



UITVOERING WERKNEMERSVERZEKERINGEN

Projectstartarchitectuur

Project Datafabriek

Versie 3.0

Datum: 11 februari 2021

VERSIEBEHEER

Versie	Datum	Status	Korte beschrijving aanleiding / wijziging
<i>(diverse eerdere versies)</i>			
2.0	17/07/2020	Definitief	Vastgesteld door AB+ (16/07/2020)
2.1	02/01/2021	Concept	Uitkomsten fase 1-4 verwerkt. Impact scope/aanpak-wijziging 2021 verwerkt. Beperkte cosmetische wijzigingen Opgeleverd aan LA GD en PM DF, ter inbreng in AB+
2.2	25/01/2021	Concept <i>(Tussenversie voor project-interne review)</i>	Opmerkingen DWH architecten en Implementatie verwerkt, m.u.v. open punten m.b.t. scope.
2.9	08/02/2021	Concept	Uitkomsten scope-afstemmingen verwerkt; voor vaststelling AB+
3.0	11/02/2021	Definitief	Door de AB+ geaccordeerde.

(zie leeswijzer voor verschillen tussen PSA 2.0 en PSA 2.1)

Versiebeheer: Gert-Jan Kooren

Opgesteld door: Gert-Jan Kooren m.m.v. Ruud Keller

Afgestemd met:	Status	Datum	Naam / Contactpersoon
Architecten DWH en Gegevensdiensten	Geen verder commentaar	25/01/2021	Ilse Klaver
Architect SOM	Geen verder commentaar	25/01/2021	Johan Blad
Projectmanager(s) Datafabriek/SOM	Geen verder commentaar	25/01/2021	Frederik Lindenaar

Bovenstaande betreft alleen afstemming m.b.t. (conceptenversies van) PSA v3.0

Besproken door:	Status	Datum
Architecten GD/DWH	Review commentaar op v2.1 verwerkt	08/01/2021
MT DWH / Implementatie	Review commentaar op v2.1 verwerkt	11/01/2021
Stuurgroep Datafabriek		
Architectuur Board (+)		

Bovenstaande betreft alleen afstemming m.b.t. (concepten voor) PSA v3.0

Distributie: Projectteams DataFabriek/SOM

Architecten DWH en Gegevensdiensten, architect SOM

Leden Architectuur Board en AB+

Leden MT DWH

Leden Stuurgroep Datafabriek

(samenstelling Stuurgroep Datafabriek/SOM nog niet bekend)

Vertrouwelijkheid:

De lezer/gebruiker van dit document wordt geacht de inhoud daarvan vertrouwelijk te behandelen, tenzij uit de toelichting of bronvermelding blijkt dat de informatie als openbaar kan worden beschouwd.

INHOUDSOPGAVE

VERSIEBEHEER	2
INHOUDSOPGAVE	3
MANAGEMENTSAMENVATTING	5
Projectscope en direct gerelateerde lijninitiatieven	5
Drivers	5
Verworven ETL+ tooling	6
Data Integratie Magazijn (DIM)	6
Impact drivers op de oplossingsarchitecturen	7
LEESWIJZER.....	10
1 INLEIDING	11
1.1 Doel project	11
1.1.1 Oorspronkelijk doel	11
1.1.2 Impact van scope changes en voortgang elders op behalen projectdoelen.....	12
1.2 Architectuurdrijfveren	16
1.3 Belangrijkste aandachtspunten.....	17
1.4 Beslis- en discussiepunten	18
2 KADERS	23
2.1 Uitgangsdocumenten	23
2.2 Aanvullende architectuorkaders.....	23
2.3 Project-specifieke uitgangspunten	24
2.4 Voor deze PSA-versie gebruikte (deel)projectplannen.....	24
3 BEDRIJFSARCHITECTUUR.....	25
3.1 Diensten.....	25
3.1.1 Afbakening Diensten	25
3.1.2 Architectuurprincipes Diensten	26
3.1.3 Afwijkingen van architectuurprincipes.....	27
3.1.4 Hoofdstructuur van de oplossing.....	27
3.2 Bedrijfsfuncties	30
3.3 Processen	30
3.3.1 Afbakening Processen.....	30
3.3.2 Architectuurprincipes Processen.....	31
3.3.3 Afwijkingen van architectuurprincipes.....	32
3.3.4 Hoofdstructuur van de oplossing.....	32
3.4 Organisatie & Management	32
4 INFORMATIEARCHITECTUUR.....	33
4.1 Gegevens	33
4.1.1 Afbakening Gegevens.....	34
4.1.2 Architectuurprincipes Gegevens.....	38
4.1.3 Afwijkingen van architectuurprincipes.....	40
4.1.4 Hoofdstructuur van de oplossing.....	40
4.2 Informatievoorziening	40
4.2.1 Afbakening Informatievoorziening.....	40
4.2.2 Architectuurprincipes Informatievoorziening	41

4.2.3	<i>Afwijkingen van architectuurprincipes</i>	43
4.2.4	<i>Hoofdstructuur van de oplossing</i>	43
5	TECHNISCHE ARCHITECTUUR	49
5.1	Software	49
5.1.1	<i>Afbakening Softwarecomponenten</i>	49
5.1.2	<i>Architectuurprincipes Softwarecomponenten</i>	50
5.1.3	<i>Afwijkingen van architectuurprincipes</i>	50
5.1.4	<i>Hoofdstructuur van de oplossing</i>	51
5.1.5	<i>Check Licentie Positie (CLP)</i>	57
5.2	Technische infrastructuur	57
5.2.1	<i>Afbakening Technische infrastructuur</i>	57
5.2.2	<i>Architectuurprincipes Technische infrastructuur</i>	58
5.2.3	<i>Afwijkingen van architectuurprincipes</i>	60
5.2.4	<i>Hoofdstructuur van de oplossing</i>	60
6	INFORMATIEBEVEILIGING EN -BEHEER	63
6.1	Informatiebeveiliging	63
6.1.1	<i>Principes Informatiebeveiliging</i>	63
6.1.2	<i>Afwijkingen van principes</i>	66
6.2	Informatiebeheer.....	66
7	ARCHITECTUURCONFLICTEN EN ONZEKERHEDEN	67
7.1	Architectuurconflicten	67
7.2	Architectuur aannames.....	67
7.3	Openstaande punten.....	67
8	METHODEN, TECHNIEKEN EN HULPMIDDELEN	68
	BIJLAGE A : RATIONALE BREDE BRONONTSLUITING	69
	BIJLAGE B : TYPEN GEGEVENSGEBRUIK DOOR DIM	71
	Gegevens als basis – brede bronontsluiting.....	71
	<i>Belang en rol RLO</i>	71
	<i>Geen weigering aan de poort</i>	72
	<i>Zeer gevoelige gegevens</i>	72
	<i>Zeer gevoelige personen/zaken</i>	72
	<i>Velden met zeer gevoelige waarden</i>	73
	Gegevens t.b.v. verdere structurering – Reference data	73
	<i>Codes + Omschrijvingen</i>	74
	<i>Cross-referencelijsten</i>	74
	<i>Groeperingshierarchieën</i>	75
	<i>Verwerking van niet via bronontsluiting geleverde reference data</i>	75
	Gegevens als parameters	75
	Gegevens t.b.v. beveiliging	76
	BIJLAGE C : AFKORTINGEN	77

MANAGEMENTSAMENVATTING

Projectscope en direct gerelateerde lijninitiatieven

Het project Datafabriek voert grote verbeteringen door in zowel het centrale data warehouse applicatielandschap (op dit moment bestaand uit drie data warehouses, alle in eigendom van Gegevensdiensten) als in de daaraan gerelateerde beheer- en gegevensleveringsprocessen.

In de **oorspronkelijke scope** zou dit resulteren in een nieuw (al dan niet virtueel) IV-organisatieonderdeel (de lijnorganisatie Datafabriek), een nieuwe IV-oplossing (het Data Integratie Magazijn, DIM), gebaseerd op nieuw verworven pakketsoftware (de ETL+ tooling).

E.e.a. zou bestuurd gaan worden door een nieuw "business" organisatieonderdeel (het Data Expertise Centrum, DEC). Inrichting van de oorspronkelijk aan dit DEC toebedeelde vraag- en ketenbesturingsprocessen werd onderdeel van de IV-transitie, en viel dus buiten scope van het project Datafabriek.¹

In de **huidige scope** (begin 2020 gewijzigd, formeel vastgelegd in projectplan 2021) is de inrichting van dat nieuwe organisatieonderdeel niet meer aan de orde; beheer van het Data Integratie Magazijn valt nu binnen de bestaande grenzen van de afdeling DWH, en inrichting van de daarvoor benodigde processen en organisatiestructuren is grotendeels onderdeel geworden van het binnen de lijn belegde "lijn-implementatietraject GD Uitvoeren en IV".

Dit lijn-traject dekt deels ook de inrichting van de processen en organisatiestructuren die oorspronkelijk aan het DEC (en dus aan de IV-transitie) toebedeeld waren. Deze zullen grotendeels binnen Gegevensdiensten Uitvoeren gaan vallen.

Per 1/1/2021 is het project opgesplitst in deelprojecten. Ook is het project SOM (Saneren, Optimaliseren en Migreren S&V-informatieproducten) nauwer met het project Datafabriek verbonden.

In dit document wordt naar het project verwezen als "project Datafabriek", naar de toekomstige IV-lijnorganisatie als "IV-organisatie DIM/DWH", naar de toekomstige business-aansturing als "Vraag- en ketensturing GD t.b.v. DIM/DWH", en naar het nieuwe organisatie-inrichtingstraject als "lijn-implementatietraject GD Uitvoering en IV".

Drivers

De voornaamste drivers van het project Datafabriek zijn:

Hoge prioriteit	Minder hoge prioriteit (maar wel noodzakelijk)
<ul style="list-style-type: none"> • Garanderen stabiliteit & continuïteit van de bestaande dienstverlening • Verkleinen time-to-market • Verbeteren gegevensbescherming ("AVG compliant") 	<ul style="list-style-type: none"> • Meten / verbeteren datakwaliteit • Verlagen TCO (Total Cost of Ownership) • Ondersteunen nieuwe typen ("self service") gegevensgebruik • Ondersteunen nieuwe typen data (b.v. ongestructureerd)

Het project Datafabriek levert hierbij de technische "enablers", en het lijn-implementatietraject "GD Uitvoering en IV" de organisatie-inrichting die, m.b.v. die enablers, de vereiste verbeteringen kan realiseren.

Uitzondering hierop is het deelproject SOM, dat, binnen het project Datafabriek, wél verbeterde organisatie-inrichting in scope heeft. Deze verbeteringen liggen echter aan de afnemerszijde (buiten de IV-organisatie DIM/DWH en het "DEC-equivalent"), en raken daarmee slechts indirect aan de Datafabriek-drijfveren.

¹ Dit "DEC" is beschreven in de adviesnota van Otten uit 2015 en de daarop voortbordurende DA GEIN.

Verworven ETL+ tooling

In de eerste fasen heeft het project ETL+ tooling verworven.

Deze tooling voldoet niet alleen aan de eisen van de applicatie DIM en de IV-organisatie DIM/DWH, maar biedt ook voldoende brede functionaliteit om inzetbaar te zijn als UWV-brede doeltechnologie voor ETL (inclusief ETL-gerelateerd metadata management), maskering en datakwaliteitsbeheer.

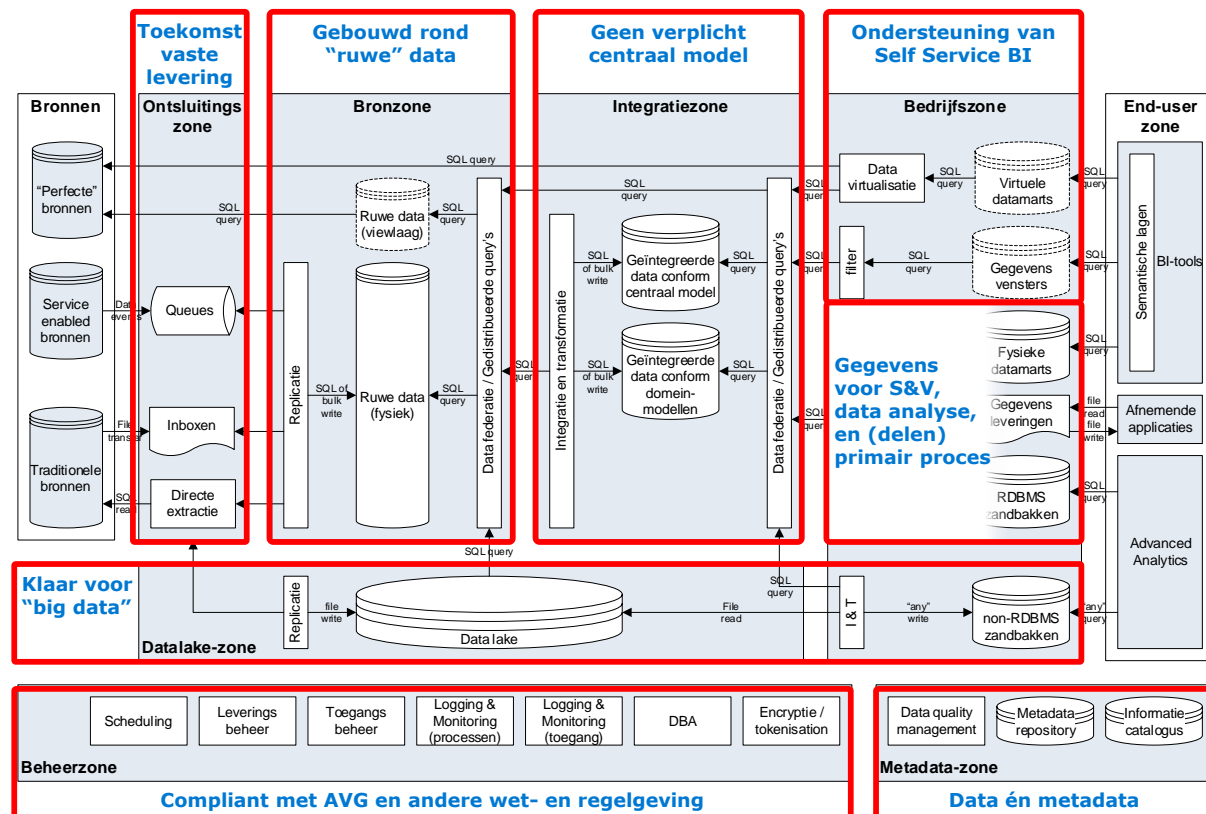
Deze UWV-brede inzet is niet in scope van het project DataFabriek, maar is wel meegenomen in de aanbestedingscontracten; bredere inzet binnen rapportage- en analyseketens is dus meestal² rechtmatig.

Data Integratie Magazijn (DIM)

Het Data Integratie Magazijn vervangt de drie bestaande centrale data warehouses.

Het is in essentie een "klassiek" enterprise datawarehouse (EDW), uitgebreid met een aantal "moderne" concepten die met name bedoeld zijn om de applicatie wendbaarder te maken (en zo het verkleinen van de time-to-market van nieuwe/gewijzigde informatievoorzieningen te ondersteunen), en om deze AVG-compliant te maken (door pseudonimisering/anonimisering en door gegevensminimalisatie). In een later stadium zal het DIM ook "big data" kunnen verwerken.³

Het DIM is dus meer dan een "DWH 4.0":



² Valideer vooraf bij Leveranciersmanagement.

N.B. Ook bij rechtmatigheid kan er impact zijn op de licentiekosten.

³ Inrichting van de daarvoor benodigde DIM-onderdelen valt buiten scope van het project Datafabriek.

