

Ketenarchitectuur

Hoofdpijnendocument

veranderopgave onderwijsgegevensplatform

Datum 20 november 2021

Opdrachtgever

Status

Versie

Colofon

Programma Onderwijsgegevensplatform

Architectuurboard Architecten van onderstaande organisaties

OCW (vz)		
PO		
VO		
MBO		
HO		
DUO		
SLO		
CITO		
Ivho		
CVTE		
CBS		

Op basis van

NORA
ROSA
HORA
FORA
MORA

Versiebeheer

Versie	Datum	Beschrijving
0.1	13-09-21	Vorm, opzet en structuur
0.2	12-10-21	Inhoud om dialoog te starten
0.3	18-10-21	Collegiale reactie directie Kennis
0.4	04-11-21	Dialoog datakamer
0.5		Dialoog standaardisatieraad
0.6		ROSA scan
0.7		Consolideren ROSA toetsen
0.8		Opdracht standaardisatieraad
0.9		Opstellen keten architectuur 0.9 door architectuurraad
1.0		Akkoord voor versie 1.0 door datakamer

Inhoudsopgave

1 Inleiding

- 1.1 Veranderopdracht Onderwijsgegevensplatform
- 1.2 Doel van ketenarchitectuur
- 1.3 Validatie, acceptatie en bijstelling van de ketenarchitectuur
- 1.4 Leeswijzer

2 Ketenarchitectuur gegevensplatform

- 2.1 Scope: Beoogde resultaten
- 2.2 Scope: wat niet onder de ketenarchitectuur valt
- 2.3 Randvoorwaarden
- 2.4 Afhankelijkheden van andere ontwikkelingen

3 Principes en uitgangspunten

- 3.1 Ketendoelen
- 3.2 Bestuurlijke principes
- 3.3 Richtende principes
- 3.4 Inrichtende principes
- 3.5 Verrichtende principes

4 Businessarchitectuur

- 4.1 Ketenproces
- 4.2 Bestaande architecturen
- 4.3 Usecase

5 Informatiearchitectuur

- 5.1 Functiegebieden
- 5.2 Eisen en beperkingen
- 5.3 Koppelvlakken

6 Plateauplanning

- 6.1 Plateau 1 Algoritme
- 6.2 Plateau 2 Governance
- 6.3 Plateau 3 Metadata
- 6.4 Plateau 4 Gegevens
- 6.5 Plateau 5 Platform

7 Informatiebeveiliging en dataminimalisatie

- 7.1 Baseline Informatiebeveiliging Overheid (BIO)
- 7.2 AVG en Archiefwet
- 7.3 Wet digitale toegankelijkheid

8 Bijlagen

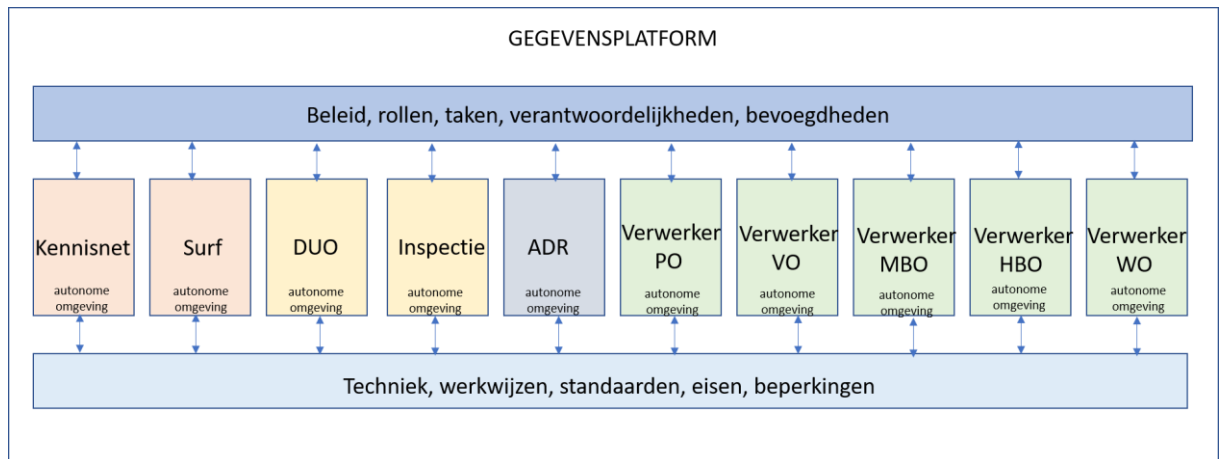
- 8.1 Begrippen en afkortingen
- 8.2 Samenwerking tussen ketenpartners

1 Inleiding

1.1 Veranderopdracht Onderwijsgegevensplatform

Bij data gedreven werken hoort een geactualiseerde architectuur met een nieuwe set van afspraken; van een architectuur waarbij de data in silo's zit naar een fundament met data in platforms.

Het onderwijsgegevensplatform¹ (hierna platform) is een laag die los staat van de proces-applicaties en die toegankelijk is voor de services en de interfaces voor de gegevensuitwisseling.



Figuur 1: beleid, normering en geharmoniseerde inrichting zijn randvoorwaardelijk voor het efficiënt en effectief verstrekken van gegevens.

Een beheerste² informatievoorziening³ (sturing en besluitvorming, processen, gegevens, interactie en óók techniek) berust op afspraken en standaarden. Belangrijke uitgangspunten daarbij zijn dat de gegevens gescheiden zijn van de processen en dat de gegevens in de bron blijven.

Het platform maakt het mogelijk dat gegevens ontsloten worden bij de bron en dat een uitwisselingsproces start met het vaststellen van datakwaliteit.

De set van afspraken en standaarden voor deze systemen is omvangrijk en wordt in Nederland vastgelegd in vijf lagen:

- de grondslagen laag (wet en regelgeving, beleid);
- de organisatorische laag (betrokken organisaties, processen);
- de informatie laag (stelsel van gegevenswoordenboeken en – modellen, BIV classificaties, op grond waarvan pseudonimisering en anonimisering);
- de applicatie laag (bouwstenen, registers), en;
- de netwerk laag (netwerken, knooppunten, technische standaarden voor gegevensuitwisseling).

¹ Het gegevensplatform betreft de interoperabiliteit (de wijze waarop en de voorwaarden waaronder gegevens compliant kunnen worden verstrekt) tussen deelnemende organisaties.

² Beheerst staat voor 'gezag uitoefenen over'. In een beheerste omgeving zien we herhaalbare processen die voorspelbare resultaten opleveren.

³ Het platform en de autonome omgevingen die daar op aansluiten vormen de informatievoorziening

Deelnemers aan het platform doen dat vanuit hun eigen autonome omgeving⁴ echter onder gedeeld beleid en met geharmoniseerde, gestandaardiseerde en uniforme methoden en werkwijzen ten aanzien van de verstrekkingen.

Metadata staan centraal, ze leveren informatie over het gegeven waardoor ieder gegeven zelfverklarend wordt. Zelfverklaring is belangrijk in iedere vorm van gegevensverstrekking maar absoluut noodzakelijk in machine machine verkeer en Artificial Intelligence toepassingen. Alle gegevens zijn daarom altijd van metadata voorzien.

Gezamenlijk anticiperen op toekomstige ontwikkelingen en daar als netwerk maximaal voordeel van hebben is de belangrijkste benefit van het platform.

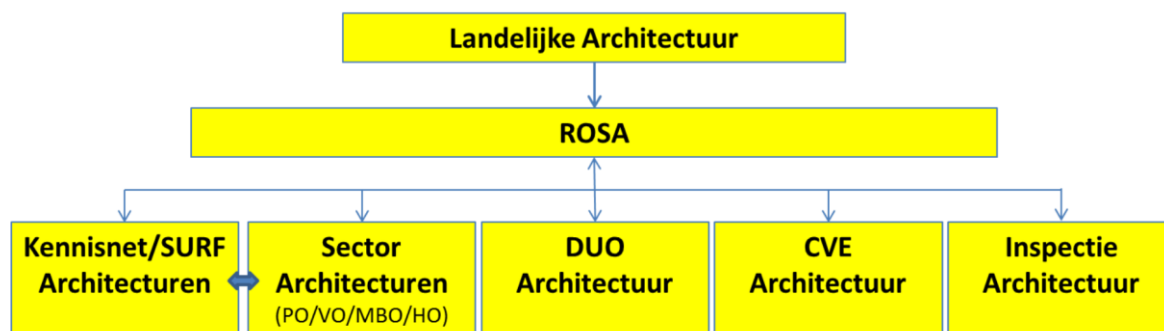
1.2 Doel van ketenarchitectuur

Voor organisaties die werken binnen een keten, en daarmee afhankelijk zijn van andere organisaties, is het van belang hetzelfde beleid en dezelfde standaarden en werkwijzen te hanteren en deze te formaliseren in gezamenlijk gebruik.

