

Informatiebeheer in een dataomgeving bij de gemeente Rotterdam.

Eerste stappen naar een integrale aanpak *by design*

Jeroen van Oss, Marco Klerks 5 oktober 2023



Gemeente
Rotterdam

Inhoud.

- a. Privacy en informatiebeheer by design bij Data gedreven Werken (Jeroen van Oss)
- b. Metadatamanagement in gegevenslandschap (Marco Klerks)

Wat is data gedreven werken?

1. Business intelligence (BI)

Dashboards voor bouwprogrammering, Instroom uitkeringen, Dienstverlening door 14010, Personeelsverloop, etc.

2. Analytics

Voorspellen op grond van data, scenario's ontwikkelen, trendanalyse, datamining

3. Data Science/advanced analytics

Algoritmen: bv. Onrechtmatigheid bij uitkeringen voor inzet toezicht en toetsing; detectie van naastplaatsingen van afval, ondersteunen van investeringsbeslissingen verkeerswegen door het inschatten van risico op ongevallen mbv een algoritme.



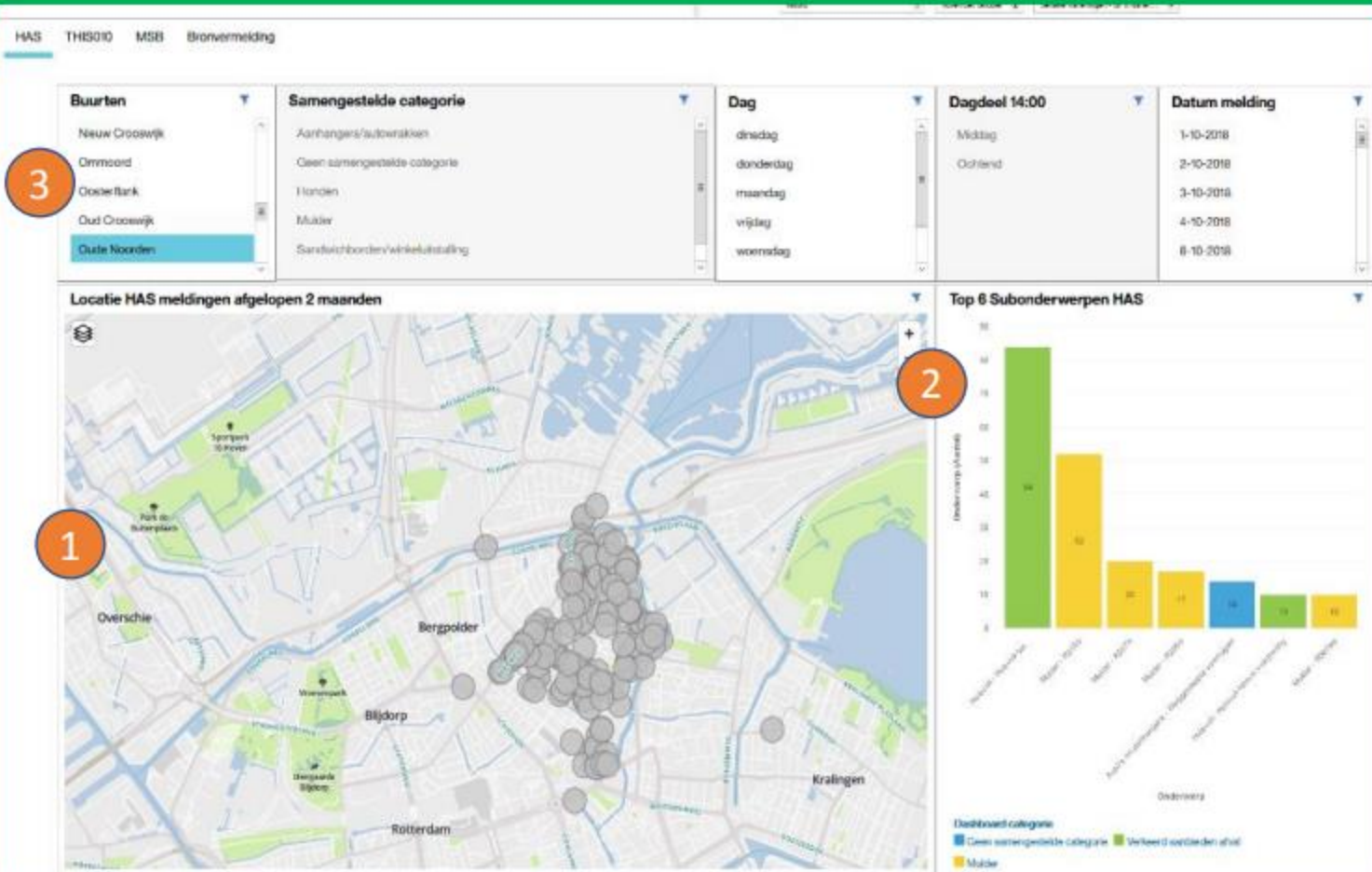
Bouwprogrammering

Stadsontwikkeling

Dit dashboard zorgt voor een dagelijkse overzicht van alle bouwprojecten per gebied, jaar, projectsoort en woningcategorie.

Het toont:

1. De status van de woningbouwproductie
2. De totale planvoorraad
3. Status van omgevingsvergunningen
4. Planologische status
5. Bouwscenario



Stadsbeheer

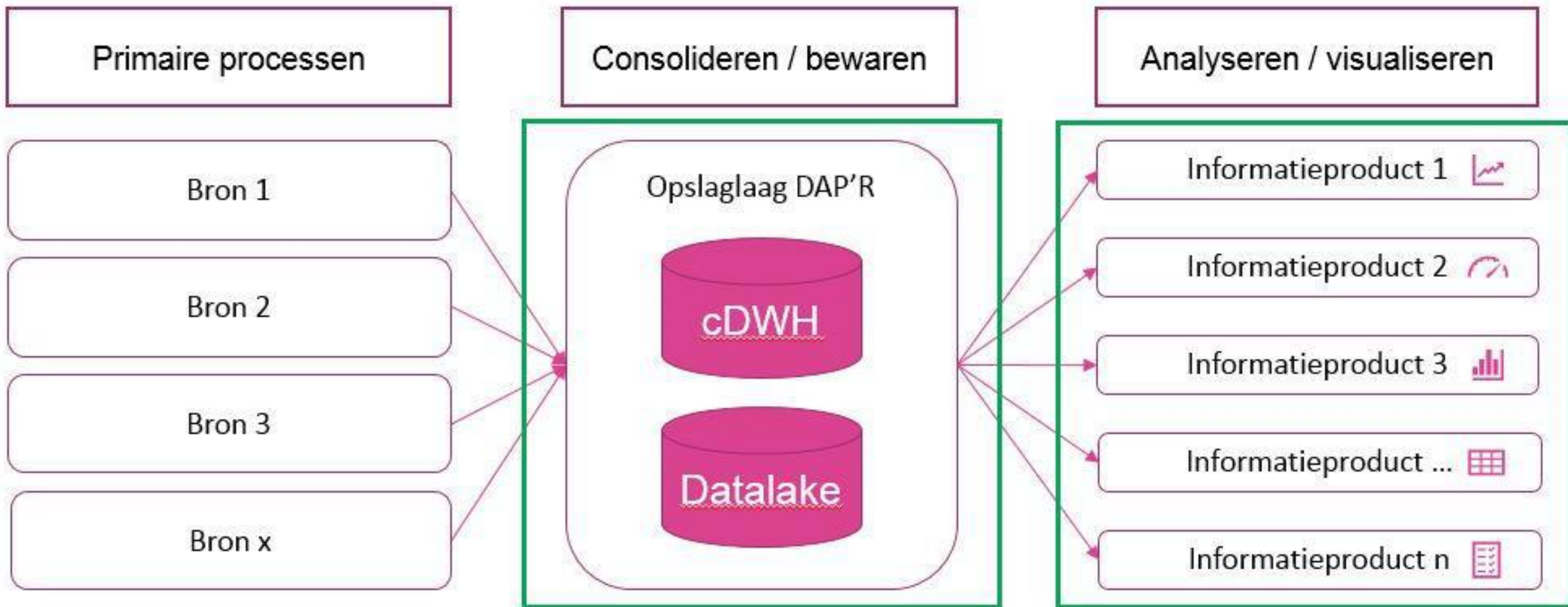
Dit dashboard is bedoeld voor de briefing Handhaving.