Rol van het DIM binnen UWV

Het Data Integratie Magazijn is:

- De verplichte leverancier (voor heel UWV) van gegevens t.b.v. verslaglegging waarvoor UWV-brede consistentie vereist is;
- De verplichte leverancier (voor heel UWV) van divisie-overstijgende analysegegevens;
- De voorkeursleverancier van geïntegreerde/afgeleide gegevens (in bulk) met lage actualiteitseisen;
- De tweede voorkeursleverancier⁴ van ruwe brongegevens (in bulk) met lage actualiteitseisen;
- Géén leverancier van "strikt vertrouwelijke" gegevens (met name medische en strafrechtelijke gegevens), tenzij er geen andere optie is (maar zie gerelateerd beslispunt in paragraaf 1.4);
- Géén leverancier van gegevens met hoge actualiteitseisen ("near realtime"). Daar ligt het primaat bij service-oplossingen zoals die binnen het KIA/KOA-concept;
- Een mogelijke leverancier van alle overige gegevens.
Er is geen beperking m.b.t. het type gegeven (anders dan de vertrouwelijkheid) of het type afnemend proces (anders dan indirect, via de actualiteitseisen)

N.B. Bovenstaande is gebaseerd op de Doelarchitectuur Gegevensintegratie (DA GEIN). Een verdere verbijzondering van de rol van het DIM binnen de architectuur van UWV zal worden vastgelegd in de nieuwe Enterprise Architectuur GegevensHuishouding (EA GH) en de Doelarchitectuur Gegevensdiensten (DA GD).

Impact drivers op de oplossingsarchitecturen

Het **garanderen van de stabiliteit & continuïteit van de bestaande dienstverlening** wordt op technisch gebied bereikt door het inzetten van de in fase 2 van het project verworven ETL+ tooling, inclusief de daarbij behorende support. Ook draagt het project zorg voor de kennisborging m.b.t. het gebruik van deze tools.

Daarnaast draagt het lijn-implementatietraject "GD Uitvoering en IV", middels verdere professionalisering van de DWH- en GLV-medewerkers, zorg voor de stabiliteit & continuïteit van de organisatie(s) die verantwoordelijk zijn (of worden) voor de run en change van het DIM.

Aan het verkleinen van de **time-to-market** zal vooral worden bijgedragen door in het eigen IV-domein barrières weg te nemen en geen overbodige functionaliteit te vereisen.

In de IV-architectuur gaat het dan bijvoorbeeld om het "kortste keten" principe, om zone-ontkoppeling, en om metadata-gedreven (en daarmee automatisch genereerbare) ETL. In de product-architectuur om de ondersteuning van halffabricaten (m.n. selfservice BI), en in de proces-architectuur om gestroomlijnde en, waar mogelijk, geautomatiseerde processen o.b.v. eenduidige afbakening van verantwoordelijkheden.

Verdere verkleining van de time-to-market vereist ook een verbetering van de "business"-aansturing (bijvoorbeeld door de inrichting van een DEC-equivalent⁵). Dit valt echter buiten scope van het project.

N.B. Procesinrichting is een belangrijke factor bij het verkleinen van de time-to-market. Deze inrichting is onderdeel van het lijn-implementatietraject "GD Uitvoering en IV". Het project Datafabriek speelt hierin slechts een ondersteunende rol.

Voor **AVG-compliance** worden zowel technische oplossingen (waaronder gebruik van standaard-maskeringsfunctionaliteit en gestandaardiseerd/geautomatiseerd data lifecycle management (DLM) van zowel gemaskeerde als ongemaskeerde gegevens) als proces-gerelateerde oplossingen

⁴ De eerste voorkeur blijft de bronapplicatie zelf (als die daartoe in staat is)

⁵ Dit "DEC" is beschreven in de adviesnota van Otten uit 2015 en de daarop voortbordurende DA GEIN.

(waaronder rigide controles op rechtsgrond/proportionaliteit/subsidiariteit en “gemaskeerde gegevenslevering tenzij”) ingezet.⁶

Het DIM (en daarmee het project Datafabriek) levert “enablers” voor AVG-compliance; volledige compliance wordt pas bereikt als de hele keten compliant is:

- DLM kan binnen het DIM alleen correct worden uitgevoerd als dat DLM ook bij de gegevensbronnen correct is ingericht;
- de controles op rechtsgrond/proportionaliteit/subsidiariteit moeten worden ingericht door het lijn-implementatietraject “GD Uitvoering en IV”;
- het DIM kan wel gemaskeerde gegevens leveren, maar de beslissing om deze ook werkelijk te gebruiken wordt genomen in de leveringsprocessen (ingericht door het lijn-implementatietraject “GD Uitvoering en IV”). Daarnaast moet ook de afnemer (vaak een decentrale analyseomgeving) qua inrichting en processen met gemaskeerde gegevens kunnen werken.

Ook **compliance met de Archiefwet** (verwijdering na verstrijken bewaartermijnen) wordt middels gestandaardiseerd/geautomatiseerd data lifecycle management (DLM) ondersteund.

N.B. Eén van de BIT-adviezen die ten grondslag liggen aan de nieuwe projectaanpak in 2021 is om na het uitzetten van DWH 3.0 te beoordelen of er nog een zakelijke rechtvaardiging bestaat voor vervanging van het veel omvangrijkere DWH 2.0. Mocht UWV, in de toekomst, inderdaad besluiten om DWH 2.0 niet naar het DIM te migreren, dan worden zowel de vereenvoudiging van het landschap als het dichten van de AVG-compliance-gaps niet, of maar zeer beperkt, verwezenlijkt.

Het **meten en verbeteren van de datakwaliteit** zal worden gebaseerd op functionaliteit van de in fase 2 van het project verworven ETL+ tooling, met, procesmatig, een duidelijke scheiding tussen de verantwoordelijkheden van de centrale organisaties (IV-organisatie DIM/DWH en “DEC-equivalent”) en/of haar gegevens-afnemers (meten en rapporteren) en die van de bron (verbeteren).

N.B. Het inrichten van de hiervoor benodigde afspraken en processen valt buiten het project, en zal elders worden belegd.

Het **verlagen van de TCO** wordt m.n. bereikt door het standaardiseren en, waar mogelijk, automatiseren van de ontwikkel- en beheerprocessen. Daarnaast zal, binnen de mogelijkheden van het nieuwe rekencentrum, gebruik gemaakt gaan worden van “hot en cold storage”, m.a.w. het opslaan van veel gebruikte gegevens op snelle/dure, en van minder gebruikte gegevens op tragere/goedkopere opslagmedia. Grootste impact op de TCO hebben de eerder genoemde verkleining van de time-to-market (en daarmee sneller leveren bedrijfswaarde) en de vereenvoudiging van het gebruik van door het DIM geleverde gegevens en informatie door goede bereikbaarheid van kwalitatief aanzienlijk verbeterde metadata.

N.B. Procesinrichting is een belangrijke factor bij het verlagen van de TCO. Deze inrichting is onderdeel van het lijn-implementatietraject “GD Uitvoering en IV”. Het project Datafabriek draagt hieraan wel bij door het leveren van “enablers”, zoals geautomatiseerde processen, rigide interfaces-standaarden en herbruikbare software-bouwstenen.

Self service gegevensgebruik zal m.n. worden ondersteund door het kunnen leveren van ruwe (al dan niet gemaskeerde) brongegevens. Gegevens kunnen hierbij zowel vanuit het DIM naar de afnemer “gepusht” worden (via gegevensleveringen) als door de afnemer zelf uit het DIM worden opgehaald (via gegevensvensters).

N.B. Self service gegevensgebruik zal m.n. relevant zijn voor decentrale analyseomgevingen. Deze zijn echter vaak (nog) niet AVG-compliant. Compliant maken van deze omgevingen (en hun processen) valt buiten scope van het project Datafabriek, en ook buiten scope van het lijn-implementatietraject “GD Uitvoering en IV”.

⁶ De term “maskeren” wordt in deze PSA gebruikt als verzamelterm voor pseudonimiseren en anonimiseren.

Voor **het ondersteunen van "big data"-achtige functionaliteit** zoals ondersteuning van "advanced analytics" en opslag/verwerking van ongestructureerde gegevens is in de architectuur voorlopig slechts een placeholder (de datalake-zone) opgenomen. Dit omdat de te implementeren oplossing sterk afhankelijk is van beslissingen buiten het project. Wel dekt het programma van eisen en wensen voor de in fase 2 van het project verworven ETL+ tooling ook functionaliteit die nodig is om het (meta)data management van een dergelijk datalake te kunnen opnemen in dat van het "klassieke" deel van het DIM.

LEESWIJZER

Voor de volgende doelgroepen zijn minimaal de volgende onderdelen relevant

Doelgroep	Minimaal relevante onderdelen
Stuurgroep	Management samenvatting, hoofdstuk 1
Overige management stakeholders	Management samenvatting, potentieel hoofdstuk 3 en 4.1
Projectmanagers	Management samenvatting, hoofdstuk 1
Projectleiders	Hoofdstukken 3, 4, 5 en 6
Bedrijfsarchitecten	Gehele document
Gegevensarchitecten	Gehele document, focus op hoofdstuk 1, 4, en 6.
Enterprise architecten	Gehele document, focus op hoofdstuk 1, 2, 4, 6.
IT architecten	Gehele document
Business consultants	Hoofdstukken 1, 3, 4 en 6
Business analisten	Hoofdstukken 1, 3, 4 en 6
Procesontwerpers en -beheerders	Hoofdstukken 1, 3, 4 en 6
Informatie analisten	Hoofdstukken 1, 3, 4 en 6
Functioneel ontwerpers	Hoofdstukken 1, 3, 4 en 6
Functioneel beheerders	Hoofdstukken 1, 3, 4, 5 en 6
Service managers	Hoofdstukken 1, 3, 4 en 6
Software engineers	Hoofdstukken 1, 4, 5 en 6
Implementatie-team	Gehele document, focus op hoofdstuk 1, 3, 4, en 6.

In zijn gewijzigde scope implementeert het project Datafabriek een IV-oplossing (DIM; het Data Integratie Magazijn), en worden de bijbehorende processen en -organisatieonderdelen door de lijn (buiten het project) ingericht. In dit document wordt naar het project verwezen als "project Datafabriek", naar de toekomstige IV-lijnorganisatie als "IV-organisatie DIM/DWH", naar de toekomstige business-aansturing als "Vraag- en ketensturing GD t.b.v. DIM/DWH", en naar het organisatie-inrichtingstraject als "lijn-implementatietraject "GD Uitvoering en IV"".

Deze versie van de PSA heeft als scope het volledige project Datafabriek, maar bevat slechts het detail dat benodigd is tot en met uitfasering DWH 3.0, en/of nu al beschikbaar is.

De wijzigingen in de projectaanpak vanaf 1/1/2021 (o.a. n.a.v. de adviezen van het BIT) zijn in deze versie verwerkt. Deze gewijzigde projectaanpak heeft overigens geen impact op de DIM/Datafabriek-architectuur an sich. Wel wordt deze architectuur nu deels door het project, en deels door de lijn verwezenlijkt. De exacte taakverdeling tussen project en lijn wordt vastgelegd in (deel)projectplannen. Deze zijn hierin leidend; deze PSA beschrijft de taakverdeling slechts in globale termen.

N.B. De Functionaris Gegevensbescherming heeft recentelijk negatief geadviseerd over de tweede versie van de GEB Datafabriek. Aannee in deze PSA-versie is dat dit niet zal resulteren in fundamentele wijzigingen in het DIM-ontwerp.

Elke volgende projectfase (of equivalent, bijvoorbeeld een groep samenhangende project-incrementen) zal vergezeld gaan van een nieuwe versie van deze PSA, met additioneel detail.

Wijzigingen t.o.v. PSA 2.0:

- Impact scopewijzigingen in 2020/2021 expliciet gemaakt*
- Kaders, open punten en gerelateerde projecten/activiteiten bijgewerkt tot status per einde fase 4*
- Verduidelijkende tekst toegevoegd (vaak als voetnoten)*
- Beperkte cosmetische wijzigingen*

1 INLEIDING

1.1 Doel project

1.1.1 Oorspronkelijk doel

(tekst vanwege traceerbaarheid ongewijzigd overgenomen uit vorige versie van de PSA, zie volgende paragraaf voor wijzigingen)

De doelstellingen van het project DataFabriek zijn op hoofdlijnen als volgt samen te vatten:

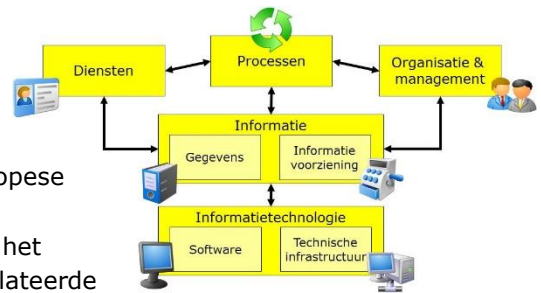
- Technologie:
 - Het verwerven van datawarehouse tooling waarmee het nieuwe datawarehouse "Data Integratie Magazijn" (DIM)" kan worden gerealiseerd (afgerond tijdens fase 2 van het project);
 - Het inrichten van dat datawarehouse; conform de doelarchitectuur DA GEIN en de verbijzonderingen daarvan in het Visie document DIA domein (fase 3 en verder);
 - Het incrementeel bouwen van een datawarehouse en business intelligence functie ten behoeve van de afnemers van de lijn-Datafabriek (fase 4 en verder).
- Processen:
 - Het inrichten van de ontwikkelstraat, inclusief processen, binnen de lijnorganisatie Datafabriek (de zogenaamde OTAP-omgevingen) ten behoeve van de realisatie van informatieproducten binnen de context van een datawarehouse (fase 3 en verder, voorwerk in fase 2);
 - Het realiseren en implementeren van werkprocessen en procedures ten behoeve van de governance van de lijnorganisatie Datafabriek (fase 3 en verder, voorwerk in fase 2).
- Medewerkers:
 - Het opleiden van (huidige en te werven) medewerkers in de in fase 2 van het project verworven datawarehouse tooling (fase 3 en verder);
 - Het opleiden van (huidige en te werven) medewerkers in de nieuwe werk- en governance processen van de lijnorganisatie Datafabriek (alle projectfasen);
 - Het voeren van regie op het werven en scholen van nieuwe medewerkers ten behoeve van de lijnorganisatie Datafabriek (alle projectfasen);
 - Het voeren van een actief verandermanagement programma gericht op het veranderen van technologie, wijze van werken en cultuur (alle projectfasen).

Doorontwikkeling van DIM en lijnorganisatie Datafabriek vindt plaats op basis van leerervaringen uit de diverse migratieprojecten. Deze doorontwikkeling is een verantwoordelijkheid van de lijnorganisatie en zal daar ook belegd worden.

Bovenstaande wordt gedreven door een aantal bedrijfsbelangen:

Hoge prioriteit	Minder hoge prioriteit (maar wel noodzakelijk)
<ul style="list-style-type: none">• Garanderen stabiliteit & continuïteit van de bestaande dienstverlening• Verkleinen time-to-market• Verbeteren gegevensbescherming	<ul style="list-style-type: none">• Meten / verbeteren datakwaliteit• Verlagen TCO• Ondersteunen nieuwe typen ("self service") gegevensgebruik• Ondersteunen nieuwe typen data (b.v. ongestructureerd)

De doelen van het project vallen in alle architectuurdomeinen:



- **Informatietechnologie:**
 - Verwerving nieuwe "ETL+" tooling via een Europese Aanbesteding
 - Installatie en inrichting van deze tooling t.b.v. het Data Integratie Magazijn (DIM), het DIA⁷-gerelateerde deel van de Gegevensintegratielaag (GIL), zoals beschreven in de DA GEIN.
- **Gegevens & Informatievoorziening**
 - Vervanging van de huidige "legacy" data warehouses van Gegevensdiensten door het DIM, met nieuwe onderliggende technologie (omwille van continuïteit en stabiliteit), een verbeterde time-to-market, en een opzet waarmee voldaan wordt aan de (nieuwe) wet- en regelgeving op het gebied van gegevensbescherming.
 - Aansluiting van dit DIM op nieuwe bron-interfaces, met rigider standards op het gebied van formaat en metadata, en zonder afnemer-specifieke filters en afleidingen. Ook hier zijn time-to-market en compliance de drivers.
 - Verbeteren metadata-management-oplossingen om bovenstaande te ondersteunen.
- **Diensten, Processen en Organisatie**
 - Stroomlijning⁸ van de bijbehorende IV-dienstverlening en -organisatie tot een "lijnorganisatie Datafabriek" qua organisatie, processen, besturing, en bemensing. E.e.a. binnen de kaders van de IV-professionalisering en de invlechting van de afdeling DWH in Gegevensdiensten.