Deze ketenarchitectuur draagt bij aan het halen van programmadoelen door de uitwisseling van gegevens in ketenprocessen en de effecten daarvan op de informatie voorziening te beschrijven. We beschrijven daarom wat gemeenschappelijk moet worden besloten en afgesproken en we sluiten uit wat autonoom is.

1.3 Validatie, acceptatie en bijstelling van de ketenarchitectuur

Bij het werken onder architectuur hoort een duidelijke positionering van architectuur, de architectuurrol en het sturen met architectuur.



Figuur 2: samenhang in de onderwijs architecturen

De voorliggende versie van de ketenarchitectuur 'onderwijsgegevensplatform' heeft uitsluitend als doel de dialoog tussen potentiële deelnemers te starten waardoor ze tot een gedragen eindversie kunnen komen.

Deze versie kan bijdragen aan de eindversie, het is aan de deelnemers dat te bepalen. De gedragen eindversie zal worden opgesteld door de architectuurraad in opdracht van de standaardisatieraad.

Architectuur is dynamisch en wordt aangepast bij voortschrijdend inzicht. Validatie van tussen versies en eindversie vindt plaats in de data kamer.

⁴ Deze ketenarchitectuur betreft niet de omgevingen binnen de organisatiegrenzen van de deelnemers an sich maar wel de koppelvlakken in de omgevingen die noodzakelijk zijn voor de verstrekkingen.

2 Ketenarchitectuur gegevensplatform

2.1 Scope: Beoogde resultaten

De ketenarchitectuur is gebaseerd op principes uit de business- en informatiearchitectuur:

- De business architectuur beschrijft het ketenproces en de services. (4.1)
- De informatiearchitectuur beschrijft de koppelvlakken en de interoperabiliteit. (5.1)

De ketenarchitectuur leidt tot een plateauplanning. Om de samenhang tussen zowel de verschillende plateaus als ook de verschillende lagen inzichtelijk te maken is een aantal plateau's beschreven. 6.1 t/m 6.5)

Deelname aan het platform betekent dat de deelnemer in gezamenlijkheid anticipeert op de effecten van trends, te voorziene wetgeving, wetswijzigingen en compliancy maatregelen. De inrichting van het platform zal rekening houden met de actuele en wijzigende condities waaronder deelnemers gegevens mogen verstrekken en gebruiken.

Deelnemers aan het platform leggen de complexe puzzel tussen de uitvoering van publieke en private taken gezamenlijk. Een gezamenlijke ketenarchitectuur maakt de uitkomsten van de puzzel inzichtelijk en herhaalbaar. Zij verkleinen de risico's daardoor maximaal.

Separaat opererende organisaties (niet deelnemend aan het platform) leggen de puzzel zelf.

2.3 Randvoorwaarden

Om de stap te kunnen maken naar realisatie binnen de kaders van de ketenarchitectuur moeten de volgende randvoorwaarden zijn ingevuld:

- De deelnemers zorgen op basis van de ketenarchitectuur voor doorvertaling in hun eigen organisatie;
- Inzet van de benodigde capaciteit uit de organisaties van de deelnemers;
- Inzicht in voortgang en knelpunten;
- Wetgeving vormt de basis voor het aanpassen van koppelvlakken en interoperabiliteit;
- Ketenprocessen zijn beschreven en hebben draagvlak bij deelnemers;
- Middelen worden tussen ketenpartners gedeeld indien mogelijk.

2.4 Relaties met andere ontwikkelingen

De ketenarchitectuur heeft raakvlakken met de architectuurontwikkelingen in de autonome omgevingen van deelnemers.

Het opstellen van de uiteindelijke ketenarchitectuur en het realiseren van de oplossingen bij deelnemers is daarom verdeeld over een langere periode waarin meerdere plateaus onderkend worden. Vanuit de plateauplanning kan worden ingespeeld op de actuele ontwikkelingen.

3 Principes en uitgangspunten

3.1 Ketendoelen

Uitgaande van overheidspublicaties⁵⁵ is een aantal aspecten en ontwerpprincipes doorlopen. Deze zijn als voorlopig referentiekader gebruikt voor deze ketenarchitectuur.

1. Vanuit van compliancy doelstellingen worden niet méér persoonsgegevens dan noodzakelijk aan niet méér geautoriseerde personen verstrekt dan noodzakelijk. Dit levert dataminimalisatie by design.
2. Vanuit effectiviteit doelstellingen worden gegevens die tussen deelnemers vrij gedeeld mogen worden aan elkaar bekend gemaakt en verstrekt. Dit levert de mogelijkheid rijkere analyses uit te voeren.
3. Vanuit efficiëntie doelstellingen wordt het verstrekken van gegevens geharmoniseerd en gestandaardiseerd. Dit levert de mogelijkheid gegevens te gebruiken vanuit de verschillende bronnen waarbij de gegevens aan de onderzoeker gepresenteerd worden als kwamen zij uit één bron.
4. Vanuit de drie bovenstaande doelstellingen vult het platform de randvoorwaarden in voor de samenwerking tussen de deelnemers die omgaan met de complexiteit tussen de stelselverantwoordelijkheid van de overheid en de autonomie van de in belangrijke mate private onderwijsinstellingen.

3.2 Bestuurlijke principes

1. Er is sprake van een duidelijk ketendoel, waarbij de deelnemers van elkaar afhankelijk zijn en zich aan de gezamenlijke opgave verbinden.
2. De veranderopgave wordt op strategisch (richting), tactisch (inrichting) en operationeel niveau (verrichten) bestuurd.
3. Het gezamenlijke meerjarenplan is gebaseerd op afspraken, er is een gezamenlijke verantwoording over de gezamenlijke ketenresultaten zodat deze op alle partners afstralen.
4. Er is overeenstemming over de architectuur(principes) en de ketenprocessen. De ketenarchitectuur geeft richting aan de informatiehuishouding van de keten.
5. Er is overeenstemming over het beheer van de ketenvoorzieningen en van de gegevens.
6. De deelnemers zijn op de hoogte van ieders belangen en respecteren elkaar daar in. Ze delen kennis en voeren de dialoog over verschillen in inzicht.
7. Een onafhankelijke ketenregisseur agendeert het ketenbelang en stemt af met (potentieel) belanghebbenden buiten de keten.
8. Elke deelnemer kan escaleren in de eigen besturingslijn.

⁵⁵ [Onderzoeksrapport OCW \(publicinnovation.nl\)](https://publicinnovation.nl)

3.3 Richtende principes

1. Autonomie en wederzijdse afhankelijkheden bepalen mede de inrichting van het platform.
2. Het platform dient een gezamenlijk erkende maatschappelijke opgave.
3. Deelname aan het platform is gebaseerd op samenwerking en vertrouwen.
4. Wettelijke taken leiden tot afbakeningen.
5. Een gegeven wordt enkelvoudig beheerd in een daartoe aangewezen omgeving.
6. Gegevens worden gevirtualiseerd verstrekt via het platform.
7. Gegevens worden meervoudig gebruikt in alle omgevingen.
8. Het platform harmoniseert, standaardiseert en uniformeert de verstrekkingen.
9. De regie op het platform is gedistribueerd onder de deelnemers op basis van gezamenlijke kaders.
10. Geen van de deelnemende organisaties mag in staat worden gesteld om een gehele procescyclus te beheersen.
11. Het platform is duurzaam en compliant.

3.4 Inrichtende principes

1. De informatievoorziening is geïntegreerd.

Gebruikers willen direct toegang tot alle voor hen relevante informatie. Applicaties worden optimaal geïntegreerd zodat gebruikers een geïntegreerde informatievoorziening wordt geboden. Processen en applicaties dienen relatief los te zijn gekoppeld zodat ze ook los van elkaar kunnen worden aangepast.

