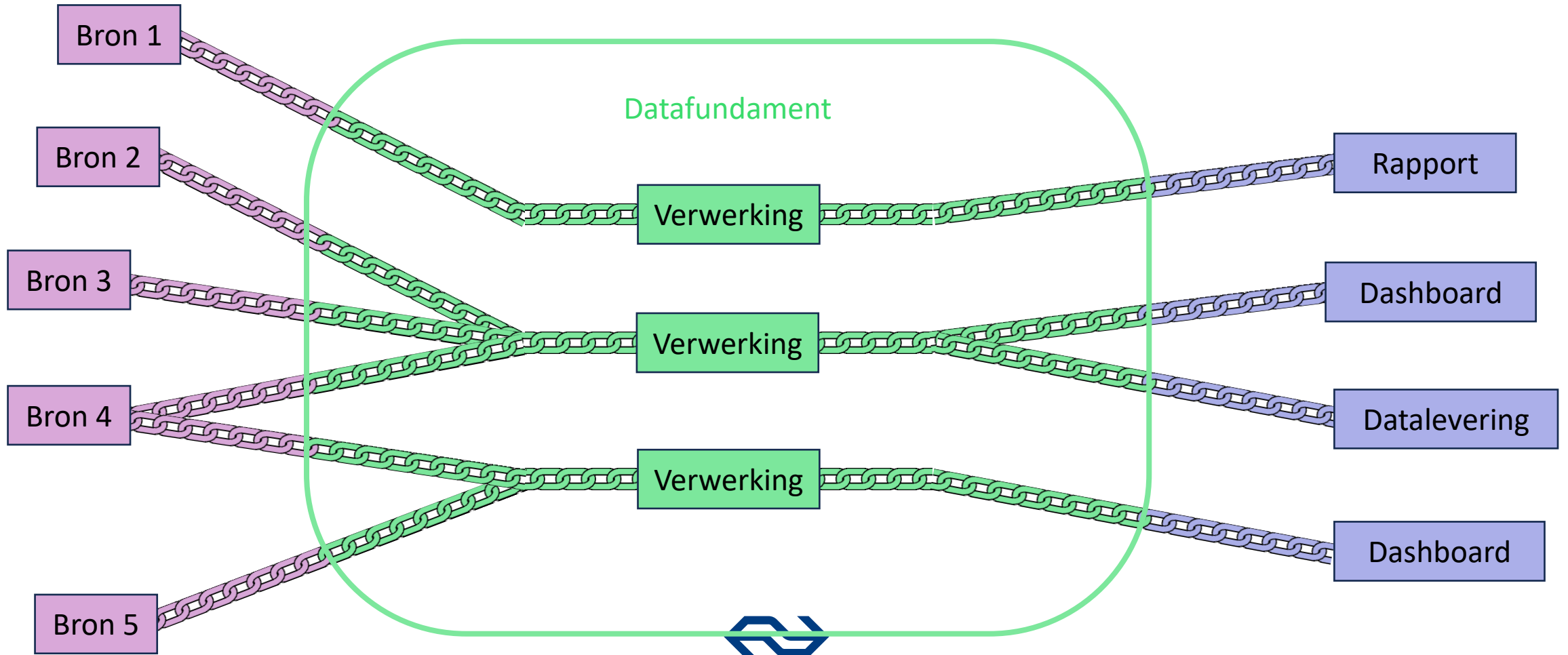




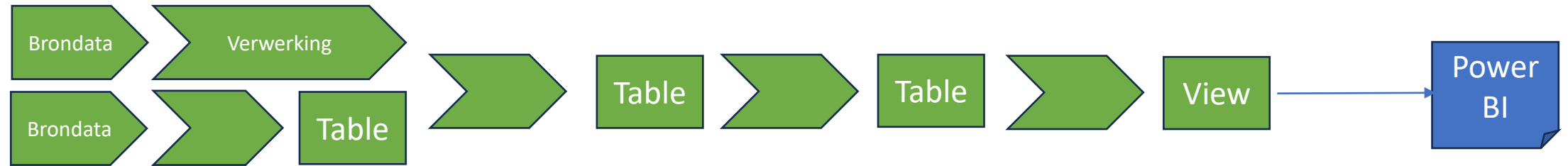
NS datafundament

- Datafundament: we ontsluiten bronnen en we leveren data- en informatieproducten
- Dataketen gaan door het NS Datafundament om van bron naar data- Infoproducten. Opwerkingen zijn dataschakels
- Een keten is het opdelen van complexiteit in behapbare delen.
 - Oftewel van brondata naar data-/informatieproduct opsplitsen in logische delen die behapbaar én herbruikbaar zijn
 - Behapbaar: een datateam kan hierop in control zijn en eigenschap nemen
 - Herbruikbaar: bevordert hergebruik, wat efficiency en consistentie/uniformiteit ten goede komt.