1.1.2 Impact van scope changes en voortgang elders op behalen projectdoelen

Het oorspronkelijke projectdoel was:

"te komen tot een goed werkende afdeling gebaseerd op basis van de nieuwe datawarehouse tooling."

In het projectplan voor 2021 is dit iets scherper verwoord:

"Het project Datafabriek heeft tot doel de datawarehouse voorziening van UWV toekomstvast te maken zodanig dat zowel interne als externe ketenpartners op een volwassen datagedreven wijze kunnen bijdragen aan de missie en visie van UWV."

- Met deze scherpere verwoording is ook de taakverdeling tussen project en lijn helderder vastgelegd: Ontwerp/realisatie van de voorbrengingsprocessen en governance vallen nu, voor het deel dat niet technologie- of DIM-specifiek is, onder lijn-implementatietraject "GD Uitvoering en IV", dat verbeteringen binnen de afdeling DWH (w.o. ook die ten behoeve van het DIM) doorvoert, en daarnaast zorg draagt voor de invlechting van de afdeling DWH (nu nog ICTS) in de divisie Gegevensdiensten (inclusief de daarvoor benodigde wijzigingen in proces en organisatie van de rest van Gegevensdiensten). Binnen dit lijn-initiatief is géén sprake meer van een "lijnorganisatie Datafabriek", maar van een "IV-organisatie DIM/DWH", aangezien alle vereiste verbeteringen in IV-proces en -organisatie worden geïmplementeerd als aanpassingen op de bestaande afdeling DWH. Het project levert nog wel input voor ontwerp/realisatie van DIM-gerelateerde processen binnen de afdeling DWH en elders binnen Gegevensdiensten (zoals doorgevoerd in het

⁷ Data Integratie en Analyse, zoals beschreven in "Visie document en plaat DIA domein, versie 0.3 mei 2018"

⁸ Voorbeelden zijn: ondersteuning van self service BI, beter/robuuster specificeren én beheren van levering vanuit de bronnen, voorraadbeheer, betere requirements-afstemming, optimaal gebruik van metadata.

"lijn-implementatietraject "GD Uitvoering en IV""), maar heeft in dezen geen sturende rol meer.

Daarnaast blijkt er op de een aantal voor de Datafabriek belangrijke dossiers de afgelopen jaren minder voortgang te zijn geboekt dan bij de start van het project verwacht:

- **Data Expertise Centrum (DEC)**

Inrichting van het (DEC) was nog wel in scope van Datafabriek-voorloper DEC/DF, maar is bij de start van het project Datafabriek buiten scope geplaatst, en onderdeel geworden van de IV-transitie.

Er is in de afgelopen jaren echter weinig tot geen voortgang op dit gebied geweest, en de term "DEC" is zelfs volledig uit het gezicht verdwenen.

Inrichting van de "vraag- en ketensturing GD t.b.v. DIM/DWH" valt nu, informeel, onder het lijn-implementatietraject "GD Uitvoering en IV". Daar zijn nu wel de eerste stappen gezet.

- **Data Lifecycle Management (DLM)**

UWV heeft de afgelopen jaren vrijwel geen voortgang geboekt bij het inrichten van DLM in de primaire systemen. Dit betekent dat UWV qua data retentie niet compliant is met Archiefwet of AVG. Begin 2021 wordt wel een "vooronderzoek DLM" gestart.

Het DIM toont de administratieve werkelijkheid van UWV, en is daarmee "in commissie incompliant" als de DIM-bronnen hun DLM niet op orde hebben.

- **AVG-compliance analyseomgevingen**

In de DIAscoop-projecten BPDA2 en Maskeren DMAP zijn hier stappen gezet, maar die beperkten zich (a) tot de centrale omgevingen en stelden (b) meer definitieve oplossingen uit tot de door het project Datafabriek gerealiseerde DIM-functionaliteit daaraan kon bijdragen.

Voor decentrale analyseomgevingen zijn nog weinig zichtbare stappen richting AVG-compliance gezet.

- **Inrichten selfservice organisatie**

Onderdeel van het programma DIAscoop, maar nog geen tastbare resultaten.

- **Metadata Management**

Onderdeel van DIAscoop, niet verder gekomen dan een vooronderzoek.

- **BI Portfolio**

Onderdeel van DIAscoop, niet verder gekomen dan een (niet door de IV-board geaccepteerd) voorstel om de gedoogtermijn voor de drie binnen UWV het meest gebruikte BI-tools (BusinessObjects, PowerBI, Qlikview) met 2 jaar te verlengen.

Er blijven dus binnen UWV 6 BI-tools en 24 analyse-tools in huis gebruikt worden, en de status van die tools (contained, doeltechnologie, wel/niet rechtmatig, etc) blijft onduidelijk.

- **Organisatie-breed gebruik/beheer PowerBI**

Het gebruik van PowerBI binnen UWV is "organisch gegroeid". Dit gebruik is nu, zie hierboven, in ieder geval tijdelijk gesanctioneerd.

Gecentraliseerd beheer voor PowerBI (door Gegevensdiensten) staat op de rol voor 2021.

Dit vereist ook aanpassingen aan de inrichting van en de security rondom PowerBI. Wat die aanpassingen precies behelzen is nog niet vastgelegd.

- **Enterprise Architectuur**

Het project Datafabriek implementeert een gemeenschappelijke, divisie-overstijgende oplossing. Dit vereist ook inbedding in de enterprise architectuur.

Bij de start van het project heeft de lead architect Gegevensdiensten hiervoor, pragmatisch, de DIAvisie geschreven, als verbijzondering van de DA GEIN.

De UWV-architectuurkolom werkt op dit moment aan een Enterprise Architectuur Gegevenshuishouding (EA GH) en een Doelarchitectuur Gegevensdiensten (DA GD), die de

DA GEIN zullen vervangen.

Beiden zijn echter nog in een zeer conceptueel stadium.

Bovenstaande scope-changes en voortgangsisues betekenen dat het behalen van de projectdoelen vaak alleen in samenwerking met andere projecten en/of lijninitiatieven mogelijk is, of pas op langere termijn (na afloop van het project) mogelijk is:

- Het **garanderen van de stabiliteit & continuïteit van de bestaande dienstverlening** valt op technisch gebied onder het project (door het inzetten van de in fase 2 van het project verworven ETL+ tooling, inclusief de daarbij behorende support). Het lijn-implementatietraject "GD Uitvoering en IV" draagt, middels verdere professionalisering van de DWH- en GLV-medewerkers, zorg voor de stabiliteit & continuïteit van de organisatie(s) die verantwoordelijk zijn (of worden) voor de run en change van het DIM.

De voor stabiele en compliant verwerking van gegevens benodigde metadata wordt door (of op bestelling) van het project vergaard, en vastgelegd in interface-specifieke documentatie (RLO's). Actualiteitseisen m.b.t. die RLO's zijn in de levercontracten (GIA's) vastgelegd.

Eenmalig vergaren van metadata t.b.v. het robuust en compliant verwerken van brongegevens (zoals vastgelegd in de RLO's), valt binnen scope van het project Datafabriek, maar metadata **management** (bij de bronnen) t.b.v. het actueel houden van die RLO's valt buiten scope van het project. Aanname is dat het toekomstige beleid m.b.t. metadata management ook de vereisten voor RLO's zal dekken.

N.B. Continuïteit van de, op DIM-data gebaseerde, BI-producten kan slechts voor de verlengde gedoogtermijn van BusinessObjects/PowerBI worden gegarandeerd.

Dit betekent dus ook dat net door het project (of de afnemers) gebouwde BI-producten potentieel moeten worden herbouwd als de gebruikte tool niet meer gedoogd en/of om andere redenen uitgefaseerd wordt.

- Het project draagt bij aan het verkleinen van de **time-to-market** door het DIM "wendbaar" te ontwerpen. Voorbeelden hiervan zijn het "kortste keten" principe, de zone-ontkoppeling, en de metadata-gedreven (en daarmee automatisch genereerbare) ETL. Procesinrichting is een belangrijke factor bij het verkleinen van de time-to-market. Deze inrichting is onderdeel van het lijn-implementatietraject "GD Uitvoering en IV", en het project Datafabriek speelt hierin slechts een ondersteunende rol. Verdere verkleining van de time-to-market vereist ook een verbetering van de "business"-aansturing (bijvoorbeeld door de inrichting van vraag- en ketensturing⁹). Dit valt echter buiten scope van het project.¹⁰

- Voor **AVG-compliance** implementeert het project Datafabriek de technische oplossingen, waaronder gebruik van standaard-maskeringsfunctionaliteit en gestandaardiseerd/geautomatiseerd data lifecycle management (DLM) van zowel gemaskeerde als ongemaskeerde gegevens.

Deze technische oplossingen zijn de "enablers" voor werkelijke AVG-compliance:

- DLM kan binnen het DIM alleen correct worden uitgevoerd als dat DLM ook bij de gegevensbronnen correct is ingericht;
- de controles op rechtsgrond/proportionaliteit/subsidiariteit moeten worden ingericht door het lijn-implementatietraject "GD Uitvoering en IV"
- het DIM levert wel gemaskeerde gegevens, maar de beslissing om deze ook werkelijk te gebruiken wordt genomen in de leveringsprocessen (ingericht door het lijn-implementatietraject "GD Uitvoering en IV").

Daarnaast moet ook de afnemer (vaak een decentrale analyseomgeving) met

⁹ Deze vraag- en ketensturing vielen oorspronkelijk onder het "DEC"... E.e.a. is beschreven in de adviesnota van Otten uit 2015 en de daarop voortbordurende DA GEIN.

¹⁰ Zie ook volgende paragraaf

gemaskeerde gegevens kunnen werken. Totdat zulks het geval is zal het DIM óf niet kunnen leveren (dit vertraagt de uitfasering van legacy DWH's), óf tijdelijk toch ongemaskeerd blijven leveren (met als gevolg IT-schuld en langer voortbestaan van AVG-incompliance).

Ook **compliance met de Archiefwet** (verwijdering na verstrijken bewaartermijnen) wordt middels gestandaardiseerd/geautomatiseerd data lifecycle management (DLM) ondersteund, en ook hier kan DLM binnen het DIM alleen correct worden uitgevoerd als dat DLM ook bij de gegevensbronnen correct is ingericht.

N.B. Voor alle "netjes" op het DIM aangesloten afnemers¹¹ geldt wel dat zij bij later inrichten van het DLM (door bron) automatisch ook hun DLM op orde krijgen; hun DLM volgt immers dat van het DIM, en daarmee indirect dat van de bron.

- Het **meten en verbeteren van de datakwaliteit** zal worden gebaseerd op functionaliteit van de in fase 2 van het project verworven ETL+ tooling.
Het inrichten van de hiervoor benodigde afspraken en processen valt onder het lijn-implementatietraject "GD Uitvoering en IV". Onderdeel hiervan is het inrichten van een duidelijke scheiding tussen de verantwoordelijkheden voor meten en rapporteren, respectievelijk verbeteren; het eerste valt deels onder verantwoordelijkheid van de centrale organisaties (IV-organisatie DIM/DWH en vraag- en ketensturing GD t.b.v. DIM/DWH) en/of haar gegevens-afnemers, het tweede (verbeteren), blijft altijd de verantwoordelijkheid van de bron.
- Het **verlagen van de TCO** wordt m.n. bereikt door het standaardiseren en, waar mogelijk, automatiseren van de ontwikkel- en beheerprocessen. Daarnaast zal, binnen de mogelijkheden van het nieuwe rekencentrum, gebruik gemaakt gaan worden van "hot en cold storage", m.a.w. het opslaan van veel gebruikte gegevens op snelle/dure, en van minder gebruikte gegevens op tragere/goedkopere opslagmedia. Grootste impact op de TCO hebben de eerder genoemde verkleining van de time-to-market (en daarmee sneller leveren bedrijfswaarde) en de vereenvoudiging van het gebruik van door het DIM geleverde gegevens en informatie door goede bereikbaarheid van kwalitatief aanzienlijk verbeterde metadata.

N.B. Procesinrichting is de bepalende factor bij het verlagen van de TCO. Deze inrichting is onderdeel van het lijn-implementatietraject "GD Uitvoering en IV". Het project Datafabriek draagt hieraan wel bij door het leveren van "enablers", zoals rigidere interfaces-standaarden en herbruikbare software-bouwstenen.

- **Self service gegevensgebruik** zal m.n. worden ondersteund door het kunnen leveren van ruwe (al dan niet gemaskeerde) brongegevens. Gegevens kunnen hierbij zowel vanuit het DIM naar de afnemer "gepusht" worden (via gegevensleveringen) als door de afnemer zelf uit het DIM worden opgehaald (via gegevensvensters).
Gegevensbeveiliging zal zo worden ingericht dat deze, binnen de beperkingen van de binnen UWV gebruikte BI-tools, zo veel mogelijk "bij de data" ligt. Hierdoor kunnen autorisatiebeslissingen en -oplossingen voor de verschillende BI-tools (in de context van het project Datafabriek m.n. BusinessObjects en PowerBI)

N.B. Self service gegevensgebruik zal m.n. relevant zijn voor decentrale analyseomgevingen.
Deze zijn echter vaak (nog) niet AVG-compliant. Compliant maken van deze omgevingen valt buiten scope van het project Datafabriek, en ook buiten scope van het lijn-implementatietraject "GD Uitvoering en IV".
Ook is er nog weinig voortgang bij het opzetten van self service (onderdeel

¹¹ Door virtueel gebruik te maken van DIM-data, door DIM-data "flush-and replace" te gebruiken, of door zelf ook weer een DLM-proces te hebben ingericht dat het DLM van het DIM volgt (net als het DIM dat doet t.o.v. de bronnen). Vastleggen van deze DLM-volgingseis dient onderdeel te zijn van de GLA.

DIAscoop).

).

Daarnaast zijn de security-eisen die volgen uit organisatie-breed gebruik/beheer van PowerBI nog onduidelijk.

- Voor **het ondersteunen van "big data"-achtige functionaliteit** zoals ondersteuning van "advanced analytics" en opslag/verwerking van ongestructureerde gegevens is in de architectuur voorlopig slechts een placeholder (de datalake-zone) opgenomen. Dit omdat de te implementeren oplossing sterk afhankelijk is van beslissingen buiten het project. Wel dekt het programma van eisen en wensen voor de in fase 2 van het project verworven ETL+ tooling ook functionaliteit die nodig is om het (meta)data management van een dergelijk datalake te kunnen opnemen in dat van het "klassieke" deel van het DIM.

Voor de volledigheid:

Temporiseren van het project Datafabriek om zo beter te kunnen "meeliften" op UWV-initiatieven elders is geen optie, aangezien dit zou resulteren in het langer blijven bestaan van de huidige continuïteits- en compliance-issues.

1.2 Architectuurdrijfveren

Ook de architectuurdrijfveren van het project vallen in alle architectuurdomeinen:

- **Informatietechnologie:**
 - De verworven nieuwe "ETL+" tooling is door UWV bestempeld als doeltechnologie.
 - De inrichting van het DIM is "cloud ready"¹².
 - De voor het DIM opgestelde¹³ interface-standaarden voor gegevensaanlevering vanuit de bronsystemen zijn herbruikbaar voor andere data-integratie-oplossingen.
- **Gegevens & Informatievoorziening**
 - Het DIM vervangt drie bestaande data warehouses, en leidt daarmee tot vereenvoudiging van het applicatielandschap.
 - Het DIM dicht een aantal AVG-compliance-gaps bij zowel de huidige DWH-omgevingen als de afnemers daarvan.
- **Diensten, Processen en Organisatie**
 - De stroomlijning van de bijbehorende IV-dienstverlening en -organisatie leidt zowel tot duidelijker gedefinieerde en meer "agile" processen, als tot een duidelijker afbakening van de verantwoordelijkheden van de IV-organisatie DIM/DWH, haar leveranciers en haar afnemers.

N.B. De activiteiten in dit domein zullen nog maar zeer gedeeltelijk door het project geleverd en bestuurd kunnen worden.