2. De informatievoorziening overschrijdt organisatiegrenzen.

Onderzoek vindt steeds meer plaats in een internationale setting van virtueel georganiseerde onderzoekers, waarbij verschillende instellingen betrokken zijn. Hiervoor is het nodig dat virtuele omgevingen en organisaties kunnen worden verbonden. Organisaties behouden hun autonomie en maken afspraken over de federatieve samenwerking op alle niveaus.

3. De informatievoorziening is tijd, plaats en device onafhankelijk toegankelijk.

Mensen leren en werken op het tijdstip en de plaats waarop het hen het beste uitkomt. Dit gedrag en de behoeften van gebruikers moeten door de informatievoorziening worden gefaciliteerd.

4. Archiefwaardige gegevens worden op een geschikte wijze bewaard.

Archivering is een belangrijke randvoorwaarde voor het kunnen realiseren van diverse gebruiksdoelen zoals FAIR, reproduceerbaarheid of controleerbaarheid van onderzoek.

5. De kwaliteit van gegevens wordt expliciet geborgd.

De kwaliteit van gegevens is de mate waarin deze geschikt zijn voor het bedoelde gebruik. Elk soort gebruik stelt eigen eisen op de verschillende dimensies zoals accuraatheid, compleetheid, actualiteit, beschikbaarheid, integriteit en vertrouwelijkheid. Om (her)bruikbaarheid van gegevens vast te stellen en te ondersteunen moet de kwaliteit expliciet worden geborgd.

6. Gegevens zijn beveiligd op basis van hun risicoclassificatie.

Beschikbaarheid, Integriteit en Vertrouwelijkheid (BIV) van gegevens worden als kernonderdeel van informatiebeveiliging gezien. Incidenten waardoor de beschikbaarheid, integriteit of vertrouwelijkheid van gegevens in gevaar wordt gebracht zorgen voor ernstige ontregeling van deelnemers. Het is daarom belangrijk de risico's expliciet te maken. Hierdoor kunnen de meest passende maatregelen worden genomen en worden overmatige maatregelen vermeden.

7 Gegevens zijn gescheiden van processen.

Proceslogica en gegevens zijn gescheiden. Op het platform wordt regie op de gegevens gevoerd door deze qua modellering en ontsluiting te standaardiseren.

3.5 Verrichtende principes

1. De gegevensverzameling is een geïntegreerd geheel.

Voor alle gegevens worden dezelfde principes gehanteerd, er wordt geen onderscheid gemaakt tussen gegevenstypen.

2. Eenduidige gegevensdefinities.

Alle gegevenstypen zijn eenduidig gedefinieerd in een semantisch gegevensmodel. Gegevens en berichten worden zo veel mogelijk volgens (inter)nationale standaarden en generieke bouwstenen gedefinieerd. Wanneer geen standaard voorhanden is, wordt door de deelnemers een gemeenschappelijke standaard nagestreefd.

3. Eén bron per gegevenstype.

Voor alle gegevenstypen is er één bron die leidend is. In de bron heeft gestructureerde opslag de voorkeur boven niet-gestructureerde opslag. De bron is geëigend voor het type gegevens.

4. Professioneel gegevenseigenaarschap.

Elk gegevenstype kent een eigenaar en een beheerder. De verantwoordelijkheid voor de kwaliteit en het beheer van de gegevens ligt bij de eigenaar.

5. Informatie over gegevens.

Gegevens en berichten worden voorzien van metadata. Metadata worden geregistreerd op het moment dat gegevens worden ontvangen of gewijzigd, bij voorkeur geschiedt dit automatisch.

6. Heldere eisen t.a.v. gegevenskwaliteit.

Eisen ten aanzien van de gegevenskwaliteit zijn helder geformuleerd. Gegevens voldoen aan de benodigde kwaliteitseisen, niet meer en niet minder.

7. Professioneel gegevensbeheer.

Het beheren van de kwaliteit van de gegevens en het naleven van principes en richtlijnen verloopt via daartoe ingerichte processen en procedures die voor alle deelnemers helder zijn. Voor de inrichting van deze beheerprocessen wordt gebruik gemaakt van bekende best practices zoals ASL, ITIL en BiSL.

8. Nakomen van wettelijke plichten m.b.t. gegevens.

Bij het opvragen, inwinnen, gebruiken, verwerken, bewaren en beschikbaar stellen van gegevens worden wettelijke regels en plichten in acht genomen.

9. Beschikbaarheid van gegevens.

Gegevensverzamelingen die gebruikt worden door meerdere overheidsorganisaties worden elektronisch beschikbaar gesteld.

4 Businessarchitectuur

4.1 Ketenproces

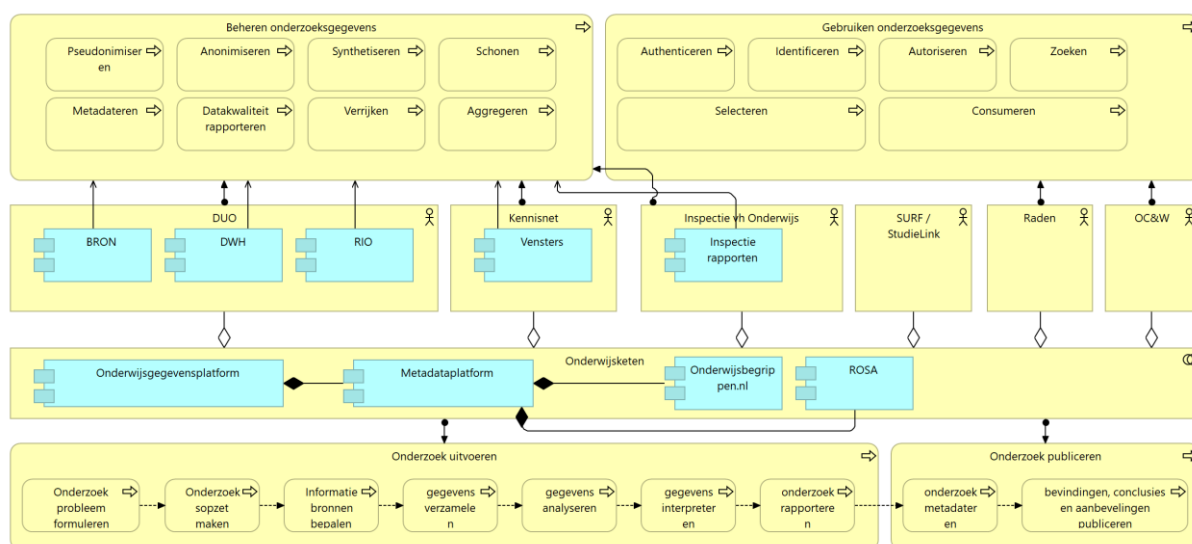
Het platform vormt een voorziening die de onderwijsketen bij het verstrekken van gegevens tussen deelnemers faciliteert. De verstrekkingen leiden in eerste instantie tot het kunnen uitvoeren van onderzoek en ondersteuning, in een latere fase wellicht ook het primaire proces.

Binnen de voorziening worden processen uitgevoerd waardoor onderzoeksgegevens worden beheerd en gebruikt.

- Sub processen binnen 'beheer' zijn onder meer: het pseudonimiseren, anonimiseren, synthetiseren, schonen, metadateren, verrijken en aggregeren van onderzoeksgegevens. Overkoepelend daaraan bestaat het proces kwaliteitsmanagement dat een 7 x 24 dashboard levert. Deze sub processen worden uitgevoerd door de bronhouder.
- Sub processen binnen 'gebruik' zijn onder meer: authenticiseren, identificeren, autoriseren van personen. Zoeken, selecteren en interpreteren van gegevens voorafgaand aan het feitelijke onderzoek.

Het publiceren van de verticale metadata (begrippen, domeintabellen en informatiemodellen) is een kernproces binnen het platform. Dit proces resulteert in het publiceren van het gehele gegevens begrippenkader aan alle deelnemers en belangstellenden. Sprekende voorbeelden daarvan zijn de website Onderwijsbegrippen.nl waarop begrippen uit het Onderwijsveld een plek hebben evenals de ROSA waar informatiemodellen voor de onderwijsketen staan gepubliceerd.