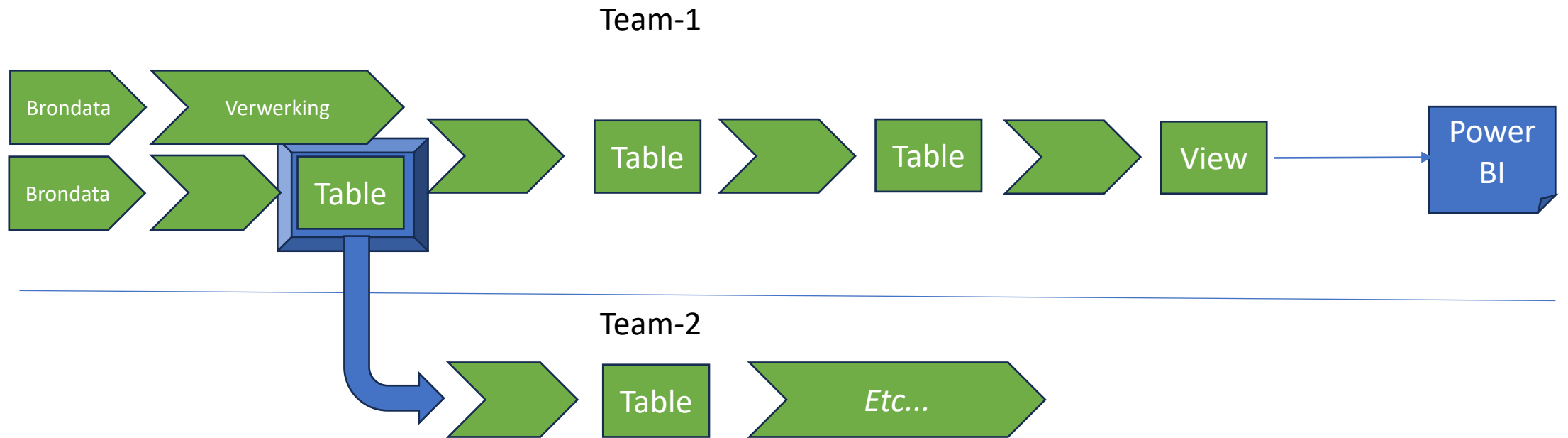
Dataketens



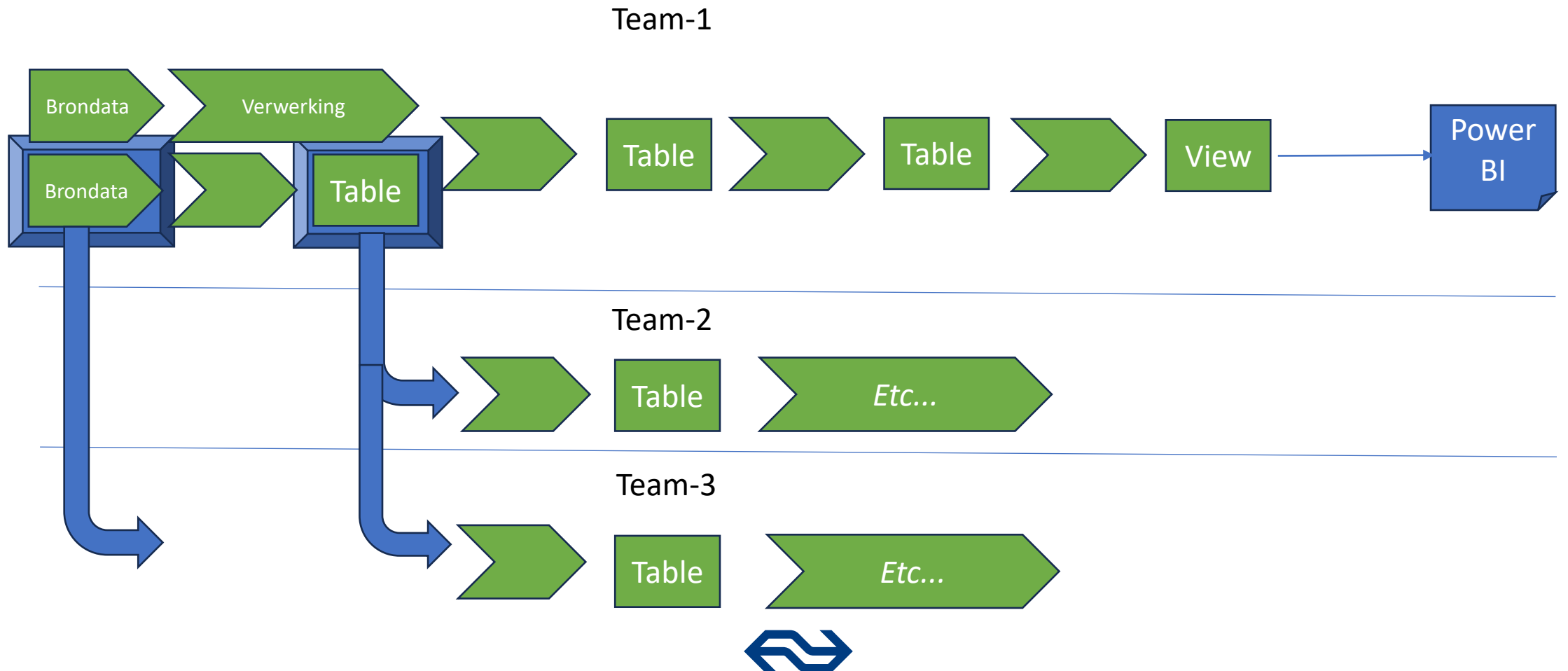
Data pipeline