Verwachting is dat dit zich zal beperken tot het leveren van globale proces-requirements aan het lijninitiatief "Implementatie bij GD Uitvoering en IV", en het monitoren van de voortgang, bij dat initiatief, van de activiteiten die noodzakelijk zijn om de door het project opgeleverde IV-producten in beheer te nemen.

*Dit betekent dat de bovengenoemde stroomlijning (zowel qua processen als qua organisatieafbakening) **niet** meer gezien kan worden als een project-doel; deze PSA beschrijft dit domein dus ook slechts in zeer algemene termen. Het*

¹² De "cloud" in deze context is de private cloud in het nieuwe HRC (bij DXC).

¹³ In fase 3 is versie 1.0 van deze standaarden geaccordeerd. Deze versie is de basis voor de interface-aanpassingen in fase 4 en 5. Geleerde lessen en/of gevraagde verduidelijkingen zullen worden verwerkt in bijgewerkte versies van die standaarden.

totaaloverzicht zal opgenomen moeten worden in de Doelarchitectuur Gegevensdiensten.

1.3 Belangrijkste aandachtspunten

Aandachtspunten van het project vallen in alle architectuurdomeinen:

- **Informatietechnologie:**
 - Het DIM mag geen end-of-life software-componenten of -pakketten bevatten.
 - Het DIM moet qua inrichting "cloud ready" zijn, maar zal in eerste instantie geïmplementeerd worden in het bestaande rekencentrum. Inrichting moet daarom zodanig zijn dat migratie van het huidige naar het nieuwe rekencentrum slechts beperkte inspanning vergt.
 - Database-structuren dienen zodanig te zijn dat AVG-compliant toegangsbeheer optimaal wordt ondersteund.
 - De maskeringsfunctionaliteit dient zodanig te zijn ingericht dat deze voldoet aan de UWV-richtlijnen op dit gebied¹⁴. Deels zijn deze richtlijnen nog niet volledig uitgekristalliseerd.
 - De gegevens- en metadata-management-componenten dienen zodanig te zijn ingericht dat deze, op de langere termijn, ook kunnen worden gebruikt bij het beheer van datavirtualisatie- en/of "big data"-omgevingen.
 - De interfaces vanuit de bronsystemen dienen voldoende metadata en gegevensredundantie (checksums, kruistellingen) te bevatten om integriteit, volledigheid en volgordeelijkheid automatisch te kunnen controleren.
- **Gegevens & Informatievoorziening**
 - Het DIM dient AVG-compliant te zijn qua toegang en gebruik.
 - De agility van gegevens- en informatielevering via het DIM dient te worden gemaximaliseerd door:
 - brede bron-interfaces¹⁵ ("voorraadbeheer")¹⁶
 - standaard interface-formaten (m.n. t.a.v. metadata), en daardoor de mogelijkheid om de inleesfunctionaliteit op basis van metadata te genereren¹⁷
 - qua datamodel ontkoppelde DIM-zones (m.b.v. data-federatie technologie)
 - ondersteuning self service BI & analyse (gegevensvensters en zandbakken)
 - toepassing "kortste keten" principe (niet meer stappen/tussenresultaten binnen het DIM dan noodzakelijk)
 - Het DIM volgt, voor ongemaskeerde brondata, het data lifecycle management (DLM) van die bronnen, en kan dus qua minimale en maximale bewaartermijnen slechts compliant zijn met AVG en andere wet- en regelgeving als de bronnen dat ook zijn. Zolang dat niet zo is zal het DIM "in commissie" ook een compliance-gap hebben. Voor gemaskeerde (gepseudonimiseerde of geanonimiseerde) data, en voor anderszins binnen het DIM afgeleide data voert het DIM het DLM zelf uit. Voor

¹⁴ Zoals beschreven in: "Richtlijn testdata anonimiseren", "Beleid Privacy by Design en by Default", "Duiding GM-beleid bij gegevensgebruik zandbakken" en "Bewaartermijnen in de datawarehouse Omgevingen"

¹⁵ Meer details in Bijlage A : Rationale brede bronontsluiting

¹⁶ In de context van het project DataFabriek beperkt dit voorraadbeheer zich tot reeds op de uit te faseren "legacy" DWH's aangesloten bronnen. Op de langere termijn zal dit waarschijnlijk gelden voor alle bronnen met een divisie-overstijgend belang. De beschrijving van de langere termijn is in scope van de Doelarchitectuur Gegevenshuishouding.

¹⁷ Interface-formaten blijven i.h.a. ongewijzigd, aangezien formaattransformaties eenvoudiger in het DIM (m.b.v. ETL-tooling) kunnen worden geïmplementeerd dan bij de bron.

ongemaskeerde data wordt hierbij het DLM van de onderliggende brondata als voornaamste sturing gebruikt, bij gemaskeerde data de specifieke DLM-standaards voor analyse-data. Deze zijn op hun beurt overigens weer deels gebaseerd op het DLM van de onderliggende brondata.

- De specificaties van de door de bronsystemen geleverde gegevens dienen voldoende gedetailleerd te zijn (zowel op functioneel als op technisch niveau) om zowel eenduidige als AVG-compliant gegevens- en informatie-leveringen via het DIM te ondersteunen. Naast gegevensdefinities (functioneel en technisch) zijn vertrouwelijkheidsclassificaties hiervoor cruciaal. De vereiste detaillering zal gewaarborgd worden binnen de bestaande standaards voor FUGEM's en FO's, met DIM-specifieke extensies hierop (als onderdeel van RLO's en GLO's) waar deze standaards de behoeften van het DIM (nog) niet dekken.

N.B. Deze informatie is afkomstig van de bronnen (en deels beschikbaar in FUGEM's en CGM), maar zal binnen de metadata-componenten van het DIM worden gestructureerd in automatisch verwerkbaar vorm.¹⁸

- **Diensten, Processen en Organisatie**

- De IV-organisatie DIM/DWH en het "DEC Equivalent" dienen AVG-compliant¹⁹ te handelen qua gegevens- en informatielevering, en ontwikkel- en beheerprocessen.
- De ontwikkel- en beheerprocessen t.b.v. het DIM, zoals die binnen de IV-organisatie DIM/DWH worden uitgevoerd, dienen gericht te zijn op maximale agility, zonder daarbij onnodige continuïteitsrisico's te creëren.

1.4 Beslis- en discussiepunten

Aandachtspunten van het project vallen in alle architectuurdomeinen:

- **Informatietechnologie:**

- De "to be" infrastructuur-componenten, en de vrijheidsgraden bij het inrichten ervan, zijn afhankelijk van de mogelijkheden binnen het nieuwe DXC-rekencentrum en -contract.

Eventuele hieruit voortvloeiende inrichtingsbeperkingen kunnen gedeeltelijke herinrichting van de voor het DIM benodigde ETL- en database-tooling vereisen.
Status: Loopt (buiten het project). DIM-vereisten m.b.t. infrastructuur zijn tijdens het aanbestedingsproces meegenomen in de validatie van de aanbesteders door vertegenwoordigers van Gegevensdiensten en DWH, en zijn ook ingebracht bij het transitieteam IBM-DXC.

Verder overleg hierover valt buiten de scope van het project, aangezien (het beheer van) de DIM-infrastructuur al is overgedragen aan de lijn (platform-team beheer binnen de IV-organisatie DIM/DWH).

- De "to be" inrichting van het technisch beheer (TAB) van het DIM, met daarin een belangrijke rol voor de IV-organisatie DIM/DWH (zie 4.2.3 Afwijkingen van architectuurprincipes), is door ICT Services opnieuw ter discussie gesteld.
Status: Loopt (buiten het project); Architecten Datafabriek, DWH en ICT Services hebben in een gezamenlijk stuk vastgelegd dat de eisen die de IV-organisatie DIM/DWH aan de TAB-processen moet stellen om haar diensten te kunnen blijven

¹⁸ Als bij de gap-analyse (door project DataFabriek) t.b.v. aansluiting van een bron op het DIM blijkt dat er geen volledig bijgewerkte FUGEM beschikbaar is, dan zal die bron gevraagd worden om zulks, conform vigerend beleid, zo snel mogelijk op te pakken, zodat de interface-metadata gebaseerd kan worden op een correcte FUGEM. Het Data Office zal, parallel daaraan, gevraagd worden om, waar al bekend is dat een DIM-bron nog bijwerking van een FUGEM behoeft, dit verbeterproces zo snel mogelijk op te (laten) starten.

¹⁹ En uiteraard ook compliant met andere wet- en regelgeving

leveren niet binnen de standaardoplossing (of zelfs de standaarddefinitie) van TAB DXC passen.

Verder overleg over de exacte inrichting van het DIM-TAB valt buiten de scope van het project, aangezien (het beheer van) de DIM-infrastructuur al is overgedragen aan de lijn lijn (platform-team beheer binnen de IV-organisatie DIM/DWH).

- **Gegevens & Informatievoorziening**

- Zeer gevoelige gegevens (bv. medische gegevens, (tussen)resultaten handhavings-processen) horen in principe niet thuis in het DIM; het is dan beter om analyse en/of rapportage hierover binnen het brondomein van de betreffende gegevens uit te voeren, en alleen de eventueel daarvoor benodigde contextuele gegevens uit het DIM te betrekken. Het kan echter zijn dat de baten van het gebruik maken van de beter ingerichte processen/tools van de IV-organisatie DIM/DWH en het DIM opwegen tegen de risico's van het opnemen van zeer gevoelige gegevens in (een extra beveiligd deel van) het DIM.
Status: Open. Vereist overleg met DIM-afnemers én -leveranciers, en met CISO en FG. De ervaring leert dat er nog veel niet geregeld en/of vastgelegd is m.b.t. het opslaan van zeer gevoelige gegevens in de DWH's (w.o. ook het DIM), en m.b.t. de doorlevering ervan als input voor analyses/rapportages. Regelen/vastleggen valt buiten scope van het project (verantwoordelijkheid voor IV-organisatie DIM/DWH en de vraag- en ketensturing GD t.b.v. DWH), maar vertraging aldaar kan wel leiden tot IT-schuld bij het project.
Opstarten afstemming voor regelen/vastleggen is belegd bij lijn- implementatietraject GD Uitvoering en IV.
- De eventuele rol van het DIA-domein (en daarmee het DIM) bij het leveren van (bulk)informatie aan operationele processen is nog niet uitgekristalliseerd. Er lopen hiervoor op het moment een aantal trajecten, zoals het opstellen van een Enterprise Architectuur GegevensHuishouding (EA GH).
In deze trajecten gemaakte keuzes kunnen m.n. impact hebben op de nu nog door DWH 3.0 aan MijnGegevensdiensten geleverde gegevens t.b.v. de SUAG.
Status: Gesloten; het DIM zal, conform de DIAvisie, ook (bulk)informatie aan operationele processen gaan leveren. .
- UWV heeft nog geen organisatie-breed beleid m.b.t. centrale en decentrale data warehouses en rapportage/analyse-omgevingen. Er lopen hiervoor op het moment een aantal trajecten (opstellen EA GH, bijvoorbeeld). In deze trajecten gemaakte keuzes kunnen m.n. impact hebben op de gegevensvoorziening (door het DIM) aan DMAP.
Status: Lopend traject. De Divisiearchitect Gegevensdiensten is gevraagd hier, als onderdeel van de EA GH, een beslissing over te (laten) nemen.
De eerste versie(s) van deze EA GH zijn echter vooral conceptueel, en eenduidig(en gedragen) beleid t.a.v. (de)centrale rapportage/analyse-omgevingen wordt op korte termijn niet verwacht.

- UWV heeft nog geen organisatie-breed beleid m.b.t. maskeren van persoonsgegevens ten behoeve van Sturen en Verantwoorden en/of advanced analytics.
Status: Lopend traject. In overleg met CISO, Bestuurszaken en de BSO Gegevensdiensten heeft het project een eerste versie van de maskeringsaanpak gemaakt (vergelijkbaar met de "beslismatrix" van TSC²⁰) en die aan hen ter review voorgelegd. Deze beslismatrix zal, na afronding, worden overgedragen aan het lijn-implementatietraject "GD Uitvoering en IV", zodat het daar kan worden meegenomen in de, nog door hen in te richten, vraag- en ketensturing GD t.b.v. DWH.
Leveringen vanuit het DIM zijn "by default" gemaskeerd. Uitzonderingen hierop (dat kunnen breed toegepaste uitzonderingen zijn) worden getoetst bij de BSO Gegevensdiensten.
- UWV heeft nog geen organisatie-breed beleid m.b.t. AVG-compliance van (decentrale) rapportage- en analyse omgevingen.
Status: Open.
Leveringen vanuit het DIM zijn "by default" gemaskeerd, maar veel "self service" omgevingen en processen zijn (nog) niet ingericht op gebruik van gemaskeerde gegevens. Waar dit de voortgang van het project (en daarmee het wegwerken van continuïteitsrisico's) belemmert zullen de betreffende omgevingen, na toetsing bij BSO/FG, ongemaskeerd belevend blijven worden. Aan de DIM-zijde is de resulterende IT-schuld beperkt; informatieproducten (en eventuele onderliggende informatiegebieden) zijn relatief eenvoudig van ongemaskeerd naar gemaskeerd "om te katten". IT-schuld aan de afnemers-zijde valt buiten verantwoordelijkheid van het project of, na overdracht, van de IV-organisatie DIM/DWH.
- UWV heeft nog geen organisatie-breed beleid m.b.t. de opslag van (al dan niet gemaskeerde) persoonsgegevens ten behoeve van Sturen en Verantwoorden en/of advanced analytics.
Status: Lopend traject. Project en JZ stellen, in overleg, een nieuwe versie op van het stuk "DIM – Grondslagen verwerking persoonsgegevens", waarin het voortschrijdend inzicht van JZ en FG is verwerkt. Dit wordt dan de basis voor het wel of niet mogen opslaan van gegevens in het DIM.
Aan Gegevensdiensten IM is gevraagd dit aspect ookmee te nemen in de "Kaders en handvatten DLM". Tot die tijd neemt het project aan dat toekomstig UWV-breed beleid consistent zal zijn met de input van JZ en FG.
- De toekomstige rol(len) van DMAP in de UWV-architectuur zijn nog onbepaald, en daarmee dus ook de toekomstige rechtsgrond van de gegevensleveringen (uit het DIM) aan DMAP, en de daaraan gerelateerde maskeringseisen.
Status: Lopend traject, deels afhankelijk van voorgaande punt. De Divisiearchitect Gegevensdiensten is gevraagd hier, als onderdeel van de EA GH, een beslissing over te (laten) nemen.
De eerste versie(s) van deze EA GH zijn echter vooral conceptueel, en eenduidig(en gedragen) beleid t.a.v. (de)centrale rapportage/analyse-omgevingen wordt op korte termijn niet verwacht.

²⁰ Formele titel: "Concretisering UWV Beleid Persoonsgegevens buiten productie"

- De legacy DWH's maken gebruik van door de afdeling DWH beheerde reference data. Eigenaarschap en afspraken rondom deze reference data zijn echter niet vastgelegd. Een duidelijke "soll" voor deze ongewenste situatie is er nog niet; het bestaande UWV-beleid m.b.t. reference data biedt hiervoor nog te weinig handvatten. Tegelijkertijd is deze reference data integraal onderdeel van de door het DIM te leveren informatieproducten, en is opname van data in het DIM dus randvoorwaardelijk voor de uitfasering van DWH 3.0 (en 2.0. en 1.0).
Status: Project voert verder onderzoek uit.
O.b.v. de uitkomsten daarvan zullen project Datafabriekricht, samen met het lijn-implementatietraject "GD Uitvoering en IV" en/of de afdeling DWH tijdelijke oplossingen inrichten om deze reference data te beheren.
Het technische deel van deze oplossingen garandeert traceerbaarheid van de reference data in het DIM, het organisatorische deel de inrichting van de (deels tijdelijke) beheer-processen (door de IV-organisatie DIM/DWH en/of de vraag- en ketensturing GD t.b.v. DIM/DWH).
Naast bovenstaande worden de interface-standaarden (bron-DIM) aangepast: een brede bronontsluiting dient ook de in de bron beschikbare reference data te bevatten.