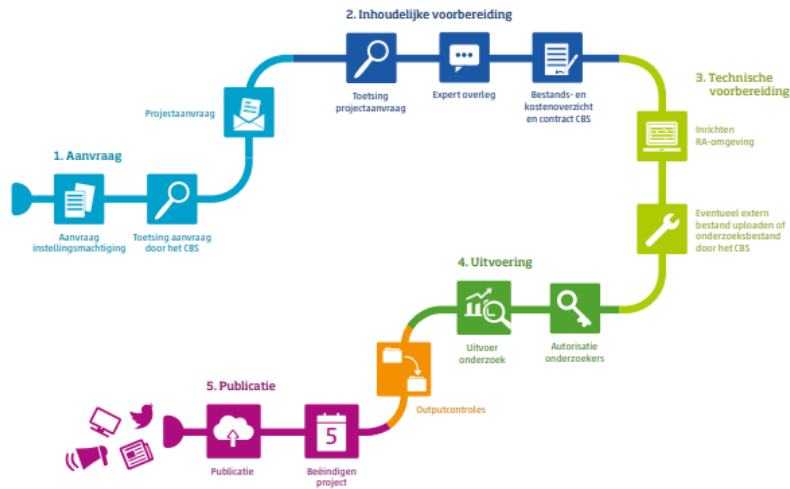
Processen worden vooraf gevalideerd op basis van standaarden en normenkaders zoals CMMI, ISO, ASL, ITIL, BSL, nationale en onderwijsarchitecturen en wetgeving waaronder prominent de AVG. Deze wijze van ontwikkelen levert transparante processen op waar eenvoudig op toegezien kan worden 'compliance by design'. Steekproeven volstaan voor de auditering, de belasting voor deelnemers is minimaal bij een maximale betrouwbaarheidsstelling.



Figuur 3: de samenstellende elementen in het herhaalbare uniforme ketenproces

Als voorbeeld van een te voorzien proces "werken met onderwijsdata" is gekozen voor het sterk vergelijkbare proces 'CBS-microdata' (figuur 4).

Stroomschema werken met CBS-microdata



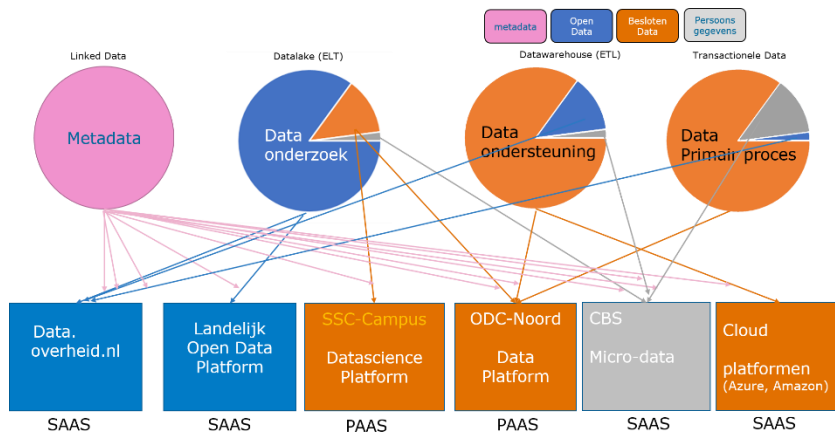
Figuur 4: proces 'werken met CBS-microdata'

4.2 Bestaande architecturen

Verondersteld wordt dat:

- indien een organisatie aan het platform wil deelnemen kan hij dat stapsgewijs in eigen tempo en belangenafweging doen;
 - de volgende acties zijn daarbij voorzien:
 - vaststellen volledigheid en niveau van de 16 vaardigheden;
 - vaststellen adequate governance (besturing en besluitvorming);
 - gegevens geschikt maken voor gebruik vanuit de bron (input / output)
 - metadateren
 - datakwaliteit
 - anonimiseren
 - pseudonimiseren
 - toepassen van standaarden.
 - procesvalidatie op basis van normenkaders
- de omgevingen die op het platform aansluiten zijn ontwikkeld onder architecturen die afgeleid zijn van de ROSA;
- omgevingen van deelnemers elkaar benaderen via het platform door middel van uniforme koppelvlakken;
- deelnemers vóór deelname getoetst zijn aan de 16 vaardigheden;
- in te zetten middelen bij voorkeur worden betrokken bij deelnemers aan het platform of commerciële aanbieders in plaats van deze zelf te ontwikkelen. (figuur 5)

Ketenarchitectuur onderwijsgegevensplatform



Figuur 5: gegevensplatform architectuur SOLL

Metadata management staat centraal in de opzet en werking van het platform. Deelnemers dienen dezelfde taal te spreken en elkaar te begrijpen. Gegevens die gebruikt worden dienen zelf verklarend te zijn in iedere vorm van gegevensverkeer zoals H2H, H2M, M2M en in Artificial Intelligence (AI) toepassingen.

4.3 Usecase

Het onderwijsgegevensplatform maakt efficiënte controles mogelijk.

Onderzocht is óf en hoe nieuwe middelen gebruikt kunnen worden in de accountantscontrole op de leerlingtelling in het voortgezet onderwijs.

Het onderzoek is uitgevoerd door middel van kwantitatief en kwalitatief onderzoek. Twee accountants die controles uitvoeren bij onderwijsinstellingen hebben aan het onderzoek deelgenomen en relevante informatie geleverd.

De accountantscontrole op de leerlingtelling in het voortgezet onderwijs kan verbeterd worden door:

- 1) Het opzetten van een koppeling tussen het Verzuimregister en BRON. Hierdoor kan een signaallijst worden opgesteld. Met deze signaallijst kan gecontroleerd worden of leerlingen die geregistreerd verzuimer zijn, bekostigd worden.

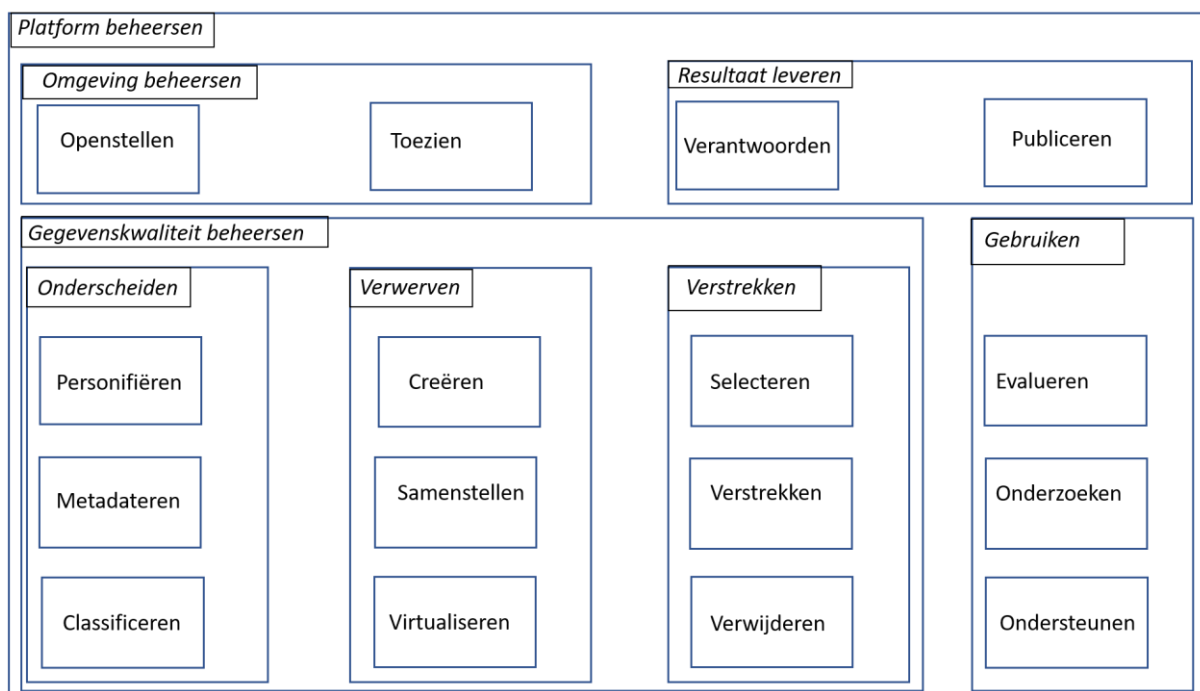
Het is hierbij van belang om de gegevens die gecontroleerd worden te inventariseren en te analyseren. De koppeling kan eenvoudig gerealiseerd worden.