Het toont:

1. De waarnemingen van de handhavers op een kaart
2. De meest voorkomende categorieën
3. Je kan filteren op handige zaken zoals de buurt en de dag van de week

Informatie Gestuurd
Handhaven

Het DataAnalytischPlatform Rotterdam (DAP'R) onder de motorkap



Het project SPIDE'R

Security

Privacy

Informatiebeheer

Datamanagement

Ethiek

,

Rotterdam

Het project Data Gedreven Werken, actielijn 4, SPIDE'R

Aanleiding

Er zijn met het programma DGW afgelopen jaren veel instrumenten en gremia ontwikkeld: DACC's, DUB's, GLO's etc., etc. Maar het proces loopt nog stroef. Er ontbreekt één helder beleid op de SPIDE'R-gebieden, kennis, een strak proces en functionaliteit in de BI-software

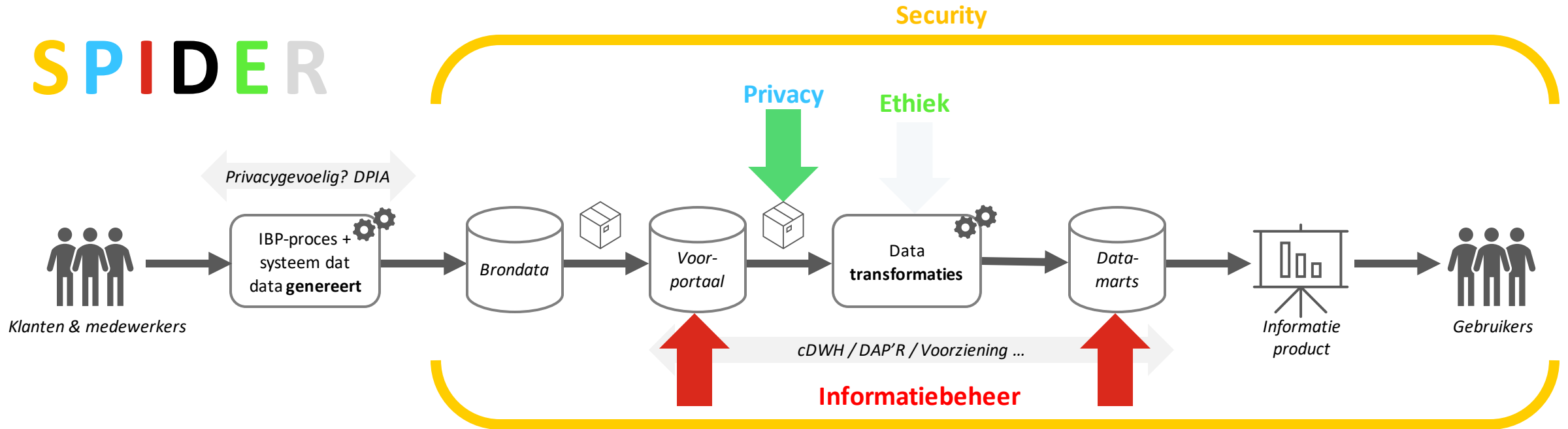
Doel

Inbedden van maatregelen m.b.t. Security, Privacy, Informatiebeheer, Datamanagement, Ethiek in de totstandkoming van producten van Data gedreven werken (DGW).

Beoogde resultaten:

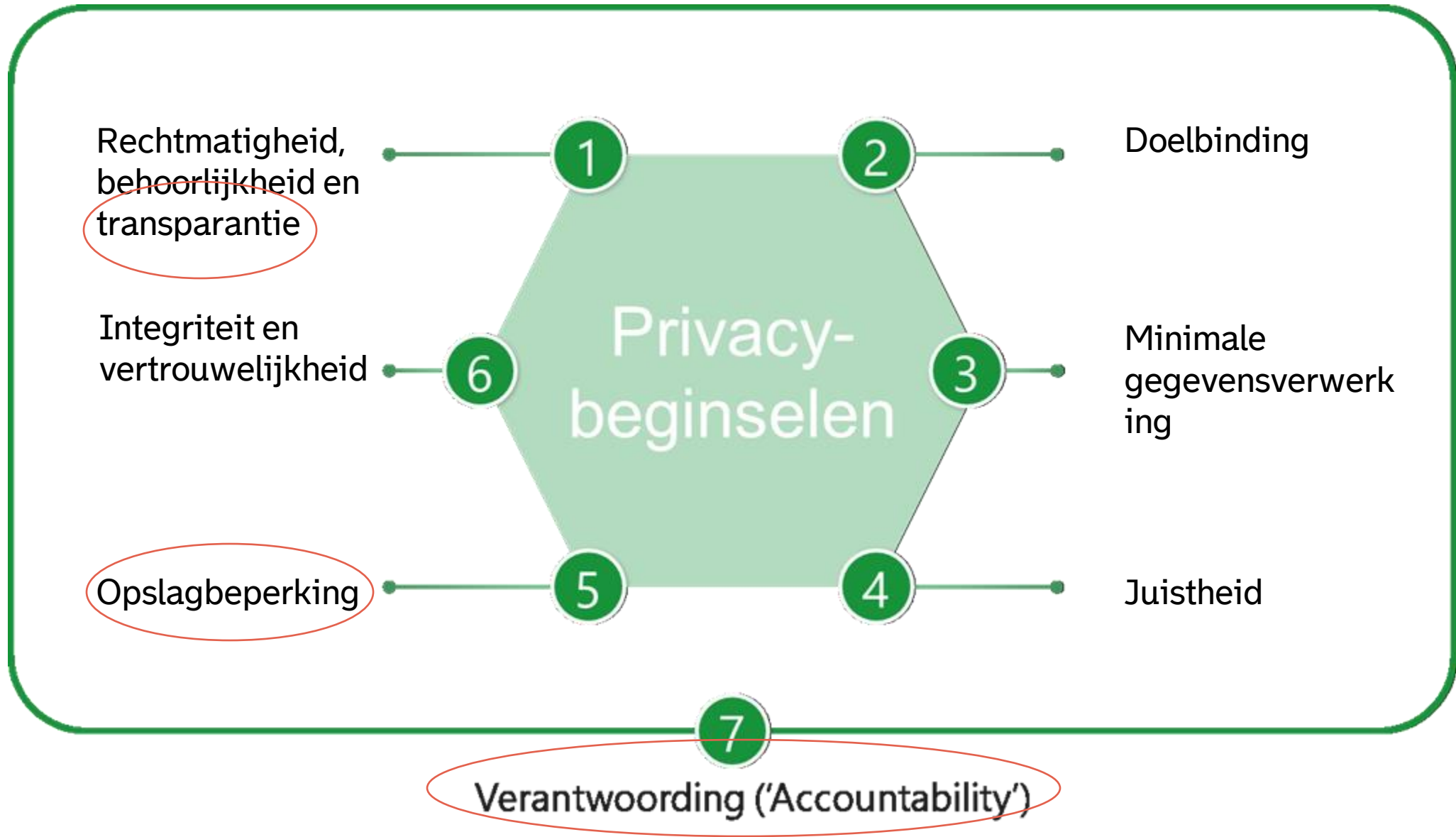
- Tactisch kader: *Privacyframework DGW*, waaronder o.a. hoofdstukken Informatiebeheer, Security, Datamanagement
- Het proces van idee naar informatieproduct (en daarna), Data Usage Board (DUB)
- Eisen voor functionaliteit voor het *DAP'R platform*, vanuit de SPIDE'R-disciplines.
- Inhaalslag vaststelling DPIA's van de applicaties die voor DGW gebruikt worden.
- Training van betrokkenen (te beginnen met ontwikkelaars en SPIDE'R disciplines)

Procesplaat: samenhang tussen de SPIDER-elementen



	Verkrijging	Gebruik	Instrument
Security	Is het proces waterdicht of kan er data weglekken?	Idem	'Stramienen' (standaarden), autorisatiestructuur
Privacy	Is er sprake van (bijzondere) persoonsgegevens?	<ul style="list-style-type: none"> • Afgeleid van verkrijging; focus op nieuwe combinaties • Is er sprake van <i>gegevensminimalisatie</i>? • Is er sprake van <i>doelbinding</i>? 	Metadata: dataclassificatie
Informatie-beheer	Wat is de levenscyclus / bewaartermijn van de data?	Afgeleid van verkrijging; focus op nieuwe combinaties	Metadata: bewaartermijnen
Datamanagement	Welke afspraken worden gemaakt over levering en datakwaliteit?	Idem	Metadata, GLO
Ethiek		Zijn de bewerkingen verantwoord of vindt er bijv. profilering plaats?	Functioneel & Technisch Ontwerp / algoritmeregister