• Diensten, Processen en Organisatie

De lijnorganisatie Datafabriek (nu: IV-organisatie DIM/DWH) is een gemeenschappelijke functie, en het DIM een gemeenschappelijke voorziening. Divisie-overstijgende vraagsturing en keten-governance (bijvoorbeeld door een Data Expertise Centrum²¹) is hiervoor noodzakelijk. Inrichting van deze vraagsturing/governance is echter **geen** onderdeel van de project-scope, maar was onderdeel van het gestopte programma IV-transitie.

Status: Loopt. Eisen aan een DEC (of equivalent) worden, door het project Datafabriek samen met het lijn-implementatietraject "GD Uitvoering en IV", verzameld en als input aan de Domeinhouder DWH opgeleverd. Informeel heeft het lijn-implementatietraject "GD Uitvoering en IV" de inrichting van een "DEC equivalent" als "in scope" geaccepteerd.

Opbouw projectstartarchitectuur

De projectstartarchitectuur is opgebouwd uit een aantal hoofdstukken. Allereerst wordt een hoofdstuk gewijd aan een overzicht welke relevante, generieke architectuurproducten geldend zijn binnen dit project.

Direct daarna volgt een hoofdstuk waarin beknopt de essentie van de verandering wordt geschetst. Vervolgens zijn de binnen UWV onderscheiden hoofd-architectuurdomeinen als aparte hoofdstukken opgenomen. Per hoofd-architectuurdomein wordt vervolgens per sub-architectuurdomein de afbakening van het verandergebied in de eenheden van het betreffende sub-architectuurdomein weergegeven.

Daarna volgt een vertaling van de algemene principes van het betreffende sub-architectuurdomein naar de project specifieke betekenis en de eventueel geconstateerde afwijkingen op de algemene architectuurprincipes.

Tot slot wordt per sub-architectuurdomein de hoofdstructuur van de oplossing beschreven.

²¹ Zoals beschreven in de "Adviesnota Data Expertise Centrum en Data Fabriek"

Tabel: motivatie ontbrekende sub-domeinen

Architectuurdomein	Motivatie voor het niet beschrijven in deze PSA	Ontheffing geaccordeerd door
Procesarchitectuur	<p>De procesarchitectuur is geen deliverable van het project Datafabriek meer, maar onderdeel geworden van het lijn-implementatietraject "GD Uitvoering en IV"..</p> <p>Deze PSA beperkt zich dus tot de globale eisen waaraan de door dat implementatietraject in te richten processen moeten voldoen.</p>	Lead architect Gegevensdiensten
Organisatie- en Management architectuur	<p>De O&M-architectuur is geen deliverable van het project Datafabriek meer, maar onderdeel geworden van het lijn-implementatietraject "GD Uitvoering en IV".</p> <p>Deze PSA beperkt zich dus tot de globale eisen waaraan de door dat implementatietraject in te richten organisatieonderdelen moeten voldoen.</p>	Lead architect Gegevensdiensten

2 KADERS

2.1 Uitgangsdocumenten

Voor deze PSA zijn de volgende referentiekaders en documenten als basis of vertrekpunt gebruikt:

Visienota's en andere strategische documenten:

- UWV Informatieplan 2018-2022 Puzzelen met prioriteiten, CIO Office, april 2018.

UWV Architectuurkaders:

- Sjabloon Projectstartarchitectuur, CIO Office, april 2016.
- Handleiding Projectstartarchitectuur, CIO Office, september 2015.
- Doelarchitectuur E-Dienstverlening, K&S, 22 november 2017, versie 1.50.
- Doelarchitectuur E-Werken, 7 januari 2016, versie 1.9
- Doelarchitectuur Gegevensintegratie, 30 april 2016, versie 1.00
- UWV Dienstenportfolio, Financieel Economische Zaken, ??? 201?.
- Bedrijfsfunctiemodel UWV, CIO Office / Martijn Bleeker, 3 februari 2020, versie 1.0
- Gegevensprincipes UWV, Data Office, februari 2013, versie 1.1.
- Omgaan met gegevensredundantie, Data Office, februari 2008, versie 1.0.
- Richtinggevende informatievoorzieningsprincipes 2016, CIO Office, april 2016.
- IT Beleid & Strategie 2010, CIO Office, juli 2010, versie 1.0.
- Richtlijn UWV Logging en Monitoring, 22 augustus 2016, versie 4.0
- IB&P richtlijn OTAP, 9 november 2016, versie 1.0
- Normenkader data omgevingen, 12 december 2017, versie 1.0
- Het dienstenmodel UWV, kaderstelling 2015, FEZ/PC&A, versie 1.4, 20 juli 2015
- BIV classificatie, v1.0, 18-04-2016
- UWV IB&P principes PSA, 28-04-2014
- Richtlijn encryptie, v1.0, 28-12-2015
- Richtlijn synchrone bronontsluiting, 8-05-2017
- UWV ICT Richtlijn Data Anonimisering en Pseudonimisering Versie 1.1 24 juli 2018
- Concretisering UWV Beleid Persoonsgegevens buiten productie, v0.6 22 juli 2015
- Architectuur Zakelijk portaallandschap MijnGegevensdiensten, v2.0, 23-08-2017
- Beleid Privacy by Design en by Default, CISO, Versie 0.81 Concept
- Duiding GM-beleid bij gegevensgebruik zandbakken, december 2016
- Duiding GM beleid bronontsluiting, januari 2018
- Beleidslijn VIP gegevens, november 2009
- Bewaartermijnen in de datawarehouse Omgevingen, v2.6, concept
- Data Lifecycle Management binnen UWV, versie 1.2
- Notitie Bewaarregels en Bewaartermijnen selectielijst UWV, 30-6-2017

Toetsingen en audits:

- Adviesnota Data Expertise Centrum en Data Fabriek, Gerard Otten, versie 1.00, 28-04-2015
- Toetsing van Architectuur Data Fabriek en Vernieuwingslijn UWV, DeltIQ, 20 december 2016
- Toetsing architectuur datafabriek en vernieuwingslijn, Capgemini, februari 2017
- Een doelbindingstoets op DWH en PAO, Noordbeek, concept versie: 0.81

2.2 Aanvullende architectuurkaders

Bestaande aanvullende architectuurkaders:

- Visie document en plaat DIA domein, versie 0.3 mei 2018
- Notitie BI-tooling dossier, 14-10-2016
- Infographic UWV doelarchitectuur
- Doorkijk DF GIL, v 0.43

2.3 Project-specifieke uitgangspunten

- Programma van Eisen en Wensen v.1.0
(Bijlage F bij UTI Openbare Europese aanbesteding ETL+ en dienstverlening 1.0)
- DIM – Grondslagen verwerking persoonsgegevens, v1.1 29 december 2020
- Interface-standaarden (bron-DIM), v1.5c 20 oktober 2020
- Hergebruik DIM-interfaces in legacy DWH's (v1.0) 21 november 2019
- UWV HLD DIM, v1.0 (final) 11 Maart 2020

2.4 Voor deze PSA-versie gebruikte (deel)projectplannen

Per 1/1/2021 (oorspronkelijk de start van fase 5) wijzigt de projectopzet van het project Datafabriek. Dit n.a.v. de adviezen van het BIT.

Conceptversies (deel)projectplannen:

- Projectplan Datafabriek na ombuiging BIT (Concept Versie 0.92, 08-02-2021)
- deelprojectplan Datafabriek Bronontsluitingen (Versie 0.7, 02-12-2020)
- Deelprojectplan Datafabriek Afnemers 1 (Versie 0.7, 02-12-2020)
- Deelprojectplan Datafabriek Afnemers 2 (Versie 0.7, 02-12-2020)
- Deelprojectplan Datafabriek SOM (Versie 0.4, 02-12-2020)
- Trajectplan Datafabriek Implementatie bij GD Uitvoering en IV (Versie 0.3.2, 01-12-2020)

Hierbij is het projectplan als leidend genomen.

Daarnaast:

- Jaarplan 2021 Gegevensdiensten (versie 2.0, 10 december 2020)

Reeds geïdentificeerde scope-issues:

- Reeds geïdentificeerde producten+aandachtspunten voor faseplan fase 5
(intern projectdocument, 20-10-2020)

3 BEDRIJFSARCHITECTUUR

Dit hoofdstuk voor de volledigheid opgenomen in deze versie van de PSA; inhoud zal uiteindelijk onderdeel worden van één van de domeinarchitecturen van Gegevensdiensten.

3.1 Diensten

Deze paragraaf heeft als input gediend voor de Producten en Diensten Catalogus (PDC) van de IV-organisatie DIM/DWH. Deze PDC is nu leidend.

3.1.1 Afbakening Diensten

De aan het DIM gerelateerde beheer- en ontwikkeltaken zullen worden ingevuld door de bestaande afdeling DWH.

De dienstverlening van de huidige afdeling DWH bestaat uit:

- Gegevensleveringen van gegevens in ruwe vorm (brongegevens) aan interne en externe partijen voor verschillende gebruiksdoeleinden.
- Gegevensleveringen van gegevens in geaggregeerde vorm aan interne en externe partijen voor verschillende gebruiksdoeleinden.
- Het leveren van expertisediensten aan interne afnemers.
- Het leveren van beheerdiensten op gegevenssets en verwerkingsmiddelen aan interne afnemers van gegevensleveringen.
- Het leveren van compensatiediensten t.b.v. bronsystemen met tekortkomingen in historie opbouw en/of gegevensvoorziening ten behoeve van (bulk) rapportage en analyse.

Het leveren van expertisediensten, beheerdiensten en compensatiediensten geschiedt uitsluitend aan interne afnemers bij UWV. Ze komen derhalve niet voor in de UWV dienstenportfolio. Gegevensleveringen worden zowel gedaan aan interne als aan externe partijen, zij het dat een levering aan een externe partij altijd via een interne partij loopt, en dus, vanuit DIM-perspectief, als "interne levering" kan worden bestuurd.

Dienst	Omschrijving	Effect
Gegevenslevering	Gegevensleveringen van gegevens in ruwe vorm (brongegevens) en geaggregeerde vorm aan interne en (indirect) externe partijen voor verschillende gebruiksdoeleinden	Volume bij afdeling DWH blijft vergelijkbaar. Wel is er een verschuiving richting leveringen in ruwe vorm, vanwege toename "Self Service" BI. Daarnaast zal een deel van gegevens niet fysiek geleverd worden (als bestand), maar virtueel, door het via een "gegevensvenster" toegankelijk maken van de data in het DIM.
Expertisedienst	Het leveren van expertisediensten, op het gebied van (meta)data, datakwaliteit, rapportage en tooling, aan interne afnemers.	Volume bij afdeling DWH neemt deels toe, aangezien metadata- en datakwaliteitsdiensten nu nog geen onderdeel zijn van de diensten-catalogus. Rapportagediensten zullen, vanwege toename "Self Service" BI, in omvang afnemen.

Dienst	Omschrijving	Effect
Beheerdienst	Het leveren van beheerdiensten op gegevenssets en verwerkingsmiddelen aan interne afnemers van gegevensleveringen	Volume bij afdeling DWH blijft vergelijkbaar.
Compensatiediensten	Het leveren van compensatiediensten t.b.v. bronsystemen met tekortkomingen in historie opbouw of beschikbaarheid ²² .	Volume bij afdeling DWH blijft vergelijkbaar.

3.1.2 Architectuurprincipes Diensten²³

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
DA/OP-P01 De ontwikkeling van een dienst vereist minimaal één combinatie van doelgroep en dienstverlening, af te leiden uit de wet dan wel uit de in de wet gelaten beleidsruimte	Gegevensleveringen vanuit het DIM zijn AVG-compliant en geschieden uitsluitend na beoordeling op rechtsgrond, proportionaliteit en subsidiariteit. Bij deze beoordeling zijn de vertrouwelijkheidsclassificaties van zowel de volledige aan de afnemer ter beschikking gestelde dataset als van de onderliggende gegevenselementen een belangrijke input.
DA/OP-P03 Kleine verschillen in dienstverlening kan UWV opnemen in één dienst	In de dienstverlening van de IV-organisatie DIM/DWH wordt gestreefd naar zo veel mogelijk meervoudig toepasbare producten en diensten. Derhalve zullen combinaties van vraagsignalen worden beoordeeld op een mogelijkheid tot generieke en/of gemeenschappelijke oplossing. N.B. Dit IV-streven dient gebaseerd te zijn op een vergelijkbaar streven binnen de, buiten het project in te richten, business-vraagsturing van de IV-organisatie DIM/DWH.

²² Het betreft hier beschikbaarheid voor bulk-bevraging. De lijnorganisatie Datafabriek en de applicatie DIM compenseren niet voor systeembeschikbaarheid in het algemeen ("uptime").

²³ Bron: 'Het dienstenmodel UWV, kaderstelling 2015, FEZ/PC&A, versie 1.4, 20 juli 2015

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
DA/OP-P04 UWV verleent diensten om een door de wetgever beoogd maatschappelijk effect te realiseren c.q. daar aan bij te dragen	De dienstverlening van de IV-organisatie DIM/DWH (via interne partijen) aan externe partijen levert altijd een bijdrage aan één van de volgende doelen: <ul style="list-style-type: none"> • administratieve lastenverlichting of procesoptimalisatie bij de afnemers; • inzicht te verschaffen in bepaalde maatschappelijke ontwikkelingen in het aandachtsgebied Werk en Inkomen. • Verantwoording aan ministeriële opdrachtgevers.
DA/OP-P05 UWV verleent een dienst ná een in de wet gefundeerde externe aanleiding	Voor wat betreft gegevenslevering aan externe partijen (via interne partijen) zie DA/OP-P01 Voor wat betreft overige leveringen van producten/diensten is dit principe niet van toepassing.
DA/OP-P06 UWV ontwerpt een dienst als het eindresultaat van een bedrijfsproces (c.q. ketenproces)	De gegevensleveringen zijn op te vatten als het resultaat van het bedrijfsproces 'Uitvoeren gegevensleveringen' uit het bedrijfsprocesmodel van divisie Gegevensdiensten. De levering van overige producten/diensten worden ontworpen als resultaat van een werkproces. Daarmee zijn ze een interne leverantie in de zin van de UPMC standaard.
DA/OP-P07 Een (deel) dienst heeft een cliënt, een ketenpartner of een derde partij als externe afnemer	Zie DA/OP-P06
DA/OP-P08 Naast de aanvrager(s) kunnen er meer belanghebbende afnemers zijn	Zie DA/OP-P03
DA/OP-P09 Een dienst is herkenbaar voor externe afnemers en opdrachtgever (lees: wetgever). Interne leveranties zijn herkenbaar voor interne afnemers/opdrachtgevers	De afdeling DWH ontwikkelt een voor alle potentiële afnemers eenvoudig raadpleegbare producten/diensten catalogus. Deze dekt ook de m.b.v. het DIM geleverde producten/diensten.

3.1.3 Afwijkingen van architectuurprincipes

Geen

3.1.4 Hoofdstructuur van de oplossing

In de Doelarchitectuur Gegevensintegratie v1.00 (DA GEIN) is een beeld geschetst van een lijnorganisatie Datafabriek²⁴ met een gegevensintegratielaag als stip op de horizon.

Het project Datafabriek zet een eerste belangrijke stap in die richting met de introductie van een ETL+ platform. Op basis van de ETL+ functionaliteit is er een producten/diensten-catalogus ontworpen die gerealiseerd kan worden als de beoogde processen/diensten van de lijnorganisatie

²⁴ Niet meer onder herkenbaar, aangezien de taken van deze beoogde lijnorganisatie onderdeel zijn geworden van het takenpakket van de afdeling DWH.

Datafabriek, zoals ingericht binnen de afdeling DWH als “IV-organisatie DIM/DWH”, operationeel en volwassen zijn. Deze processen/diensten zien er sterk samengevat als volgt uit:

Vier dienst/productcategorieën:

- Gegevensleveringen;
- Compensatiediensten;
- Beheer- & inrichtingsdiensten;
- Expertisediensten

Vaak is de IV-organisatie DIM/DWH binnen UWV de “preferred supplier”, maar zelden de “verplichte leverancier”. De precieze afspraken op dit gebied zijn afhankelijk van de organisatorische positionering van de afdeling DWH (en de IV-organisatie DIM/DWH daarbinnen) t.o.v. de rest van Gegevensdiensten (w.o. ook de vraag- en ketensturing GD t.b.v. DIM/DWH) en de overige divisies van UWV. Deze positionering valt buiten scope van het project Datafabriek, en wordt in deze PSA dan ook niet verder beschreven.