- 2) Het wijzigen van de controle op de beschikkingen voor Praktijk Onderwijs (PRO) en Leerweg Ondersteunend Onderwijs (LWOO).

Deze kunnen vervangen worden door steekproeven, uitgevoerd door de Accountants Dienst Rijk (ADR) of de Inspectie van het Onderwijs (IvO).

5 Informatiearchitectuur

5.1 Functiegebieden



Figuur 6: de 4 bedrijfsfuncties en 16 deelfuncties binnen het platform

Platform beheersen

Het platform heeft een samenwerkingsverband tussen deelnemers. Anders dan bij commerciële platformen waar één partij de voorwaarden bepaalt wordt in het netwerk in gezamenlijkheid bepaald wie welke rol inneemt en hoe de governance wordt ingericht.

Het platform betreft het gegevensverkeer tussen de autonome omgevingen van de deelnemers. De autonome omgevingen maken onderdeel uit van het platform via hun koppelvlakken die bepalend zijn voor de interoperabiliteit van de omgevingen.

Door de koppelvlakken volgens actuele (inter) nationale standaarden en wetgeving te harmoniseren en te standaardiseren wordt het gegevensverkeer uitlegbaar, transparant en compliant gemaakt. Met efficiëntie- en effectiviteitseffecten (zoals rijkere analyses door het breder kunnen gebruiken van gegevens die gedeeld mogen worden) als resultaat.

De optelsom van alle koppelvlakken vormt, op basis van het gezamenlijke beleid, standaarden, wetgeving, methoden en technieken, het platform.

Een omgeving die aansluit op het platform moet daarom aantoonbaar in staat zijn de 4 bedrijfsfuncties en daarin de 16 deelfuncties uit te voeren. Ontbreekt de functie in de huidige situatie of is deze onvoldoende ontwikkeld volgens de normen van het platform dan worden deze respectievelijk ontwikkeld of naar een hoger niveau gebracht. Vaststellen van het niveau vindt plaats door middel van een ROSA scan.

De vier hoofdfuncties die aantoonbaar in een herhaalbaar proces (figuur 4) uitgevoerd moeten kunnen worden zijn:

- Omgeving (koppelvlakken) beheersen;
- Gegevens (kwaliteit) beheersen;
- Gebruiken;
- Publiceren.

Voor deze versie van de ketenarchitectuur is uit ieder van de 4 hoofdfuncties een aansprekend voorbeeld van een bedrijfsfunctie genoemd. Een volledige beschrijving volgt in naast hogere versies. Een korte typering van de 16 bedrijfsfuncties vindt u in de bijlage 'begrippen'.

Omgeving beheersen

Voorbeeld van bedrijfsfunctie: *Openstellen*

Openstellen geeft invulling aan het principe dat niet méér gegevens dan noodzakelijk worden gedeeld met niet méér personen dan noodzakelijk.

Onder 'openstellen' vallen 'authenticiseren', 'identificeren' en 'autoriseren':

- Authenticiseren: Aan de hand van een authenticatiemiddel controleren van de geclaimde identiteit van een gebruiker.
- Identificeren: gebruikers op basis van een identiteitsverklaring koppelen aan een digitale identiteit met betrouwbare kenmerken.
- Autoriseren: het proces om een recht toe te kennen aan een geauthentiseerde identiteit; het recht van een geauthentiseerde identiteit om gebruik te maken van bepaalde functionaliteit.

Onderzoekers krijgen op basis van verstrekte autorisaties door middel van transacties toegang tot de gegevens die zij nodig hebben voor het uitvoeren van hun taak. Niet anders dan in de primaire onderwijsprocessen wordt verhinderd dat zij op eenvoudige wijze grootschalig gegevens van het platform kunnen kopiëren naar hun autonome omgeving.

Openstellen staat voor 'openstellen voor gebruik' en sluit 'openstellen voor dubbel beheer' uit.

Gegevens(kwaliteit) beheersen

Voorbeeld van bedrijfsfunctie: *Metadateren*

Metadata (informatie over gegevens) zijn noodzakelijk om de gegevens die gebruikt worden voor analyse doeleinden zelfverklarend te laten zijn. Een must in machine machine verkeer en Artificial Intelligence toepassingen waar het platform op anticipeert.

Metadateren vormt een kernpunt in de inrichting van het platform. Voorzien wordt een centrale bedrijfsfunctie hiervoor in te richten. Metadata hebben immers alleen nut wanneer zij bij alle deelnemers hetzelfde inzicht in de betekenis van gegevens geven.

Een greep uit het scala van potentiële metadata:

- Herkomst van het gegeven;
- Bron waarin het gegeven vastgelegd is;
- Doel waarvoor het gegeven gebruikt mag worden;
- Grondslag de wettelijke taak waarvoor het gegeven gebruikt mag worden;
- Verwerkingsdatum;
- Vervaldatum.

Gebruiken

Voorbeeld van bedrijfsfunctie: *Evalueren*

Het verzamelen, interpreteren en presenteren van informatie om de waarde van een resultaat of proces te bepalen:

- Verzamelen: een reeks instructie waarmee (op basis van een vooraf gestel doel) gegevens die noodzakelijk zijn voor het kunnen uitvoeren van de evaluatie op het platform worden uitgevraagd
- Interpreteren: op basis van metadata bepalen welke gegevens voor het onderzoeksdoel geschikt zijn.
- Presenteren: het geschikt maken van de onderzoeksresultaten voor het tonen (grafisch, rapportage, site) aan belanghebbenden en belangstellenden.

In de sector onderwijs staat de term 'evalueren' voor het gebruiken van gegevens met als doel het vaststellen van de effecten van het onderwijsbeleid in de maatschappij (AWB).

Evalueren is door de wetgever gerelateerd aan het bekostigen van instellingen van onderwijs. Verdere grondslagen liggen vast in onderwijswetgeving.

'Evalueren' beperkt zich niet tot de wettelijke taak. Ook voor deelnemers die private taken uitvoeren en voor belanghebbenden zoals ouders is het zeer interessant te weten wat de effecten van hun uitvoering op de maatschappij zijn.

Resultaat leveren

Voorbeeld van bedrijfsfunctie: *Publiceren*

Publiceren betreft het wereldkundig maken van de onderzoeksresultaten.

Het publiceren van de onderzoeksresultaten is een verbijzonderde bedrijfsfunctie die onder verantwoordelijkheid van de opdracht gevende deelnemer valt.

De onderzoeker is niet noodzakelijkerwijs de publicist doordat de gepresenteerde onderzoeksresultaten naar de norm van opdracht gevende deelnemer wel of niet publicabel kunnen zijn. De bedrijfsfunctie 'publiceren' wordt om die reden los van de functies 'onderzoek', 'ondersteuning' en 'evaluatie' ontworpen en ingericht.

6 Plateauplanning

De succesvolle ontwikkeling van het bol.com platform⁶ (klein beginnen met boeken, risicoloos uitbreiden met tweedehands boeken, daarna detailhandel aanbieders en nu alles wat maar te bedenken is) heeft de inspiratie geleverd voor de plateauplanning van het onderwijsgegevensplatform. Bol.com is een Nederlands succesverhaal en weet al vele jaren lang vriend en vijand te verbazen. Een fascinerend verhaal over groot durven denken, omgaan met succes en teleurstellingen en de drang naar continue vernieuwing. Het bol.com devies 'klein beginnen vanuit groot denken' klinkt in deze ketenarchitectuur en de plateau planning door.



Het Onderwijsgegevensplatform is namelijk véél meer dan alleen een technische omgeving, vanuit het perspectief van de NORA⁷ is het essentieel dat er voldoende aandacht is voor alle 5 lagen : grondslagenlaag, organisatorische laag, informatielaag, applicatielaag én de netwerklaag.

In deze plateauplanning beginnen we dan ook vanuit hetgeen het meest belangrijk is in plaats van, zoals te doen gebruikelijk, vanuit het implementeren van een tool.

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
(Data + Integratie architectuur, inrichting Datahub / proxy)

Applicatielaag
(bouwstenen, registers)

Plateau 5 : Omgeving op orde
(data platform)

Netwerklaag
(netwerken, knooppunten)

Figuur 7: veranderplateau's conform NORA

Een belangrijk aspect bij de keuze voor een aanpak en veranderplateau's is een geschikte usecase en de vertaling daarvan naar de plateau's. De gekozen usecase betreft het efficiënter inrichten van de controle op de leerlingtelling in het voortgezet onderwijs (zie 4.3).