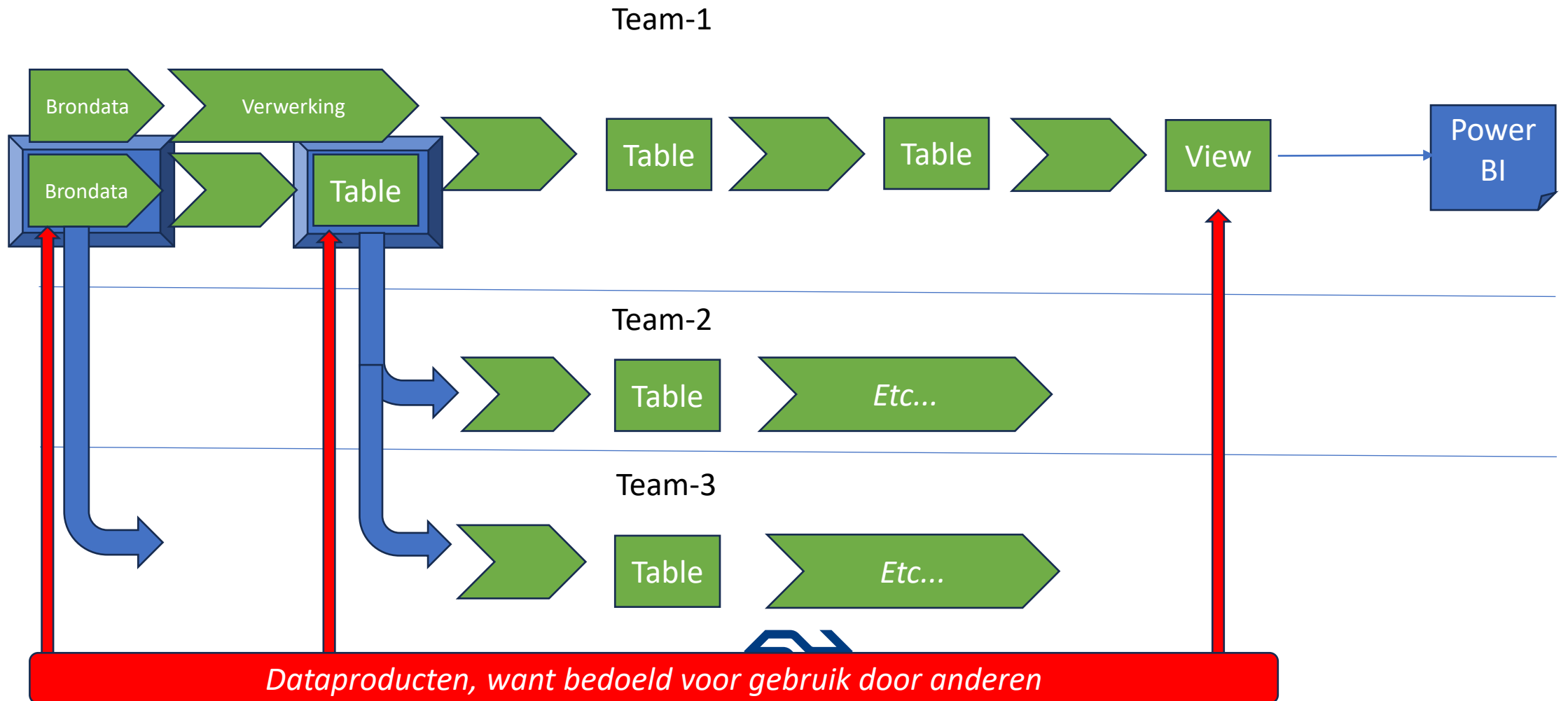
Data pipeline



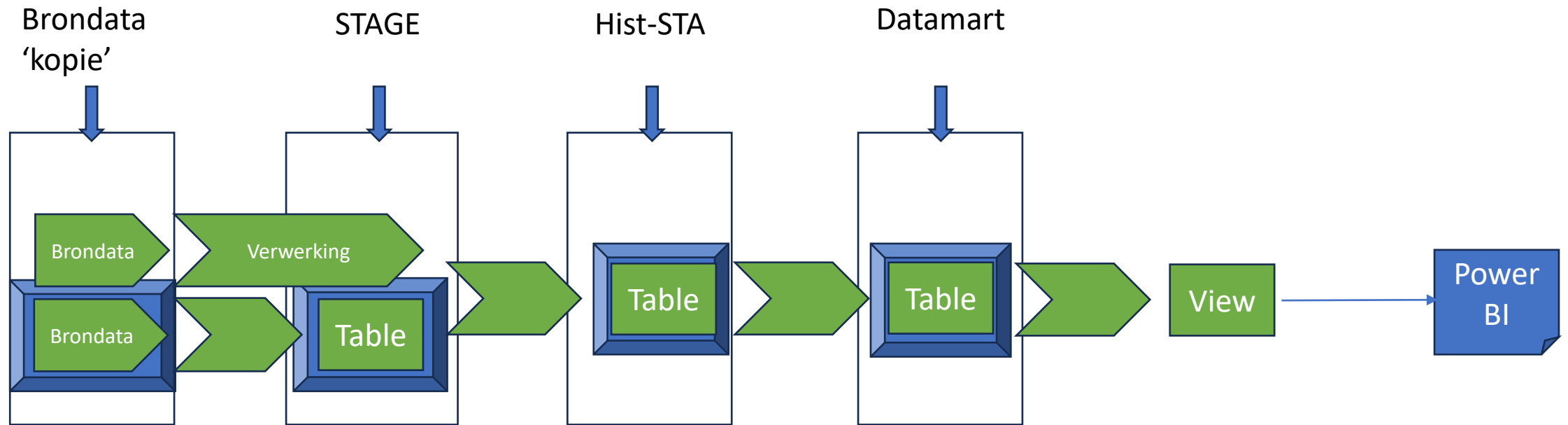
Data pipeline



Data pipeline