Productgroep Gegevensleveringen

- Datamarts
Op maat gemaakte gegevensverzamelingen, vaak dimensioneel gestructureerd, ten behoeve van specifieke gebruikersgroepen en/of -processen (meestal rapportages, OLAP, dashboards, of vergelijkbaar), inclusief self service BI. Vaak vereist het op maat maken complexe transformaties en afleidingen.
- Gegevensvensters
Virtuele gegevensverzamelingen (deelverzamelingen van de in het DIM beschikbare gegevens) die ontsloten worden ten behoeve van self-service BI en/of self service analytics, d.w.z. rechtstreeks gebruik door afnemers, bijvoorbeeld t.b.v. querying en ad-hoc rapportage. Anders dan bij data marts is er geen sprake van complexe transformaties en afleidingen. Wel bevatten gegevensvensters filters op tabel-, rij- en attribuutniveau. Dit om te garanderen dat de getoonde deelverzameling alleen gegevens bevat waarvoor rechtsgrond, proportionaliteit en subsidiariteit geldt.
- Gegevensbestanden
Levering van in meer of mindere mate bewerkte gegevens, doorgaans in de vorm van platte bestanden. Voorbeelden zijn leveringen aan pensioenfondsen en aan analyse-omgevingen.

Dienstgroep Beheer & Inrichting²⁵

- Inrichten gebruiksomgeving
Het initieel inrichten van een omgeving waarin een gegevensverzameling beschikbaar wordt gesteld. Bijvoorbeeld het inrichten van een zandbak of een omgeving waarin een datamart geplaatst gaat worden, inclusief de bijbehorende IT middelen en BI- of analyse-tooling
- Beheer op eindgebruiksmiddelen
Het uitvoeren van operationele beheertaken op front end tools voor rapportage en analyse, zoals BusinessObjects.
- Beheer op levering
Het uitvoeren van operationele beheertaken op de inhoud van de geleverde (virtuele) gegevensverzamelingen (datamarts, gegevensvensters, etc.).

²⁵ In de context van het project DataFabriek beperkt de scope van deze diensten zich tot tools/technieken/middelen waar de lijnorganisatie Datafabriek, reeds (middels ervaring met bestaande oplossingen of in het project DataFabriek) expertise over heeft opgebouwd. Uitbreiding van die expertise tot een algemene kenniscentrum is mogelijk, maar valt buiten de scope van het project.

Dienstgroep Compensatie

- Compensatie historie
De opbouw van historie in brongegevens indien een bronsysteem dit zelf niet kan.
- Compensatie beschikbaarheid
Het bieden van beschikbaarheid van brongegevens voor wat betreft bulkbevragingen.²⁶
- Compensatie snelheid
Snellere beschikbaarstelling van nieuwe en/of gewijzigde gegevens- en/of informatieproducten, door het ontkoppelen van de implementatie van deze nieuwe wensen van de releasekalender van de bronsystemen. Ook wel bekend als "voorraadbeheer".

Dienstgroep expertisediensten

- Inzicht kwaliteit
Het bieden van inzicht in de kwaliteit van bronsysteemgegevens en/of afleidingen daarvan.
- Inzicht metagegevens
Het bieden van inzicht in de semantiek en syntax van gegevens die de IV-organisatie DIM/DWH kan leveren, alsmede inzicht in de relaties tussen deze gegevens en de onderliggende brongegevens (horizontale lineage), en tussen functionele gegevensbeschrijvingen en hun technische implementatie (verticale lineage).²⁷
- Rapportagedienst
Het uitvoeren van een rapportage-activiteit uit naam van de afnemer waarbij het resultaat van de rapportage beschikbaar wordt gesteld aan die afnemer.
- Expertisedienst
Het beschikbaarstellen van BI-expertise aan afnemers om hen bij te staan in BI gerelateerde activiteiten.

N.B. Het deelproject SOM definieert potentieel nog extra dienstgroepen, buiten de afdeling DWH (en de IV-organisatie DIM/DWH daarbinnen). Deze diensten zullen óf worden beschreven in een volgende versie van deze PSA, óf in een aparte PSA voor het deelproject SOM.

²⁶ De IV-organisatie DIM/DWH biedt géén compensatie voor de beschikbaarheid van brongegevens in het algemeen ("uptime"). Dit is een verantwoordelijkheid van de bron zelf.

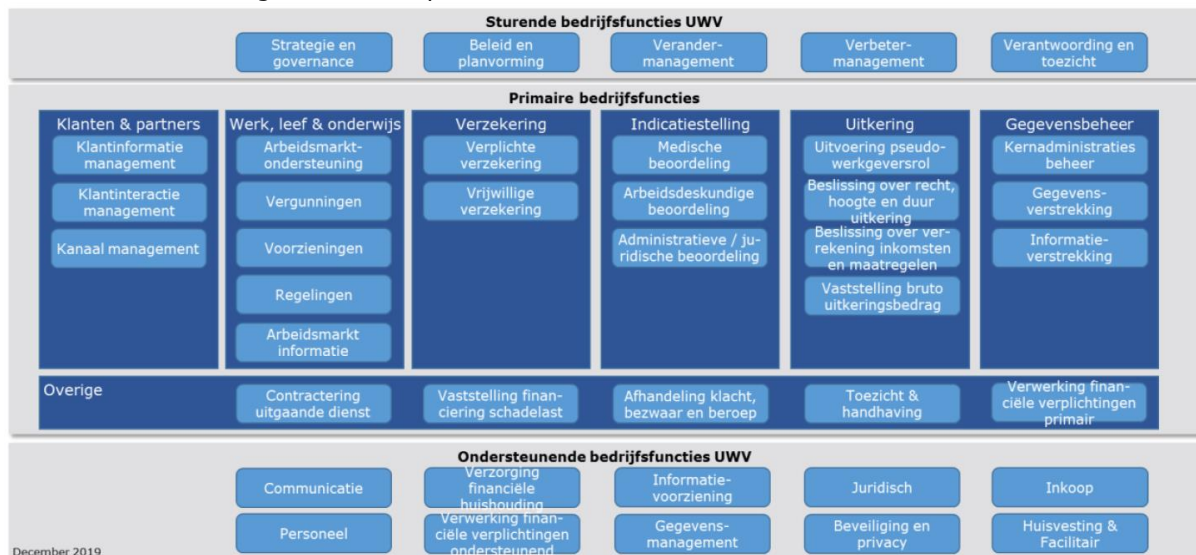
²⁷ Het functionele deel van deze expertise is, voor een groot deel, het "doorleveren" van kennis die reeds in HUDSV, FUGEM's en/of het CGM is vastgelegd.

3.2 Bedrijfsfuncties

De IV-organisatie DIM/DWH heeft enkele kerntaken die toegevoegde waarde leveren voor haar afnemers:

1. Het leveren aan afnemers van gegevens (in ruwe vorm of in enige mate bewerkt)
2. Het leveren aan afnemers van beheerdiensten op geleverde gegevenssets en/of doelomgevingen van deze geleverde gegevenssets
3. Het leveren van expertise op het gebied van Business Intelligence.

Indien we de taken van de IV-organisatie DIM/DWH projecteren op het Bedrijfsfunctiemodel UWV dan levert dit het volgende beeld op:



De IV-organisatie DIM/DWH valt onder de ondersteunende bedrijfsfunctie Informatievoorziening. De door de IV-organisatie DIM/DWH ter beschikking gestelde gegevens en informatie ondersteunen vooral:

- alle sturende bedrijfsfuncties;
- alle primaire functies binnen Gegevensbeheer
- de primaire functies Klantinformatiemanagement en Arbeidsmarkt informatie
- de ondersteunende functies Verzorging financiële huishouding en Gegevensmanagement

In theorie kan de IV-organisatie DIM/DWH overigens ondersteunende (bulk)gegevens leveren aan elke bedrijfsfunctie.

3.3 Processen

3.3.1 Afbakening Processen

(in deze PSA niet beschreven, aangezien procesontwerp buiten scope is geplaatst, en onderdeel wordt van de invlechting van de afdeling DWH in Gegevensdiensten; het lijn-implementatietraject "GD Uitvoering en IV")

3.3.2 Architectuurprincipes Processen

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
GEAUTOMATISEERD UWV voert zijn bedrijfs- en werkprocessen geautomatiseerd uit. Alleen indien strikt noodzakelijk maakt menselijke handeling (ondersteund door ICT) deel uit van het proces.	Het inlezen en verder verwerken van gegevens zal, in het overgrote deel van de gevallen, volledig automatisch gebeuren. Alleen in uitwijksituaties en bij zeer in frequente processen zal er sprake zijn van handmatige start van, verder geautomatiseerde, processen. N.B. Het gebruik van gegevens door afnemers van de DIM, bijvoorbeeld in analyses, zal vaak ad hoc en/of interactief zijn. Menselijke handelingen zijn hier cruciaal.
WERK BIJ DE KLANT UWV legt het werk waar mogelijk neer bij de klant.	Beperkt relevant: De IV-organisatie DIM/DWH levert primair interne diensten, en alleen klantdiensten aan ketenpartners (bv. pensioenfondsen).
STURING OP RESULTAAT UWV richt processen in met sturing op resultaat en een sluitende werkoverdracht.	De grenzen van de verantwoordelijkheid van de IV-organisatie DIM/DWH, en daarmee de aansluiting tussen de processen bij de bronnen en/of afnemers en die bij de IV-organisatie DIM/DWH zullen. door het lijn-implementatietraject GD Uitvoering en IV, eenduidig worden vastgelegd. Dit geldt zowel in de ontwikkelprocessen ("change") als in de reguliere processen ("run"). Sturing op deze laatste processen zal gebaseerd zijn op GegevensInwinningsAfspraken (GIA's) en GegevensLeveringsAfspraken (GLA's), en aansluiten op reeds binnen Gegevensdiensten beschikbare (en binnen het lijn-implementatietraject "GD Uitvoering en IV" verder verbijzonderde) processen en applicaties voor Contract Management.
WERKEN MET ZAKEN Elke hoeveelheid werk die voor de externe klant wordt uitgevoerd, waarvan kwaliteit en doorlooptijd bewaakt moeten worden, wordt vastgelegd en afgehandeld als zaak.	Beperkt relevant: De IV-organisatie DIM/DWH levert primair interne diensten, en alleen klantdiensten aan ketenpartners (bv. pensioenfondsen). Wel blijkt dat beperkt vastgelegde koppelingen (in de bronnen) tussen een zaak en de daarbinnen gebruikte gegevens het DLM van die brondata te compliceren. Dit raakt ook het DLM binnen het DIM, aangezien dat het DLM van de bron volgt.
RECONSTRUCTIE UWV kan de wijze waarop een dienst geleverd is verantwoord en reproduceren	Wordt geïmplementeerd door voor alle gegevens binnen het DIM de administratieve tijdslijn (de momenten dat een gegeven in het DIM werd geladen/gewijzigd/verwijderd) vast te leggen.

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
STANDAARD PROCESVERLOOP UWV kiest voor standaardisatie van het procesverloop van de primaire en secundaire bedrijfsprocessen en de IV-processen.	Het inleesproces van gegevens zal, in het overgrote deel van de gevallen, worden gestandaardiseerd. Dit vereist standaardisering van interface-formaten en –metadata. De overige gegevensverwerkende processen zullen zoveel mogelijk gestandaardiseerd worden, m.n. door het inzetten van standaard ETL+ tooling en door het definiëren, vastleggen en delen van standaards en “best practices” bij het gebruik hiervan.

3.3.3 Afwijkingen van architectuurprincipes

Geen

3.3.4 Hoofdstructuur van de oplossing

(in deze PSA niet beschreven, aangezien procesontwerp buiten scope is geplaatst, en onderdeel wordt van de invlechting van de afdeling DWH in Gegevensdiensten; het lijn-implementatietraject “GD Uitvoering en IV”)

N.B. Procesontwerp van BI-processen en hun besturing is wél in scope van het deelproject SOM, maar is daar een deliverable, die dus nog niet in deze PSA-versie kan worden beschreven. Door SOM opgeleverde procesontwerpen zullen óf worden beschreven in een volgende versie van deze PSA, óf in een aparte PSA voor het deelproject SOM.

3.4 Organisatie & Management

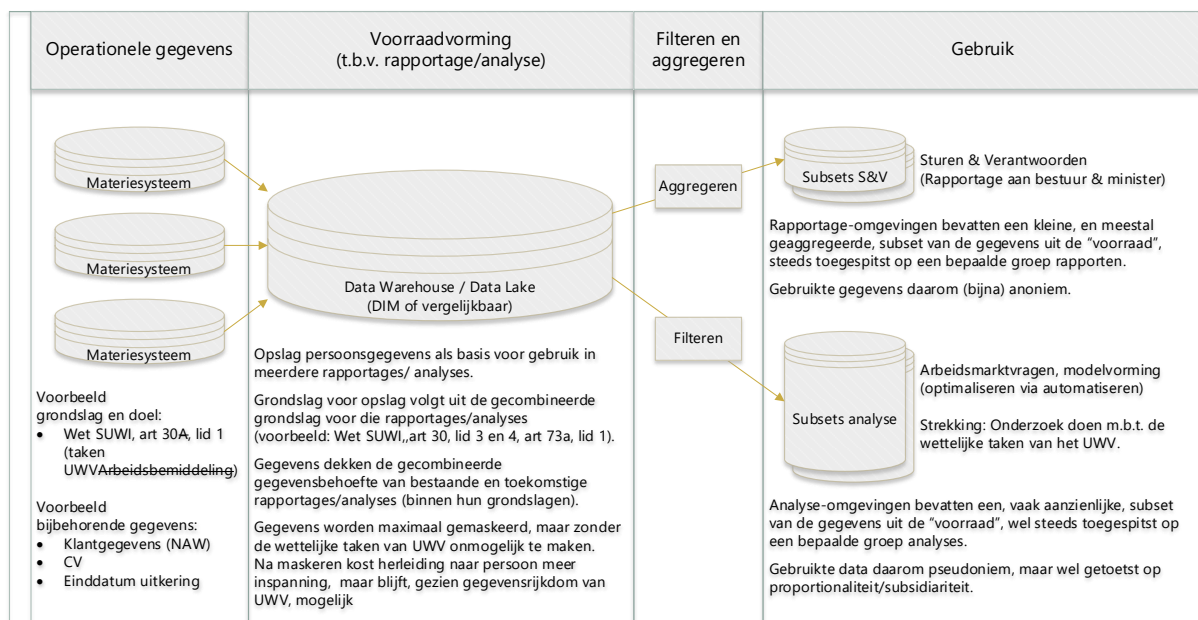
(in deze PSA niet beschreven, aangezien de O&M-architectuur buiten scope is geplaatst, en onderdeel wordt van de invlechting van de afdeling DWH in Gegevensdiensten; het lijn-implementatietraject “GD Uitvoering en IV”)

4 INFORMATIEARCHITECTUUR

4.1 Gegevens

In versie 1.0 van deze PSA werden de termen “anonimiseren” en “pseudonimiseren” gebruikt, conform de “UWV ICT Richtlijn Test Data Anonimisering 1.0”. Deze richtlijn is inmiddels vervangen door de “UWV ICT Richtlijn Data Anonimisering en Pseudonimisering 1.1”, die beter aansluit op de AVG. Daarnaast heeft Juridische Zaken, op verzoek van het project Datafabriek, onderzocht op welke wijze de voornaamste voor UWV geldende wet- en regelgeving (WetSUWI, Archiefwet, AVG) het DIM raken. De weerslag daarvan is beschreven in het document “DIM – Grondslagen verwerking persoonsgegevens, v1.0”²⁸.