Raakvlakken privacybescherming en informatiebeheer



Het Privacyframework

H 1 Privacy beginselen

H 2 Privacygovernance

- rollen, verantwoordelijkheden
- Het DUB process)
- Gegevensleveringsovereenkomsten (GLO's)

H 3 Levenscyclusbeheer (Informatiebeheerkaders)

H 4 Security

Inzet vanuit Informatiebeheer

- 1. Bewaren en vernietigen: Wat valt onder het informatiebeheer? hoe lang bewaar je data van een informatieproduct? Hoe pas je bewaartermijnen toe op alle verschijningsvormen? Masterdata, testdata, rapportages, lijsten.**
2. Goed documenteren t.b.v. transparantie, hoe zijn beslissingen tot stand gekomen, bijvoorbeeld bij gebruik algoritmes? Documenteren van hoe het BI Informatieproduct werkt.
3. Beheerverantwoordelijkheid voor data bij leveringen intern (GLO) en extern (DTA).
Archiefovereenkomst.
4. Metadatamanagement: bijdrage van Marco hierna

Uitgangspunten toepassing bewaartermijnen

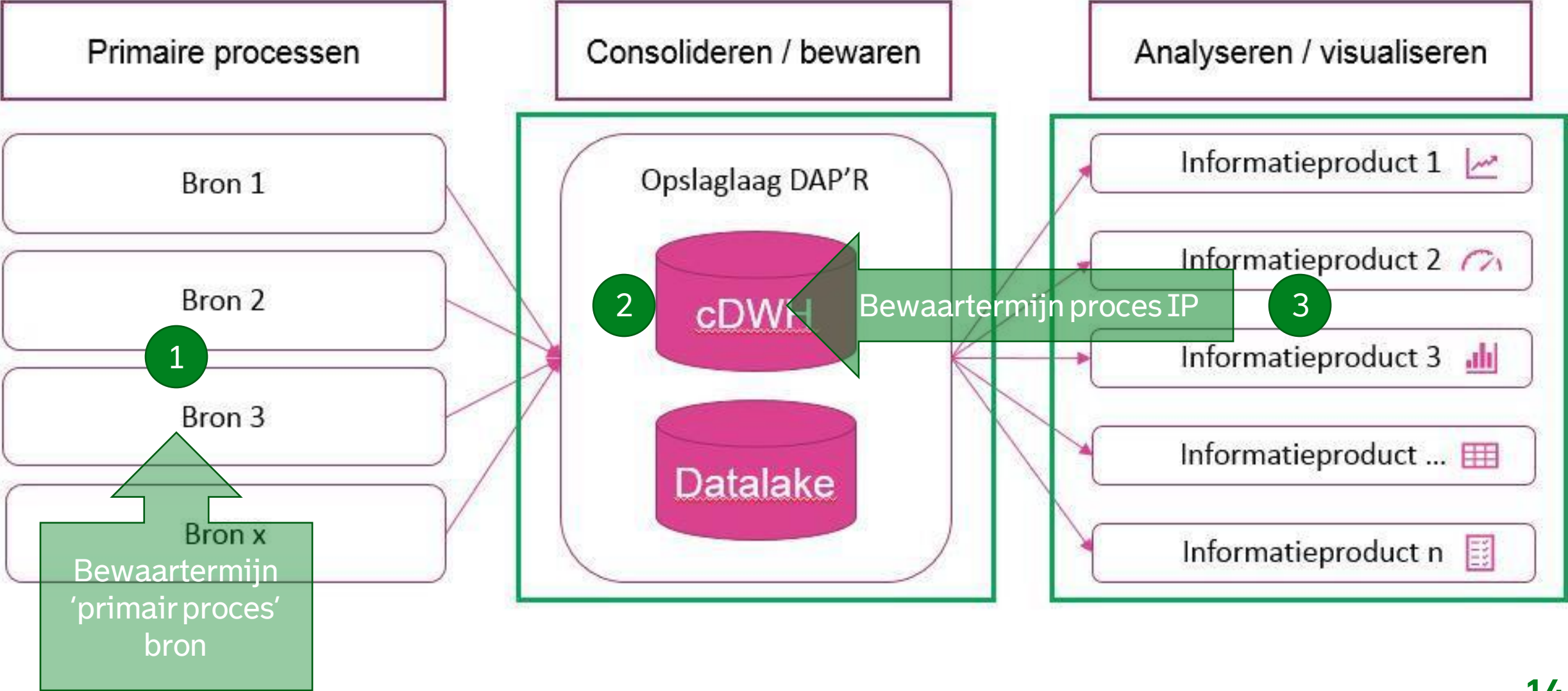
- Informatie ('archief') ontstaat vanuit een bedrijfsproces
- Doel is 'kunnen reconstrueren hoe het proces verlopen is' (op grond waarvan zijn besluiten genomen?)
- Informatie uit een of meerdere bron(en), die hergebruikt word(en)t voor DGW, is neerslag van een ander proces dan het proces waar het oorspronkelijk in gevormd is (de bron).
- Die hergebruikte informatie, heeft daarmee een ander doel, ondersteunt een ander proces en heeft daarmee een andere bewaartermijn.