Data pipeline



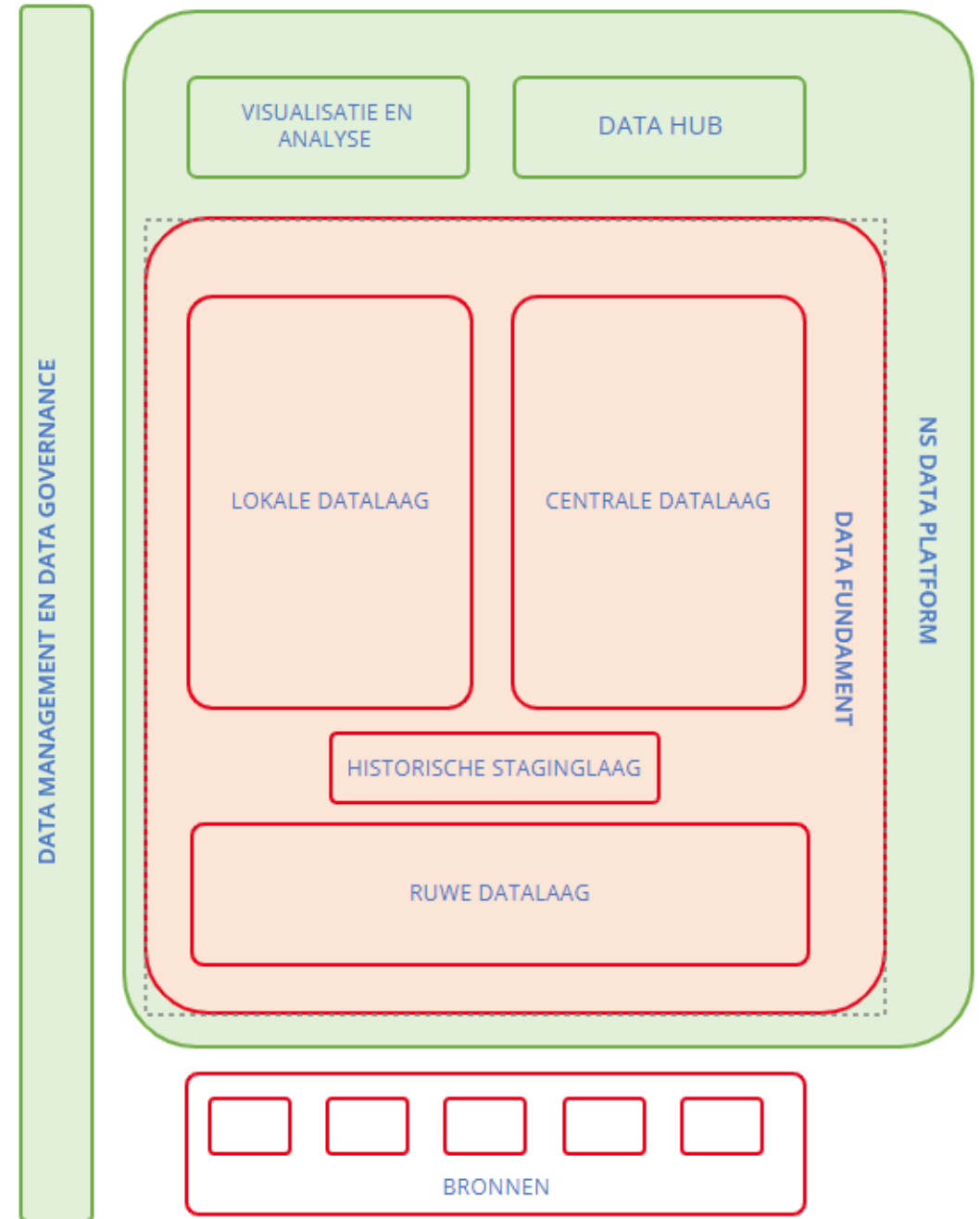
*Generieke data
(zoals in bron-systeem, evt. +historie)*