In de context van deze PSA is onderstaand plaatje (overgenomen uit het grondslagen-document) m.n. relevant:



De gegevensverwerking in het DIM volgt conform bovenstaande, drie sporen:

- Gegevens die binnen UWV om operationele redenen beschikbaar zijn, worden in ongemaskeerde vorm ook in het DIM vastgelegd. De bewaartermijn van die gerepliceerde gegevens volgt daarbij de bewaartermijn van de originele gegevens. De bron blijft verantwoordelijk voor het uitvoeren van het Data Lifecycle Management (DLM) op die originele gegevens²⁹; het DIM volgt slechts en is daarmee “in commissie niet compliant” als een bron een gegeven te vroeg of te laat verwijdert.
- Gegevens die binnen UWV, na verloop van de operationele bewaartermijn, beschikbaar dienen te blijven t.b.v. rapportage en analyse blijven conform die langere bewaartermijn³⁰ beschikbaar in het DIM, zij het alleen in gemaskeerde vorm. Gezien de gezamenlijke

²⁸ Begin 2021 verschijnt een nieuwe versie van dit document, met voortschrijdend inzicht van JZ erin verwerkt.

²⁹ Conform de hiervoor relevante bewaartermijnen uit de Selectielijst, typisch 5-7 jaar

³⁰ Typisch 20-25 jaar, zie Selectielijst onder “Bewaartermijnen voor ondersteunende processen, punt 15 en 16

detailbehoefte van de achterliggende rapportage- en analyseprocessen blijft deze data (indirect) herleidbaar op personen, en moet dus beschouwd worden als pseudoniem.

- Gegevenssets die vanuit het DIM worden doorgeleverd aan een specifiek rapportage- of analyseproces beperken zich tot het voor dat proces noodzakelijke detail³¹. Deels is dit vereiste detail zodanig dat de data pseudoniem blijft, deels zal aggregatie (of vergelijkbaar) mogelijk zijn, zodat de betreffende gegevensset beschouwd kan worden als anoniem.

4.1.1 Afbakening Gegevens

Het DIM verwerkt in principe alle bedrijfsobjecten van UWV. Daarom wordt hieronder het effect op diverse categorieën van gegevens beschreven.

Gegevenscategorie	Effect
Identificerende gegevens Alle brongegevens die in één of andere vorm identificerend zijn voor een natuurlijk persoon ³² .	De huidige DWH's voeren geen data lifecycle management uit. Het DIM doet dat wel: daar zullen ongemaskeerde persoonsgegevens nooit langer beschikbaar zijn dan in de bron, zodat aan archiverings- en vernietigingsverplichtingen ³³ kan worden voldaan.
Gemaskeerde identificerende gegevens Alle brongegevens die ongemaskeerd in één of andere vorm identificerend zouden zijn voor een natuurlijk persoon, maar nu niet meer (of slechts zeer beperkt). N.B. In een beperkt aantal gevallen is maskering van een identificerend gegeven niet mogelijk, omdat dat de bruikbaarheid van het gegeven voor rapportage/analyse teniet doet. Denk hierbij, bijvoorbeeld, aan de geboortedatum.	De huidige DWH's maskeren ³⁴ niet. Het DIM doet dat wel: alle persoonsgegevens zijn daar initieel zowel in ongemaskeerde als gemaskeerde vorm beschikbaar. Zowel opslag binnen als levering door het DIM (aan haar afnemers) volgen het "privacy by design" principe: <ul style="list-style-type: none"> • Ongemaskeerde persoonsgegevens worden gescheiden opgeslagen van hun gemaskeerde equivalent. • Geleverde data is gemaskeerd, tenzij er rechtsgrond, proportionaliteit en subsidiariteit geldt voor het gebruik van ongemaskeerde data. Zolang persoonsgegevens ongemaskeerd in de bron beschikbaar zijn, bestaat er ook een ongemaskeerde replica van in het DIM. De maskering binnen het DIM moet dan dus worden beschouwd als pseudonimisering . N.B. De (verdere) maskering en/of filtering van de door het DIM aan de afnemers geleverde/ontsloten gegevens volgt rechtsgrond, proportionaliteit en

³¹ E.e.a. conform een door afnemer uit te voeren controleproces m.b.t. rechtsgrond, proportionaliteit en subsidiariteit

³² Het gaat hierbij om primair en secundair identificerende gegevens. Zie "UWV Richtlijn Data Anonimisering en pseudonimisering" voor details.

³³ Zoals onder andere beschreven in "Bewaartermijnen in datawarehouse omgevingen" en "Data Lifecycle Management binnen UWV"

³⁴ De term "maskeren" wordt in deze PSA gebruikt als verzamelterm voor pseudonimiseren en anonimiseren.

Gegevenscategorie	Effect
	<p>subsidiariteit. Hier zal dus wel sprake kunnen zijn van anonimisering.</p> <p>Bij verwijdering van de persoonsgegevens uit de bron verdwijnt ook de ongemaskeerde replica uit het DIM.³⁵ Vanaf dat moment is er binnen het DIM alleen nog een gemaskeerde versie beschikbaar. Gezien de rijkdom aan gegevens binnen UWV, en daarmee binnen het DIM, kan niet worden gegarandeerd dat deze gegevens op geen enkele wijze herleid kunnen worden op personen. Voor de AVG blijven ze dus pseudoniem. Bij gebruik in (minder rijke) deelverzamelingen van de DIM-data, zoals gebruikt bij (en toegankelijk voor) de afnemers van het DIM kunnen ze echter beschouwd worden als "de facto anoniem". Het gegevensgebruik valt daarmee nog steeds onder de AVG, maar de privacy-risico's zijn geminimaliseerd.³⁶</p>
<p>Niet-identificerende gegevens</p> <p>Alle brongegevens die, als de relatie met persoonsgegevens is verwijderd, niet tot een persoon te herleiden zijn.</p>	<p>Bijna elk gegevenselement dat te relateren is aan (niet geanonimiseerde) persoonsgegevens is te classificeren als "gevoelig" of zelfs "zeer gevoelig" (bv. medische gegevens, opsporingsgegevens).</p> <p>De huidige DWH's beveiligen dit slechts door toegangsbeheer.</p> <p>In het DIM:</p> <ul style="list-style-type: none"> • Worden alle niet-identificerende gegevens gescheiden opgeslagen van, al dan niet gemaskeerde, persoonsgegevens. • Kunnen deze gegevens zowel worden gekoppeld aan persoonsgegevens als aan gemaskeerde persoonsgegevens. <p>Dit laatste is de default-optie ("privacy by design"), en bij verwijdering van de ongemaskeerde gegevens (zie hierboven) zelfs de enige optie.</p>
<p>Voorbewerkte gegevens</p> <p>Alle gegevens die zijn afgeleid uit brongegevens om ze eenvoudiger bruikbaar te maken voor groepen gegevens-afnemers en/of -producten. Als die afnemers/producten worden gebruikt voor keten/proces-integratie, dan zijn ze in de regel niet anoniem, als ze worden gebruikt voor analyse en/of rapportage juist wél</p>	<p>De huidige DWH's maken geen onderscheid tussen voorbewerking voor keten-integratie en die voor analyse/rapportage; beiden lopen via hetzelfde model.</p> <p>Het DIM maakt dat onderscheid wel.</p> <p>De strikte scheiding tussen identificerende en niet-identificerende gegevens zoals die bij de brongegevens wordt gehanteerd is hier niet mogelijk. Daarom geldt voor ongemaskeerde voorbewerkte persoonsgegevens dat ze in hun geheel de Data Lifecycle van de brongegevens volgen. Verwijdering uit de bron betekent dus ook verwijdering van de ervan afgeleide persoonsgegevens.</p>

³⁵ De fysieke verwijdering uit het DIM wordt bepaald door het moment waarop de bron die verwijdering doorgeeft, en dat is weer afhankelijk van de functionaliteit die de bron op dit gebied biedt.

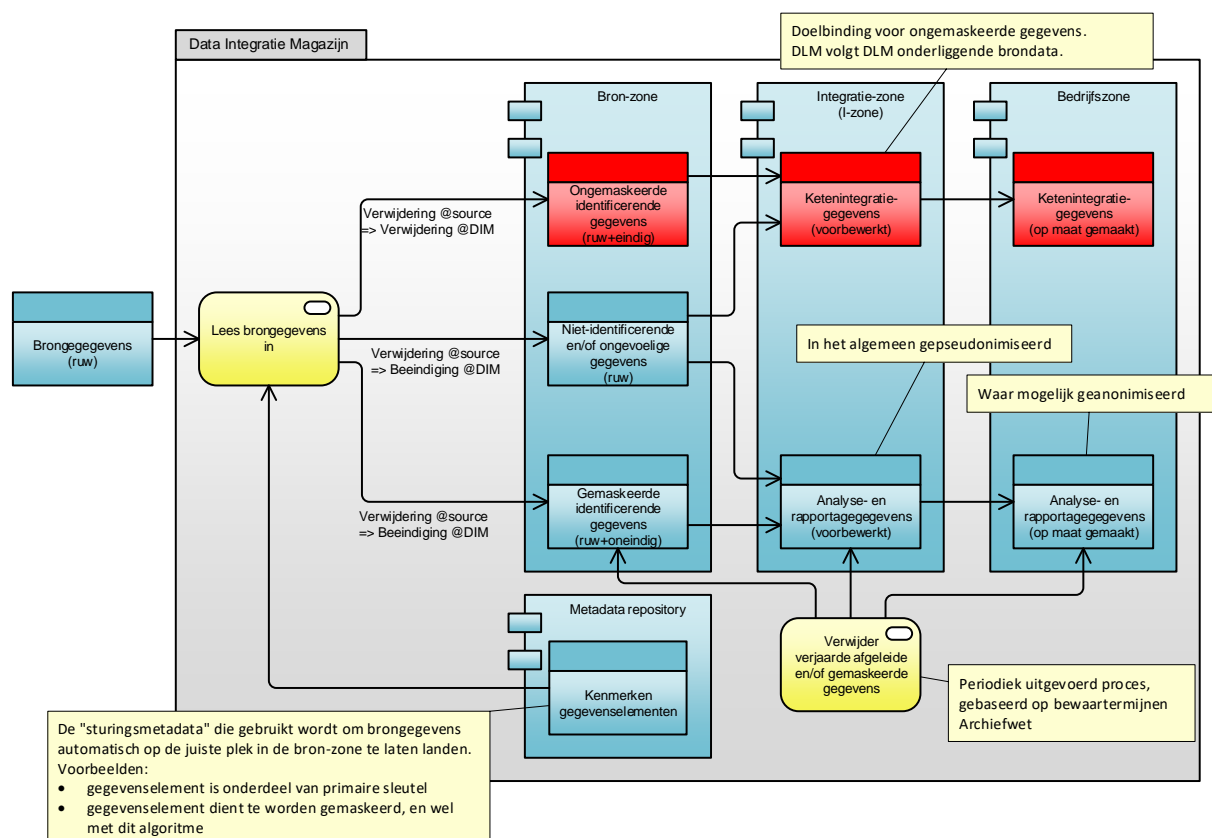
³⁶ Ook voor gebruik van gemaskeerde data blijft een GEB (of tenminste een GEB-check) vereist.

Gegevenscategorie	Effect
Op maat gemaakte gegevens Als voorbereekte gegevens, maar dan toegespitst op één specifieke gebruikersgroep of gegevensproduct.	De huidige DWH's maken geen expliciet onderscheid tussen voorbereeking voor keten-integratie en die voor analyse/rapportage; impliciet gebeurt dit wel, bijvoorbeeld doordat de gegevens via separate datamarts en/of interfaces lopen. Binnen het DIM zal elk informatieproduct geclassificeerd worden qua privacy-gevoeligheid, en zal de Data Lifecycle daarop afgestemd worden (zie onder Voorbereekte gegevens)

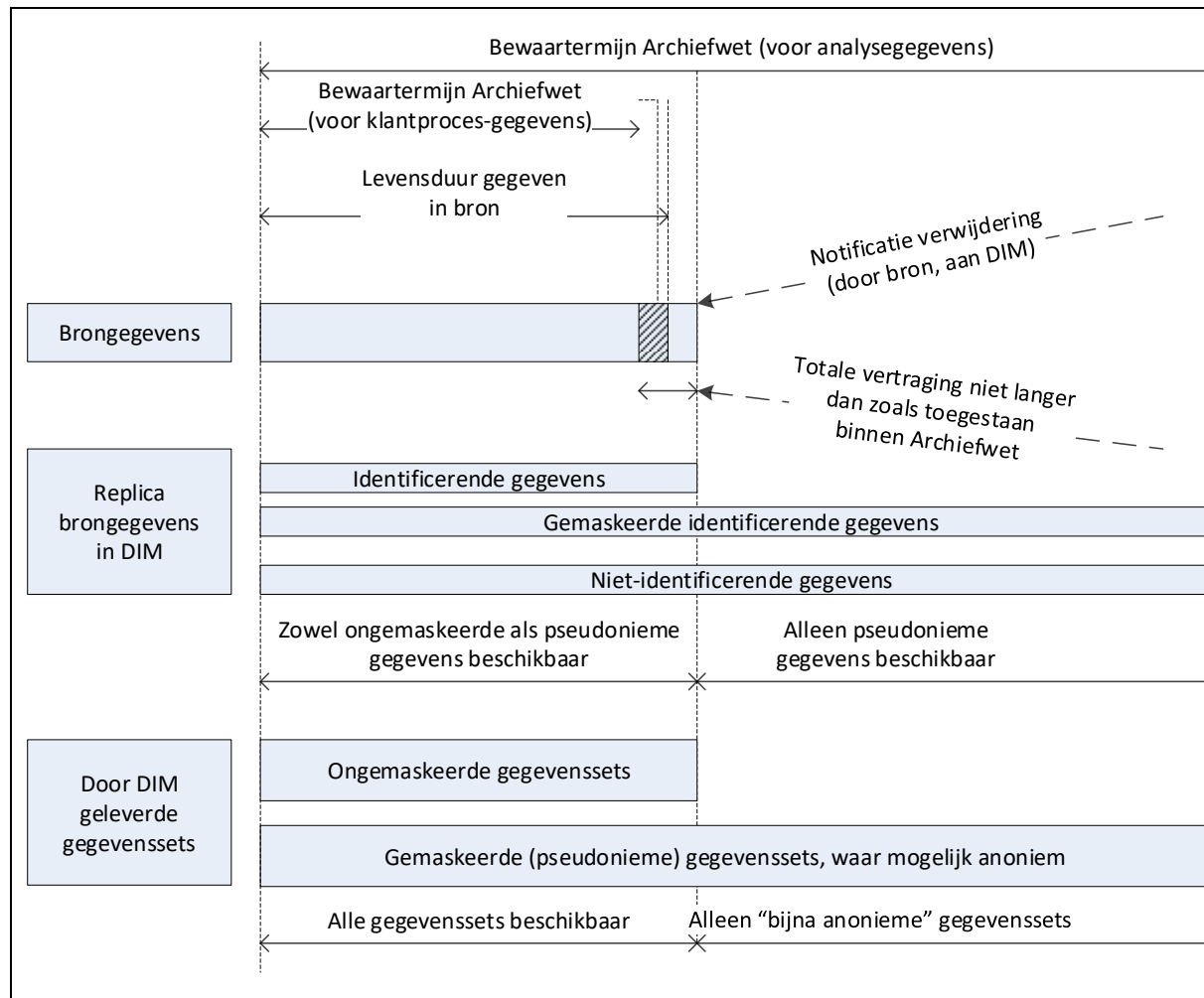
Voor alle gegevens geldt dat ze uiteindelijk uit het DIM worden verwijderd. Dit om compliant te blijven met de Archiefwet.

N.B. In bovenstaande moet "Natuurlijk persoon" gelezen worden als "Klant van UWV" en/of "Burger", dus niet als "Medewerker van UWV". Maskering van personeelsgegevens gebeurt niet bij opslag, maar, indien nodig, bij levering aan de afnemers van het DIM.

Gecombineerd met de zonering van het DIM (zie 4.2 Informatievoorziening) levert dit het volgende plaatje op:



E.e.a. resulteert in de volgende "data lifecycles":



N.B. De interfaces van bron naar DIM zullen zo worden ingericht dat ook de verwijdering van gegevens uit die bron kan worden gerepliceerd in het DIM, bijvoorbeeld door het doorgeven van "verwijder-gebeurtenissen", (door het leveren van records met een "verwijder-vlag"), of door een periodieke volledige levering ter reconciliatie.