6.1 Plateau 1 : Algoritme + doel op orde

Als overheid zorgen we dat onze werkwijze transparant is⁸ en dat de verantwoording aan belanghebbenden mogelijk is. We omarmen nieuwe technologieën wanneer die ons werk beter en efficiënter maken.

⁶ [Het geheim van bol.com - het boek](#)

⁷ <https://www.noraonline.nl/wiki/Vijflaagsmodel>

⁸ [Algoritmen register Amsterdam](#)

Een algoritme is een wiskundige formule waarmee aangetoond wordt welke gegevens waarom voor welk doel gebruikt worden. In programmeertaal is het een instructie, een stukje code, om een probleem op te lossen. We noemen iets kunstmatige intelligentie (Artificial Intelligence) wanneer die algoritmes op basis van gegevens of signalen uit hun omgeving zelfstandig beslissingen nemen én daarvan leren.

Plateau 1 heeft als focus om grip op algoritmes⁹ te krijgen. Dat betekent voor het Onderwijsgegevensplatform óók dat er grip op de wet nodig is. In dit plateau wordt relevante (Onderwijs)wetgeving gekoppeld aan doelen zodat dit een goede basis vormt voor een verwerkingsregister¹⁰.

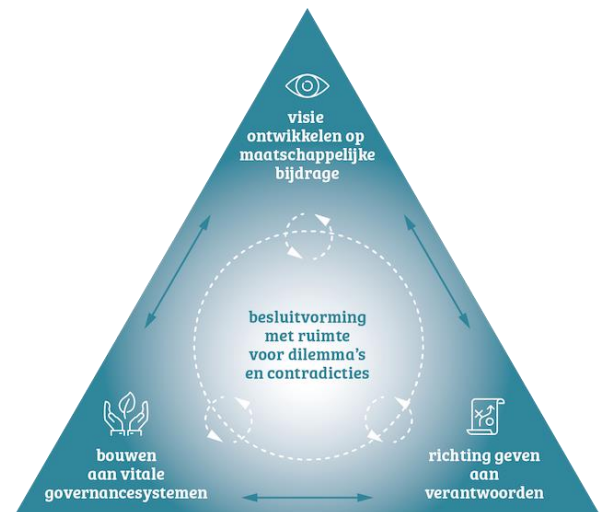
6.2 Plateau 2 : Governance op orde

Good Governance¹¹ is een dragende kracht voor groei en ontwikkeling van samenwerkingsverbanden in ketens. Dit is nodig om relevante maatschappelijke doelen en economische waarde te realiseren. Voor de keten is deze noodzakelijk om transparant te kunnen functioneren waarbij helder is welke partij wat toevoegt in de waardeketen.

In de vakliteratuur zijn twee aspecten van Good governance prominent aan de orde. Ten eerste de procedures voor de besluitvorming, zoals de inrichting van de keten. Ten tweede de omgangsvormen en spelregels, de cultuur.

We onderscheiden drie governance domeinen:

1. Visieontwikkeling:
Er is een toenemende oproep aan private organisaties om hun maatschappelijke verantwoordelijkheid te nemen. Dat houdt in dat alle organisaties een visie ontwikkelen op de bijdrage die ze willen realiseren, gebruik makend van hun kerncompetenties.
2. Vitale governance systemen:
In sommige organisaties zijn zoveel regels en procedures dat mensen hun werk niet meer goed kunnen doen. In andere is de besluitvorming zo informeel dat je er eigenlijk ook niet goed mee uit de voeten kunt. De kunst is om de formele regels en de cultuur zo op elkaar af te stemmen dat een flexibel, open systeem ontstaat.
3. Verantwoorden:
Externe toezichthouders verlangen verantwoording. Nieuwe manieren van verantwoorden richten zich steeds minder op cijfers over het verleden en steeds meer op stuurinformatie voor de toekomst. Nieuwe inzichten werken door in de organisatie, waardoor ook intern nieuwe manieren van rapporteren ontstaan.



Figuur 8: Good Governance

Deze drie domeinen worden in dit plateau uitgewerkt vanuit de usecase. Als leidraad en voorbeeld van best-practices willen we gebruik maken van het governance model van SSC-Campus inclusief de daar opgedane lessons-learned.

Het model beschrijft afspraken over de sturing die waarborgen dat een goede en op de behoeften toegesneden ondersteuning geboden wordt aan de (primaire) processen van

⁹ [Grip op algoritmische besluitvorming bij de overheid \(rathena.nl\)](https://rathena.nl/)

¹⁰ [Verantwoordingsplicht | Autoriteit Persoonsgegevens](#)

¹¹ [Visie op Good Governance - Good Governance](#)

de deelnemers. Deze afspraken bevorderen een efficiënte en effectieve realisatie van de doelstelling en worden gekenmerkt door een open wijze van communiceren en verantwoording afleggen ten behoeve van belanghebbenden.

Het model brengt evenwicht aan tussen de belangen van de opdrachtgevers en opdrachtnemers. Tenslotte draagt de governance bij aan samenwerking tussen de opdrachtgevers.

6.3 Plateau 3 : Metadata op orde

Deelnemers willen kwalitatief hoogwaardige¹² gegevens. Daartoe richten we een rechtmatig en doelmatig metadata platform in dat tegemoet komt aan de verwachtingen van opdrachtgevers en het publiek.

Ook dit metadata platform voldoet auditeerbaar aan wettelijke kaders. Uitgangspunten (besluiten, regels, processen, organisatie e.d.) worden daartoe bijeengebracht voor de (toekomstige) inrichting van het Onderwijsgegevensplatform zodat de metagegevens volledig, juist en actueel beschikbaar zijn. Alle vakdisciplines op het gebied van ontwikkeling en beheer worden hierbij betrokken.

Metagegevens zijn essentieel voor het vinden, begrijpen en gebruiken van gegevens. Dit is analoog aan de FAIR-principes¹³.

Vanuit Archief- en Informatiemanagement (AIM) worden het nut en de noodzaak van metagegevens als volgt gedefinieerd: "Metadata vormen een hulpmiddel om documenten juist te kunnen vinden, te kunnen begrijpen en er op te kunnen vertrouwen dat zorgvuldig wordt omgegaan met gevoelige informatie en dat informatie tijdig vernietigd wordt of juist bewaard blijft."

DMBOK, 'Data Management Body of Knowledge' deelt soorten metagegevens in de categorieën business, technical en operational in, verwijzend naar de plek waar metagegevens ontstaan.

De NORA definieert de term als volgt: "Gegevens die context, inhoud, structuur en vorm van informatie en het beheer ervan door de tijd heen beschrijven".

Metagegevens zijn, in de definitie van Broekmans, 1987: "Metagegevens zijn beschrijvingen van de gegevens". De auteurs geven de volgende voorbeelden:

- De betekenis of definitie van gegevens;
- De vorm (lengte) of waardenbereik;
- De eigenaar/beheerder;
- De structuur (de relatie tussen de gegevens onderling);
- Herkomst en gebruik (datalineage).