*Specifiek
(t.b.v. een specifiek Informatieproduct)*

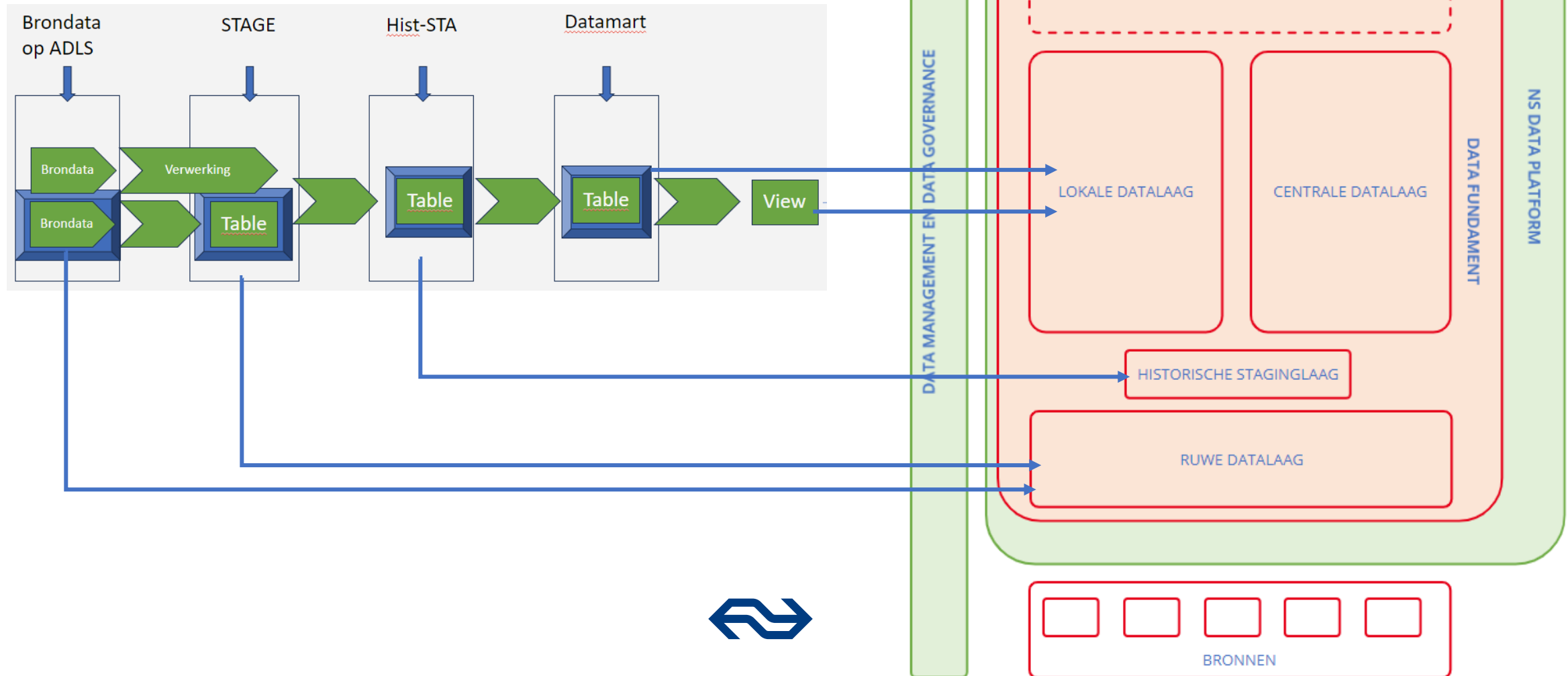


Insteek: dataproducten

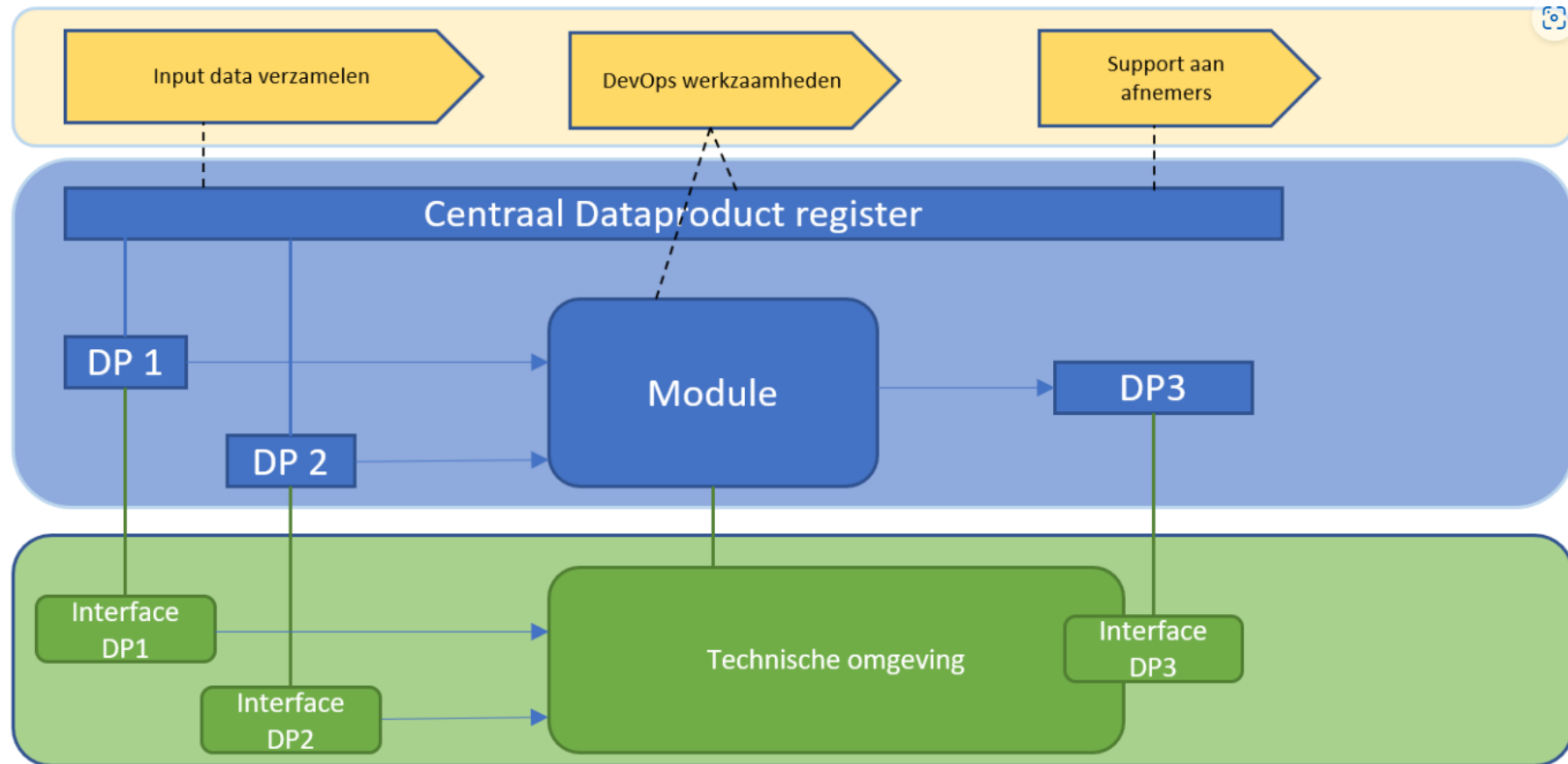
- Datalagen zijn dataproduct categorieën
- Categorie geeft duidelijkheid over mate dat data is opgewerkt + stelt kaders.
- Dataproduct: voor een afnemende dataschakel (applicatie).
- Informatieproduct: gebruikersinterface (apart product – *out-of-scope*)