Voorbeelden:

IBP-Proces	Bewaartermijn	Grondslag in selectielijst
Verstrekken uitkeringen	10 jaar na einde uitkering	8.1 Voorzieningen verstrekken, generiek resultaat
Metten of het uitkeringsproces effectief verloopt	10 jaar na einde informatieproduct	4.1 Evaluatie uitvoeren, generiek resultaat

[handreiking-gebruik-selectielijst_20200506.pdf \(vng.nl\)](#) par. 5.6

Hoe werkt dit door in het datalandschap?



Bezwaren vanuit Privacy

- Ruwe data kunnen persoonsgegevens bevatten.
- Deze worden dan bijvoorbeeld 10 jaar of blijvend bewaard, conform de bewaartermijn die bij het proces van het informatieproduct hoort.
- Dit wringt met het AVG beginsel van opslagbeperking en lijkt weinig meerwaarde voor transparantie en erfgoed te hebben.

Afweging bewaartermijn ruwe data tussen AVG en Archiefwet: transparantie versus privacy?

1. Pseudonimiseren

- (Beperkte) privacybescherming met behoud van transparantie

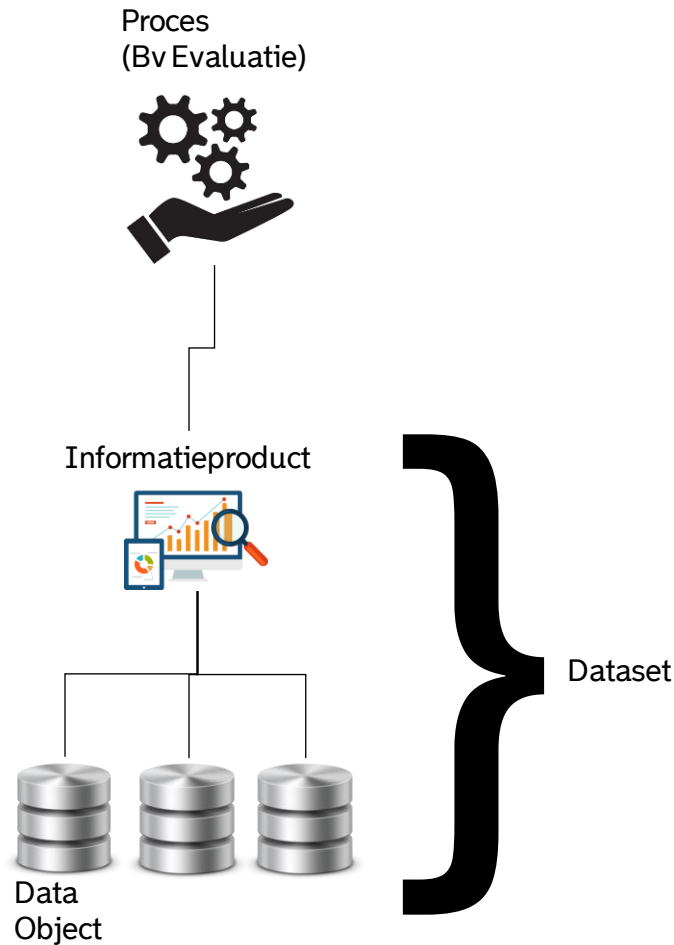
2. Anonimiseren

- Mogelijk verlies aan betekenis, kan transparantie in de weg staan

3. Vernietigen volgens alternatieve, niet-proces gebonden bewaartermijn: 19.1 'Gegevens verwerken' 5 jaar.

- Nog steeds erg lang....
- Aanpassing VNG Selectielijst is gewenst!
- Documenteren hoe het Informatieproduct tot stand is gekomen t.b.v. transparantie!