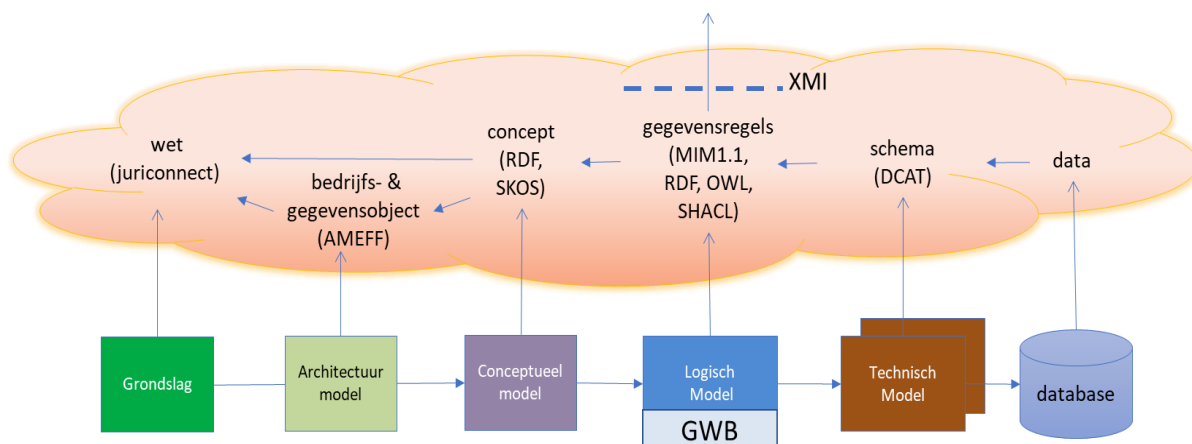
Er zijn 3 soorten datalineage:

1. Verticale datalineage toont het ontstaan van een datavereiste vanuit regelgeving naar implementatie in een datamodel op metadataniveau.
2. Horizontale datalineage toont de toewijzing van brongegevens aan doeluitvoer op metagegevensniveau. Het toont de functionele transformatielogica van hoe brongegevens worden getransformeerd naar een doeleindtoestand.
3. Fysieke datalineage demonstreert de feitelijke gegevensstroom van bronsysteem naar rapportageoplossing, ter ondersteuning van de metadata-architectuur van de datagestuurde rapportageomgeving

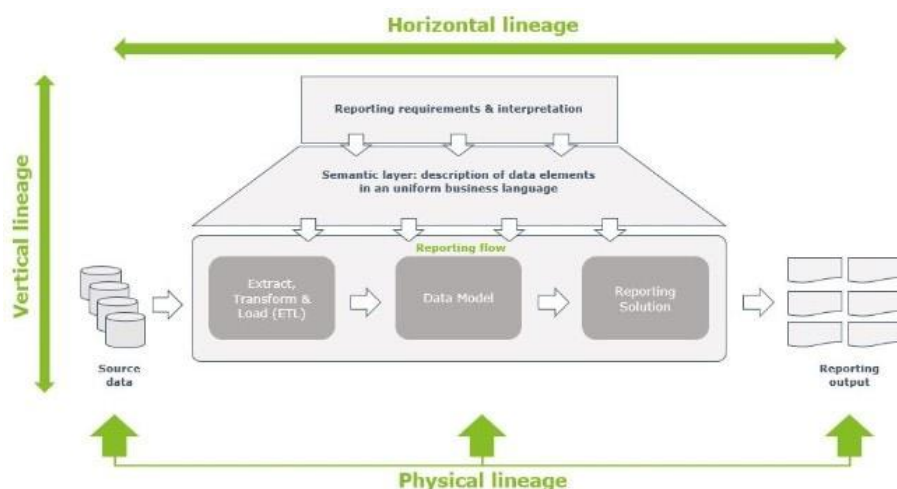
¹² [Gegevenskwaliteitsraamwerk NORA](#)

¹³ <https://www.go-fair.org>

Ketenarchitectuur onderwijsgegevensplatform



Figuur 9 : verticale datalineage over metadata



Figuur 10: In de metadata voorziening worden verticale data lineage technieken gebruikt.

De Parlementaire Onderzoekingscommissie Kinderopvangtoeslag(POK), het rapport van de Tijdelijke commissie Uitvoeringsorganisaties (TCU) en de rapporten Werk aan Uitvoering (WaU) wijzen op veranderende maatschappelijke verwachtingen over transparantie en herleidbaarheid van wettelijke uitvoeringsprocessen. Onderwijs wordt in deze rapporten met name genoemd.

Bij de inrichting van het platform anticiperen we daarom op de thema's:

- Gegevensmodellering
De betekenis en structuur van gegevens zijn vastgelegd in modellen (met een centrale rol voor het gegevenswoordenboek);
- Contextmetadata
Dit gaat over een beschrijving van de context waarin gegevens zijn ontstaan, bijvoorbeeld het herkomst, historie van gegevens, logging e.d.;
- Gegevensportfolio
Dit is een centrale index die beschrijft waar gegevens zijn opgeslagen en wanneer ze op reis gaan. Deze index bevat de gecreëerde artefacten (bijvoorbeeld databases, gegevensleveringen etc.).

Deze thema's gaan over de transparantie en datalineage vanaf wet/grondslag, via architectuur, betekenisbepaling en structuur tot en met een veld in de database of een distributie.

In dit plateau komen alle deelnemers gezamenlijk tot ingerichte methoden, technieken, handboeken en voorschriften (MTHV) en brengen zo hun metadata op orde.

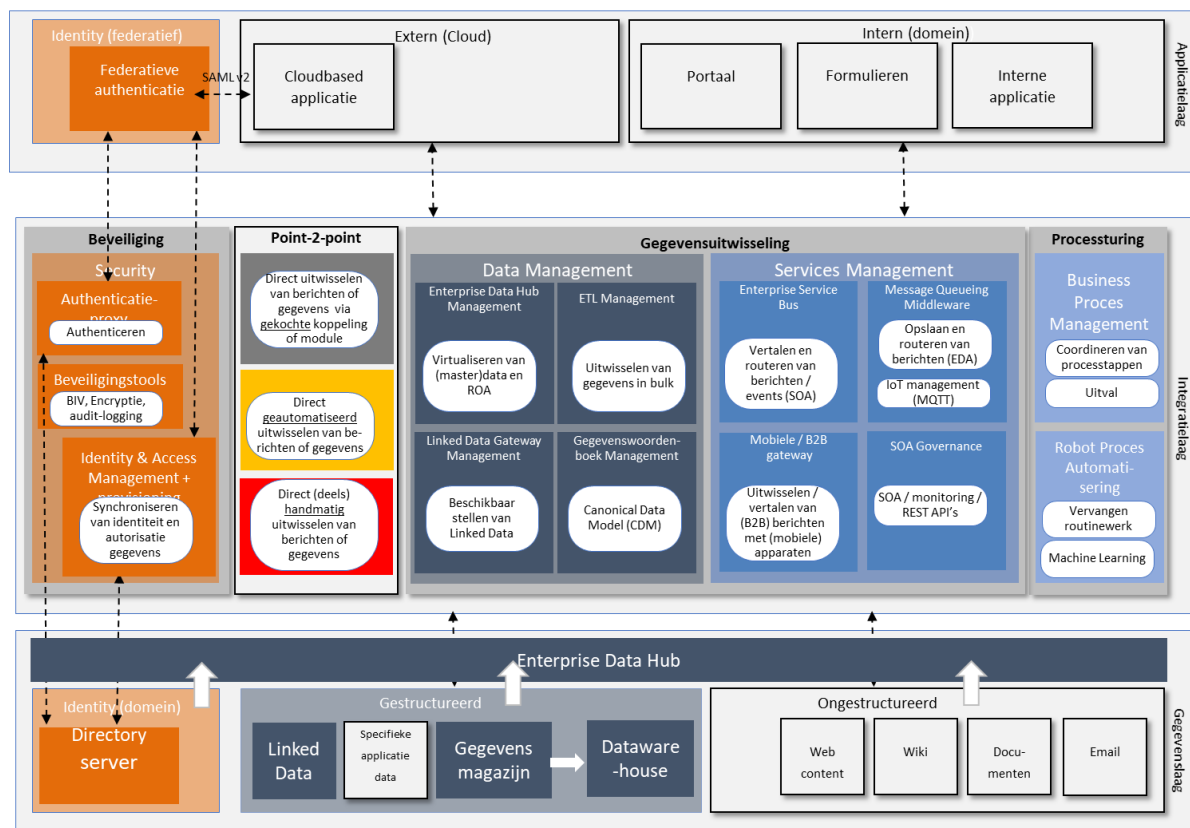
6.4 Plateau 4 : Gegevens op orde

Naast de Data Architectuur wordt invulling gegeven aan de Integratie Architectuur. Er zijn uiteenlopende vormen van integratie beschikbaar, al dan niet gebruik makend van ondersteunende middleware-oplossingen. De omvang van integratieprojecten kan enorm verschillen van een simpele API-aanroep tot de implementatie van een complete en grootschalige ESB-stack die een volledige integratie van een pluriform applicatielandschap moet realiseren.

De kans op succes is sterk afhankelijk van de keuzes die in een integratieproject gemaakt worden. Vanuit het oogpunt risico's te mitigeren is het belangrijk op de hoogte te zijn van de bestaande soorten applicatie-integratie, wat de kenmerkende verschillen zijn, en waar op te letten bij het maken van keuzes.