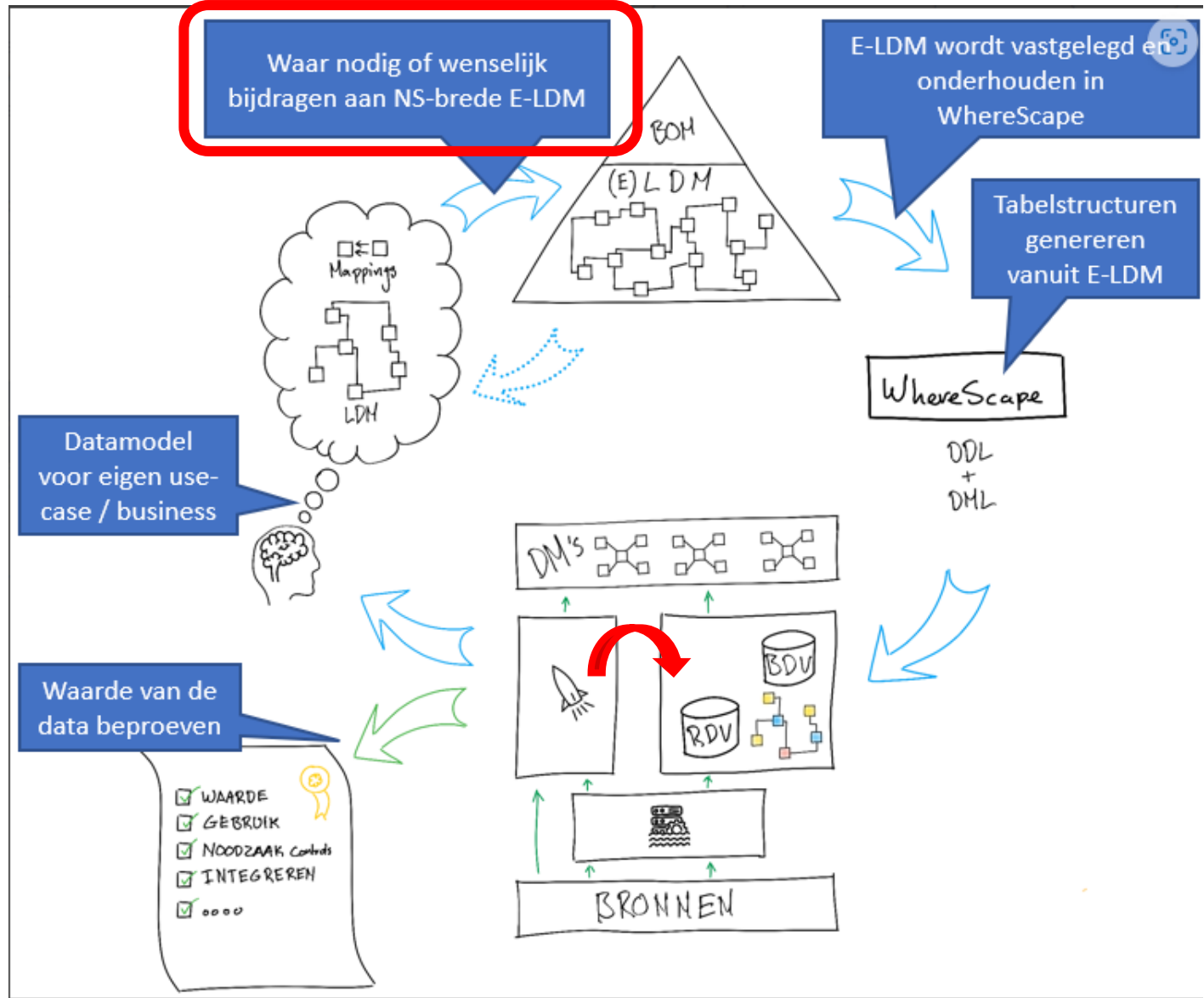


Datalagen zijn dataproduct-categorieën



Dataproduct(en)
consumeren en
leveren