Bepalen en effectueren bewaartermijnen op dataobjecten- (conceptueel)



Een **informatieproduct** ondersteunt een (data analyse) **proces**.

Een informatieproduct gebruikt één of meerdere **dataobjecten**.

Het informatieproduct gezamenlijk met de dataobjecten noemen we een **dataset**.

De dataset in zijn geheel wordt toegepast ten behoeve van een proces.

Per proces is een bewaartermijn van toepassing (de selectie categorie).
Voorbeelden zijn

- 4.1 Evaluatie (**10 jaar**)
- 4.1.1 Evaluatie met externe werking (**blijvend**)
- 12.1 Toezicht en handhaving (**5 jaar**)
- Etc.

Dataobjecten kunnen in meerdere informatieproducten worden gebruikt. Bij het informatieproduct leggen we de datum vast wanneer deze in gebruik wordt genomen en ook wanneer deze buiten gebruik wordt gesteld. Dit is de startdatum voor het toepassen van de bewaartermijn voor het informatieproduct en de onderliggende dataobjecten.

Bewaartermijnen -> dataobjecten



IP 1 - Proces X
 Bewaartermijn 10 jaar
 Datum buiten gebruik 01-10-2013
 IP+Objecten verwijderen op 01-10-2023



IP 2 - Proces Y
 Bewaartermijn 2 jaar
 Datum buiten gebruik 01-01-2021
 IP+Objecten verwijderen op 01-01-2023



IP 3 - Proces Y
 Trendanalyse afgelopen 15 jaar
 Bewaartermijn 2 jaar
 Datum buiten gebruik 01-11-2020
 IP+Objecten verwijderen op 01-11-2022

Huidige voorstel is alle rijen verwijderen als object niet meer in actief IP wordt gebruikt.
 Op tabel niveau datum 01-10-23 (hoogste datum).
 Je kunt niet zien welk record precies in welk IP wordt gebruikt.

Persoon	Naam	Start datum object	Eind datum object	Persoon Actief	Start datum record	Eind datum record	Bewaar Categorie vanuit IP	Datum buiten gebruik ivm IP	Datum verwijderen vanuit IP	Bewaar Categorie vanuit Object	Datum verwijderen vanuit Object	Record Actief	Verwijderen historie (ouder dan 20 jaar)	Opmerking
150250	Mevr. X-D****nie	01-01-2000		Ja	10-08-2000		10 jaar	01-10-2013	01-10-2023	12 jaar	Geen einddatum bekend nog	Ja	NvT	Wordt verwijderd wanneer laatste IP verwijderd wordt
150250	Mevr. D****nie	01-01-2000		Ja	01-01-2000	09-08-2000	10 jaar	01-10-2013	01-10-2023	12 jaar	Geen einddatum bekend nog	Historisch	08-08-2020	Niet gebruikt in P1 en P2, maar kun je niet zien (ligt aan selectie). Wordt eerder verwijderd op basis van historie.
110000	...	01-01-2000	31-12-2010	Nee	01-12-2010		10 jaar	01-10-2013	01-10-2023	12 jaar	31-12-2022	Ja	NvT	Niet gebruikt in P1 en P2, maar kun je niet zien (ligt aan selectie). Wordt verwijderd voor IP3 terwijl mogelijk nog in gebruik.
110000	...	01-01-2000	30-11-2010	Nee	01-01-2000	30-11-2010	10 jaar	01-10-2013	01-10-2023	12 jaar	30-11-2022	Historisch	30-11-2030	Niet gebruikt in P1 en P2, maar kun je niet zien (ligt aan selectie). Wordt verwijderd voor IP3 terwijl mogelijk nog in gebruik.

Hoe bepaal je bewaartermijn op objectniveau? Dus los van een proces waarin het gebruikt wordt?
 NB: De 20 jaar voor historie is gebaseerd op trend analyses zijn altijd < 20 jaar

Implementeren bewaartermijnen in het DAP'R

- Onderzoek naar mogelijkheden van Purview op het Azure platform
- Wat kunnen we procesmatig borgen, wat technisch?
- De structurele oplossing vraagt om volwassen metadatamanagement
- Nog een flinke puzzel....