De figuren 11 en 12 laten zien wat nodig is om invulling te geven aan de integratie-architectuur.

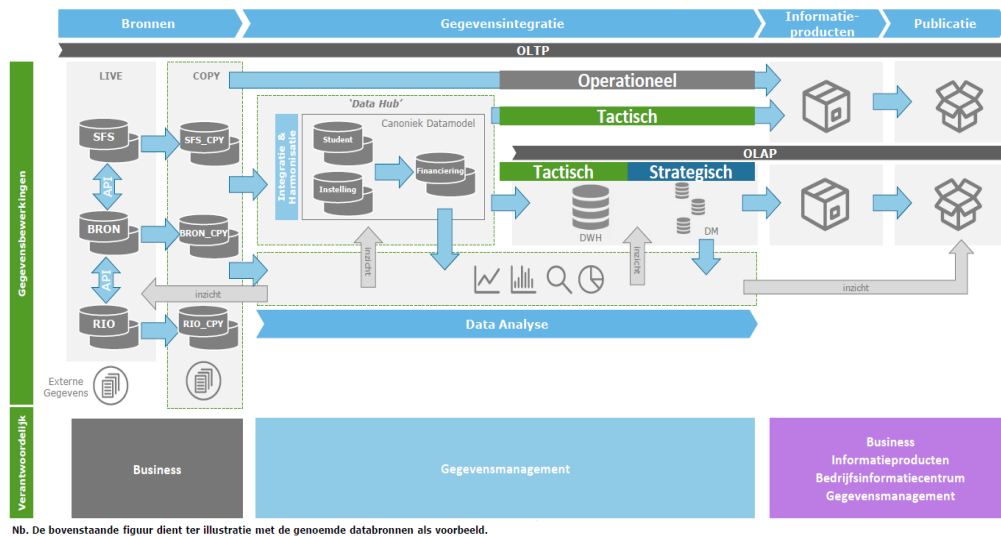
Integratie-architectuur



Figuur 11: referentie integratie architectuur

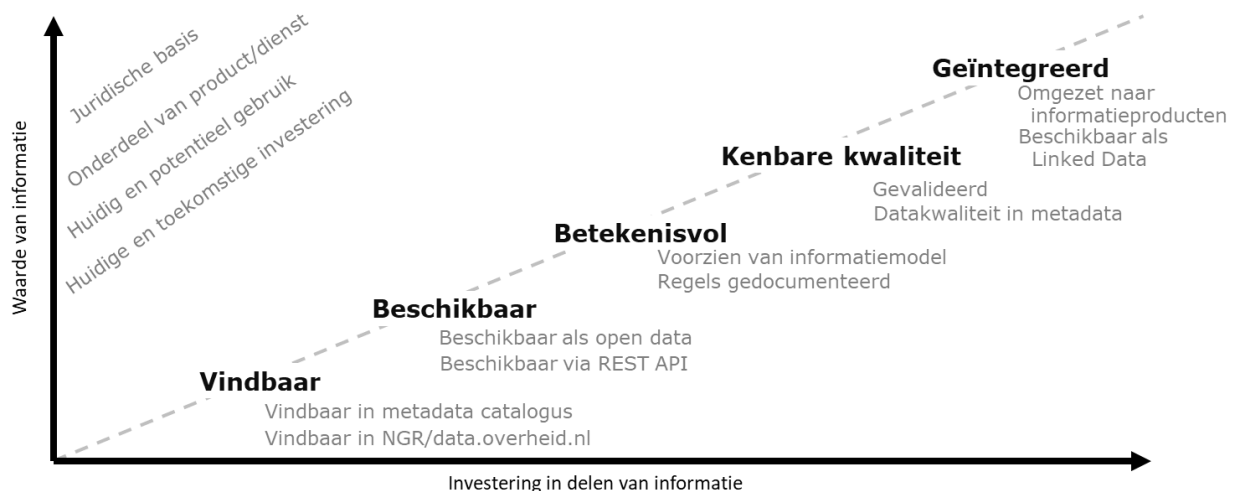
Ketenarchitectuur onderwijsgegevensplatform

Een ingevuld voorbeeld:



Figuur 12: voorbeeld van gegevensstromen, inclusief gebruik van gegevensmagazijn, Data Hub en Data Warehouse

Dit plateau betreft het op orde brengen (per deelnemer) van de gegevens en de daarbij horende voorzieningen (brongegevens blijven immers zoveel mogelijk decentraal staan) voor zowel de opslag als de integratie van gegevens. Figuur 13 geeft inzicht in de richting.

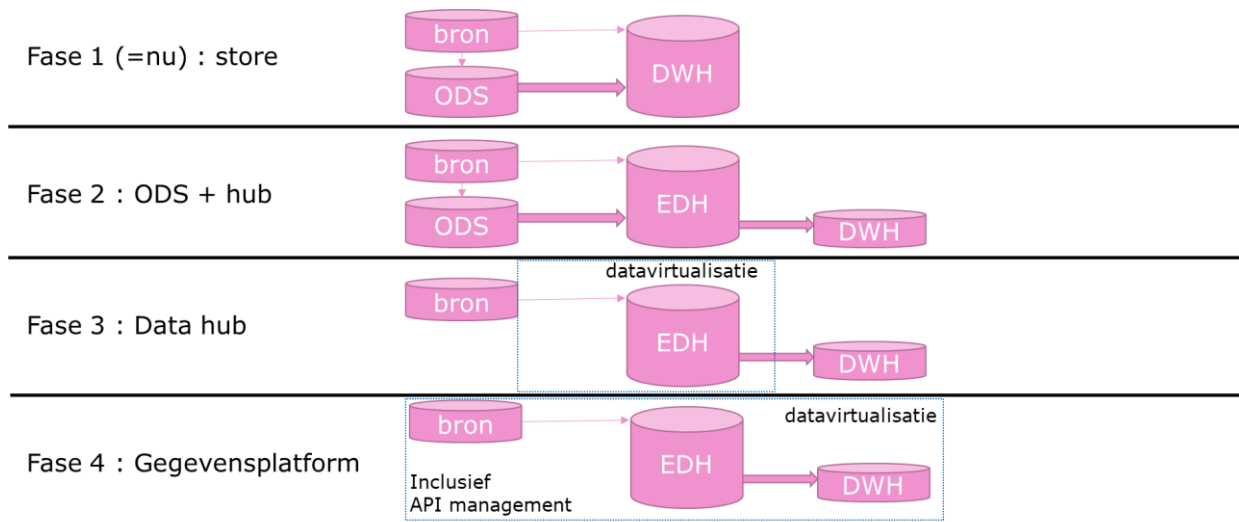


Figuur 13: Data op orde, de waarde van informatie versus de investering

6.5 Plateau 5 : Platform op orde

Dit plateau betreft het op orde brengen van de onderscheiden platformen die beheerd / gebruikt gaan worden in de context van het Onderwijsgegevensplatform. Ook hiervoor geldt dat dit niet één platform is, maar dat dit kan bestaan uit meerdere platformen die specifiek ingericht zijn voor een usecase, bijvoorbeeld een Doorstroommonitor.

Het kan nodig zijn om per deelnemer de eigen gegevensomgeving dusdanig te moderniseren, dat het past bij de requirements voor aansluiting op het Onderwijsgegevensplatform. Daarin kan de inrichting van een Data Hub, Datavirtualisatie en API management een belangrijke rol spelen met als doel om gegevens uiteindelijk zo dicht mogelijk bij de bron te beheeren / gebruiken.



Figuur 14: voorbeeld fasering inrichting modern gegevensplatform

Vanuit:

- Fase 1 (huidige situatie) zetten we de stap naar;
- Fase 2 met daarin een nieuwe omgeving (Data Hub) die naast de huidige situatie wordt gezet en welke voldoet aan alle eisen rondom een modern gegevensplatform;
- Fase 3 waarin de dubbeling wordt verwijderd en een Data Hub zonder Operational Data Store gerealiseerd wordt;
- Fase 4 waarin deze methodiek door alle deelnemers wordt ingezet en voor alle verstrekkingen het gegevensplatform beschikbaar is;
- Fase 5 tenslotte levert onder meer API management omdat ook de bronnen zelf nu onderdeel vormen van het gegevensplatform.