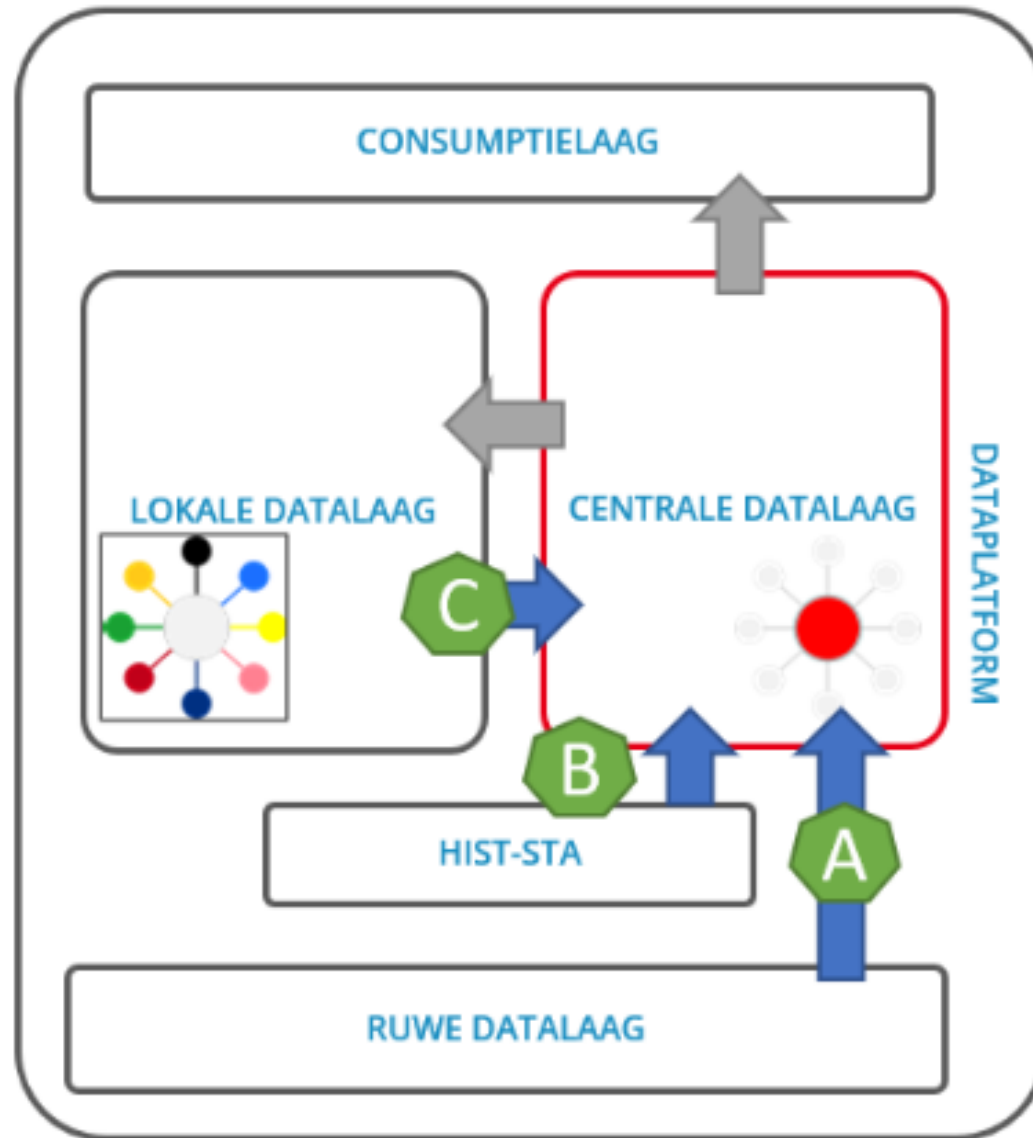




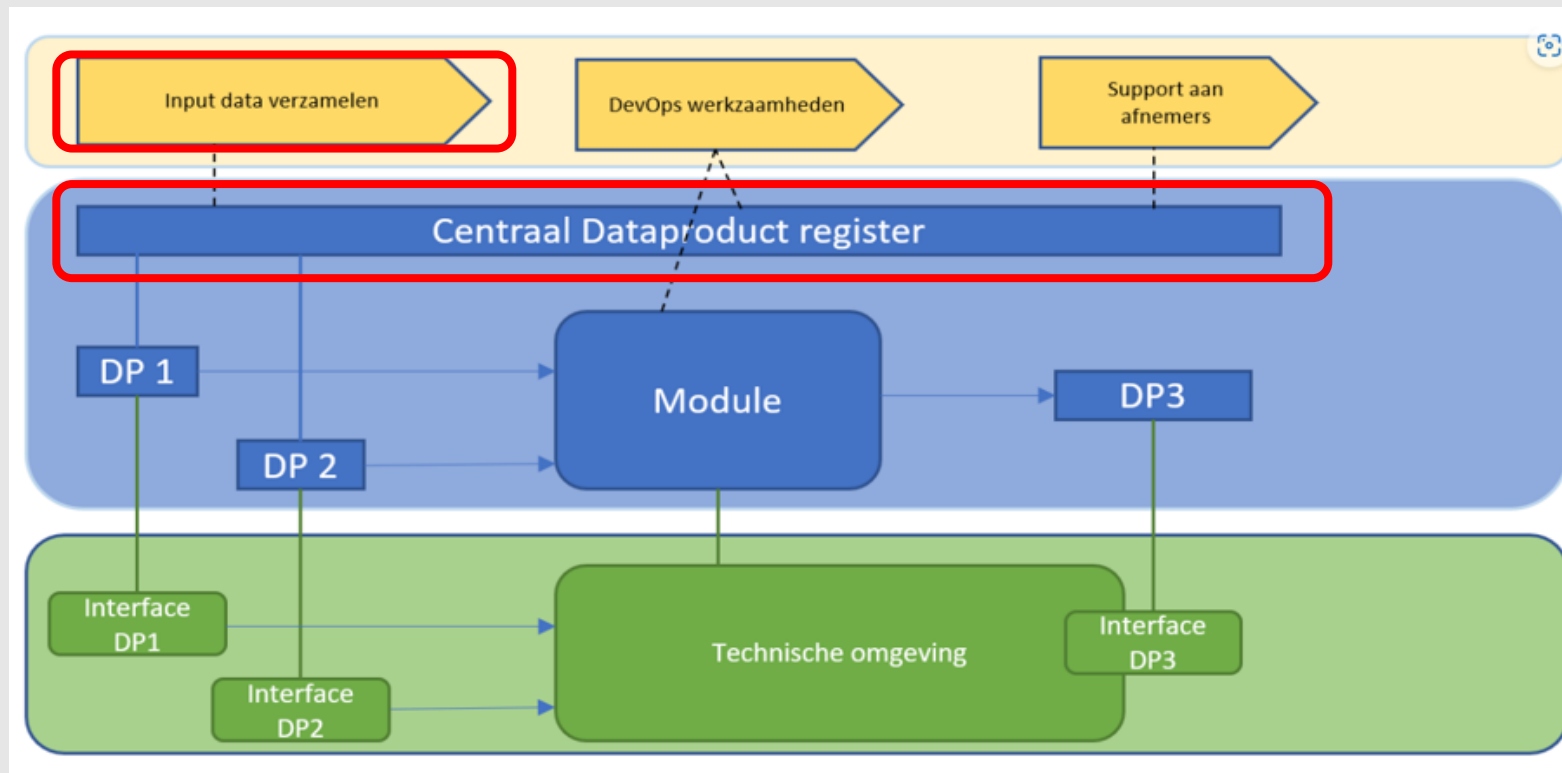
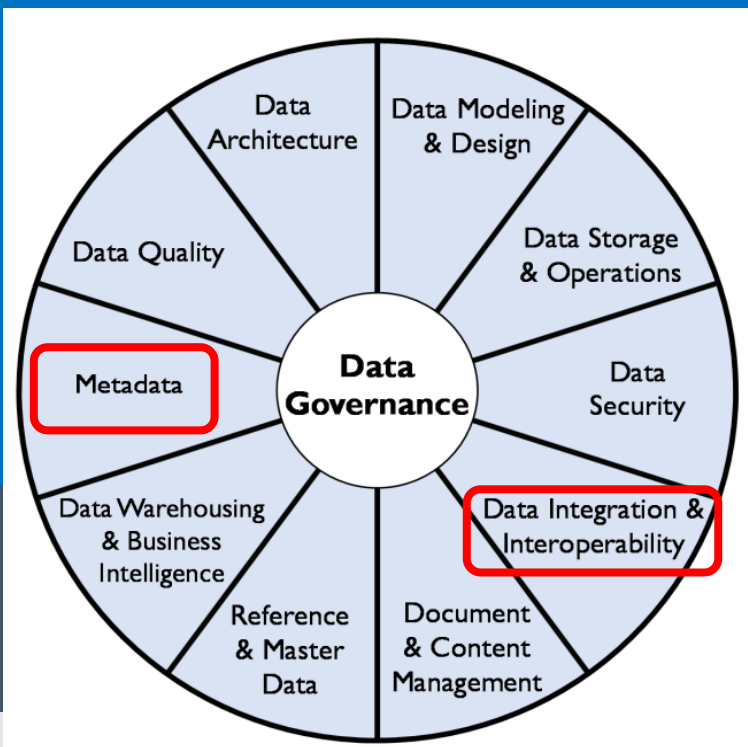
Centraal datamodel

- E-LDM: Enterprise Logisch Datamodel
 - Datateams hebben (business)datakennis en dragen deze bij via deel-LDM uit eigen bedieningsgebied. Dit mag in verschillende vormen.
 - Bijdrages van datateam worden getoetst en geïntegreerd in E-LDM door Kernteam-IA; zij vertegenwoordigen heel NS en bewaken kwaliteit E-LDM.
 - Fysieke datamodel (datastructuur) wordt gegenereerd vanuit E-LDM zodat fysiek en logisch altijd in lijn zijn. Dit is in IST een issue!





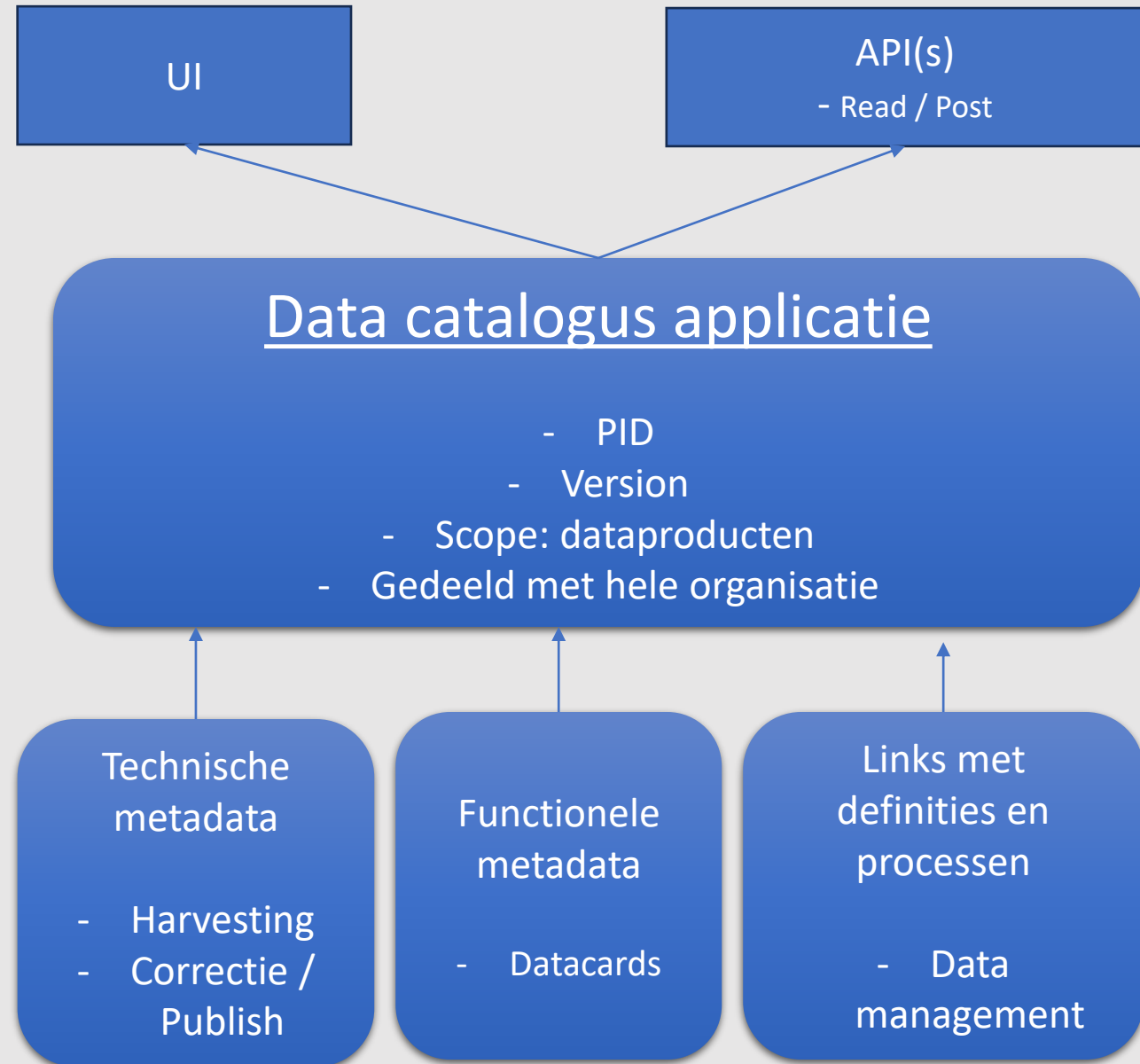
Data vindbaarheid



Rijke metadata in Datacatalogus

- Centrale data catalogus interface
- Shared responsibility om datacatalogus te vullen

➔
Release /
handmatig



Bedankt!