Waar het doorgeven van verwijderingen niet mogelijk is zal het DIM, op basis van door bron geleverde specificaties, de verwijderlogica van de bron "naspelen". De verantwoordelijkheid voor de specificatie van die logica, en de validatie dat e.e.a. correct in het DIM is geïmplementeerd, blijft in dergelijk geval overigens bij de bron.

N.B. Voor vernietiging moet er altijd een controle door een mens plaatsvinden, dus nooit volledig automatisch. Dit om te voorkomen dat per ongeluk data wordt vernietigd die nog nodig was voor een ander proces. Deze controle-stap is onderdeel van het verwijderproces voor gemaskeerde en niet-identificerende gegevens; voor identificerende gegevens wordt deze controlestap vóór vernietiging uitgevoerd bij de bron. Omdat het DIM voor ongemaskeerde data het DLM van de bron volgt is deze controle-stap voor deze data dus al impliciet onderdeel van het proces.

4.1.2 Architectuurprincipes Gegevens

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
1) De UWV-gegevenshuishouding is gestandaardiseerd.	<p>De ruwe gegevens in het DIM zijn identiek aan die in de bron Eventuele maskering is identiek binnen alle deelstores van het DIM.</p> <p>De functionele gegevensdefinities in (de metadata-componenten van) het DIM worden, waar beschikbaar, overgenomen uit FUGEM/CGM.³⁷</p> <p>N.B. De technische implementatie van deze functionele gegevensdefinities verschilt per bron, en is onderdeel van de technische documentatie van die bron. Het (door de bron) leveren van een gedetailleerde beschrijving van de functionele betekenis (in FUGEM/CGM-termen) van de aan het DIM geleverde technische velden is daarom integraal onderdeel van de interface-standaarden³⁸.</p> <p>Naast FUGEM/CGM gebruikt UWV ook het Centraal Handboek Uniforme Definities Stuur- en Verantwoordingsinformatie UWV (HUDSV). Ook de standaard-definities in dit document worden opgenomen in de metadata-componenten van het DIM.</p>
2) UWV geeft inzicht in en draagt zorg voor de kwaliteit van haar gegevens.	<p>Het DIM biedt mogelijkheden om datakwaliteit te meten/toetsen, zowel incidenteel als binnen het reguliere proces. De verantwoordelijkheid voor datakwaliteit blijft echter bij de bron.</p>
3) Gegevensregistraties voldoen aan archiveringsverplichtingen.	<p>In de bronzone van het DIM zijn alle gegevens langs zowel de geldigheidsdimensie als de transactiedimensie vastgelegd. Voor de overige zones is dit afhankelijk van de eisen van de afnemer(s).</p> <p>Voor ongemaskeerde persoonsgegevens volgt het DIM de retentiestrategie (het Data Lifecycle Management) van de bron, gemaskeerde persoonsgegevens en overige gegevens worden conform de langere bewaartermijnen voor analyse- en rapportage-gegevens opgeslagen.</p> <p>N.B. De replicatie van gegevens naar het DIM ontslaat de bron niet van haar archiveringsverplichtingen; de bron dient hier nog steeds (in eigen applicatie en/of m.b.v. het Elektronisch Archief) zorg voor te dragen. Het DIM draagt wél zorg voor het beschikbaar blijven van gemaskeerde gegevens t.b.v. statistische analyse en verslaglegging, ook nadat de onderliggende (ongemaskeerde) gegevens uit de bron zijn verwijderd.</p>

³⁷ Voor binnen het DIM afgeleide gegevens wordt óf de FUGEM van de bron leidend, óf, voor afleidingen, functionele deelmodellen binnen het DIM (al dan niet gebaseerd op het HUDSV).

³⁸ Deze gedetailleerde beschrijving wordt vastgelegd in een RLO (RecordLayout)

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
4) Gegevens worden verzameld, verstrekt en toegankelijk gemaakt voor zover als daarvoor een wettelijke grondslag bestaat.	Het DIM verwijdert ongemaskeerde gegevens zodra de bewaartermijn ervoor verstreken is. Dit verwijderen gebeurt door verwijderingen in de bron ook te laten resulteren in een verwijdering in het DIM. ³⁹ De IV-organisatie DIM/DWH stelt alleen gegevens beschikbaar (aan afnemers) als daarvoor rechtsgrond, proportionaliteit en subsidiariteit geldt. Binnen het DIM worden gegevens daarnaast waar mogelijk gemaskeerd; het DIM volgt daarmee het "privacy by design" principe.
5) UWV stelt uitsluitend gegevens beschikbaar waarvan UWV zelf de bron is of die UWV als authentiek gegeven beheert.	De processen van de IV-organisatie DIM/DWH zullen zo zijn ingericht dat (indirect) aan externe partijen geleverde gegevens altijd voldoen aan dit principe.
6) UWV vraagt geen gegevens uit, die al beschikbaar zijn binnen de keten of binnen het stelsel van basisregistraties.	<i>(niet van toepassing; de IV-organisatie DIM/DWH vraagt extern UWV geen gegevens uit⁴⁰)</i>
7) UWV legt geen redundante gegevensverzamelingen aan.	De IV-organisatie DIM/DWH legt, in het DIM, juist uitsluitend redundante gegevensverzamelingen aan. Het betreft hier echter "gecontroleerde" redundantie conform het vigerend UWV-beleid. De beschikbaarheid van het DIM als gemeenschappelijke voorziening verkleint de noodzaak voor redundantie elders in UWV.
8) UWV past de waarde van een gegeven uitsluitend aan op aangeven van de bron waaruit het gegeven is verkregen.	Door de IV-organisatie DIM/DWH (of in het DIM) worden geen gegevens gewijzigd, anders dan op aangeven van de bron. Wel worden in het DIM afgeleide gegevens berekend en vastgelegd. Een bijzonder geval zijn de door IV-organisatie DIM/DWH (of elders binnen GD Uitvoeren) beheerde reference data. Deze worden buiten het DIM beheerd (maar dus wel potentieel door de IV-organisatie DIM/DWH), en vervolgens met een reguliere interface in het DIM geladen. Dit garandeert dat ook voor deze reference data de wijzigingsgeschiedenis volledig traceerbaar is.
9) UWV biedt degene over wie UWV gegevens registreert of verwerkt, op verzoek inzicht in de inhoud en het gebruik van die geregistreerde gegevens.	De IV-organisatie DIM/DWH houdt van alle gegevens de herkomst én het gebruik vast (middels een combinatie van inwinnings- en leverafspraken, gecombineerd met binnen het DIM vastgelegde metadata), en biedt middelen om deze te analyseren en/of te exporteren.
10) Beschikbaarstelling van gegevens aan of door UWV vindt uitsluitend plaats op basis van een vastgelegde afspraak	De IV-organisatie DIM/DWH betreft alleen gegevens uit een bron als daarvoor een GIA (GegevensInwinningsAfspraak) is vastgelegd ⁴¹ , en levert alleen conform een GLA (GegevensIbwinningAfspraak), waarin onder andere rechtsgrond, proportionaliteit en subsidiariteit zijn vastgelegd. Een bijzonder geval zijn de door IV-organisatie DIM/DWH (of elders binnen GD Uitvoeren) beheerde reference data. Afspraken hierover worden wel formeel vastgelegd, maar potentieel gebeurt dat niet via GIA's.

³⁹ De fysieke verwijdering uit het DIM wordt bepaald door het moment waarop de bron die verwijdering doorgeeft, en dat is weer afhankelijk van de functionaliteit die de bron op dit gebied biedt.

⁴⁰ Het DIM slaat wel elders binnen UWV extern opgevraagde gegevens op, maar deze zijn dan dus al "beschikbaar binnen de keten".

⁴¹ Ook eventuele gebruiksbeperkingen voor de geleverde data zijn in deze GIA vastgelegd.

4.1.3 Afwijkingen van architectuurprincipes

Generiek architectuurprincipe	Aard van de afwijking	Reden voor de afwijking	Besluitvorming
Geen			

4.1.4 Hoofdstructuur van de oplossing

Het DIM creëert geen nieuwe bedrijfsobjecten, en wijzigt geen bestaande.

Wel wordt:

- het Data Lifecycle Management van de replica-objecten in het DIM nauw aangesloten op dat van de originele objecten in de bron
- Waar relevant ook een gemaskeerde versie van die replica gemaakt.

(zie 4.1.1 Afbakening Gegevens)

N.B.

Door de IV-organisatie DIM/DWH en haar afnemers worden wel afleidingen gemaakt, en deels worden deze ook opgeslagen in het DIM. Dit resulteert wél in informatieobjecten, maar niet in nieuwe/gewijzigde bedrijfsobjecten.

4.2 Informatievoorziening

4.2.1 Afbakening Informatievoorziening

Het DIM vervangt de huidige drie, elkaar deels overlappende, centrale data warehouses:

- DWH 1.0 (DWARFS)
Gebaseerd op COBOL-maatwerksoftware
- DWH 2.0 (NGO)
Gebaseerd op maatwerk PL/SQL-generatoren
- DWH 3.0 (UDS)
Gebaseerd op Oracle Warehouse Builder (OWB)

Voor alle drie deze data warehouses bestaan aanzienlijke continuïteitsrisico's:

- Kennis van de onderliggende tooling van DWH 1.0 en 2.0 is nog maar zeer beperkt aanwezig;
- De voor DWH 3.0 gebruikte tooling wordt vanaf juli 2022 nog maar beperkt door de leverancier ondersteund.

Het project Datafabriek heeft daarom de vervanging van alle drie de systemen in scope. Hierbij zal door meer dan één systeem geleverde functionaliteit en/of data uiteraard vervangen worden door één gemeenschappelijke oplossing.

Systeem	Omschrijving	Effect
DWH 1.0	COBOL Datawarehouse	Wordt vervangen door het DIM
DWH 2.0	Oracle/maatwerk Datawarehouse	Wordt vervangen door het DIM
DWH 3.0	Oracle/OWB Datawarehouse	Wordt vervangen door het DIM
DMAP	Analyse-omgeving K&S en Handhaven	Wordt gevoed door het DIM, kan daardoor: <ul style="list-style-type: none"> • Maskering uitbesteden aan het DIM • Data gedeeltelijk virtueel ontsluiten

Systeem	Omschrijving	Effect
Overige DWH's	(PAO, AMI, ADM, etc.)	Kunnen deels vervangen worden door (gegevensvensters op) het DIM (in later stadium te beslissen; geen urgentie)
BI-omgevingen	(BusinessObjects, etc.)	In het project worden bestaande, deels binnen BI-omgevingen geïmplementeerde, informatieproducten vervangen door functioneel equivalente alternatieven, meestal binnen diezelfde BI-omgeving. ⁴²⁴³
Binnen BI-omgevingen gerepliceerde gegevens	(m.n. Qlikview-tabellen)	Gerepliceerde gegevens kunnen deels vervangen worden door (gegevensvensters op) het DIM (in later stadium te beslissen; geen urgentie)

N.B. Alleen DWH 1.0/2.0/3.0 zijn in scope van het project Datafabriek.

4.2.2 Architectuurprincipes Informatievoorziening

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
P1 - UWV stelt stabiliteit, continuïteit en informatiebeveiliging voorop	Het project vermindert de continuïteits- en stabiliteitsrisico's van de huidige DWH-omgevingen aanzienlijk, door het uitfasen van end-of-life software en het, bij vervanging, expliciet meenemen van beheerbaarheid. De nieuwe IV-omgeving (het Data Integratie Magazijn) wordt van de grond af opnieuw ontworpen. Informatiebeveiliging is hierbij één van de belangrijkste ontwerp-aspecten.
P2 - UWV kiest voor besturing van de informatievoorziening in IV ketens, en binnen centrale kaders	De IV-organisatie DIM/DWH is een centrale organisatie voor (DIA-gerelateerde) gegevensvoorziening, die de gemeenschappelijke IV-oplossing DIM beheert, en daarnaast expertise biedt m.b.t de generieke ETL+ tooling.
P3 - UWV kiest voor tijdig en geleidelijk vernieuwen	Het project Datafabriek voert als voornaamste activiteit een urgente "technology refresh" uit. Geleidelijkheid uit zich hier met name in de gefaseerde migratie van legacy DWH's naar het nieuwe Data Integratie Magazijn.
P4 - UWV zet gepaste outsourcing in voor applicatieontwikkeling en -beheer en volledige outsourcing voor exploitatie van infrastructurele voorzieningen	Het DIM zal maximaal gebruik maken van standaardpakketten.

⁴² Uitkomst van het project BI Portfolio (buiten scope project DataFabriek) was dat voor de drie voornaamste op dit moment gebruikte BI-tools (BusinessObjects, Qlikview en PowerBI) verlengde bedrijfsgarantie geboden gaat worden (status bij schrijven van deze PSA: voorlegger onderweg naar RvB).

⁴³ Het implementeren van nieuwe BI-tooling of het verbeteren van de (technische) ondersteuning van bestaande BI-omgevingen valt buiten scope van het project DataFabriek.

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
P5 - UWV voert zijn bedrijfs- en werkprocessen geautomatiseerd uit	<p>De interfaces met de bronsystemen worden zo ontworpen dat het inlezen en verwerken van brongegevens in het DIM, inclusief controles op volledigheid/volgordelijkheid volledig automatisch kan worden uitgevoerd.</p> <p>Het DIM maakt waar mogelijk gebruik van metadata-gedreven oplossingen. Hierdoor wordt zelfs een deel van het ontwikkelproces geautomatiseerd.</p>
P6 - UWV kiest voor het persoonlijk maken van haar digitale dienstverlening	Niet relevant voor IIV-organisatie DIM/DWH / DIM.
P7 - UWV kiest voor het gebruik van gemeenschappelijke en generieke voorzieningen	De IV-organisatie DIM/DWH is een centrale organisatie voor (DIA-gerelateerde) gegevensvoorziening, die de gemeenschappelijke IV-oplossing DIM beheert, en daarnaast expertise biedt m.b.t de generieke ETL+ tooling. Deze ETL+ tooling is doeltechnologie.
P8 - UWV kiest voor een inrichting van het ICT-landschap met gestandaardiseerde en vervangbare bouwblokken, zowel in de infrastructuur als in de software	<p>Het DIM zal maximaal gebruik maken van standaardpakketten.</p> <p>Het DIM zal opgebouwd worden uit ontkoppelde "zones".</p> <p>Qua infrastructuur zal initieel gebruik gemaakt worden van het bestaande IBM-rekencentrum, omdat implementatie direct in het nieuwe DXC-rekencentrum een te grote impact heeft op de tijdslijnen van het project Datafabriek. Migratie naar DXC valt overigens wel binnen de doorlooptijd van het project.</p> <p>Aandachtspunt bij migratie naar DXC is dat de door dat nieuwe rekencentrum geleverde standaardblokken voldoende vrijheidsgraden moeten bieden om aan de specifieke eisen die DWH's en BI- / analyse-omgevingen stellen te voldoen.</p>
P9 - UWV kiest bij de selectie van een IV-oplossing voor hergebruik boven standaardoplossing, en standaardoplossing boven maatwerk	Het DIM zal maximaal gebruik maken van standaardpakketten.
P10 - UWV kiest uitwisselingsstandaarden op basis van overheids-, open- en Marktstandaarden	<p>Het DIM zal, qua inlezen van brongegevens, aansluiten op de "market best practices" op dit gebied.</p> <p>Voor gegevensleveringen aan afnemers zal het DIM zich conformeren aan de hiervoor vereiste standaarden.</p>

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
P11 - UWV kiest ervoor om het eigenaarschap van applicaties integraal te beleggen	<p>Het DIM is eigendom van Gegevensdiensten.</p> <p>NB: De gegevens in het DIM hebben wél verschillende eigenaren:</p> <ul style="list-style-type: none"> • Ruwe data is van de broneigenaar • Afgeleide data is van de afnemer (of van Gegevensdiensten als gedelegeerd eigenaar namens de afnemers)⁴⁴

4.2.3 Afwijkingen van architectuurprincipes

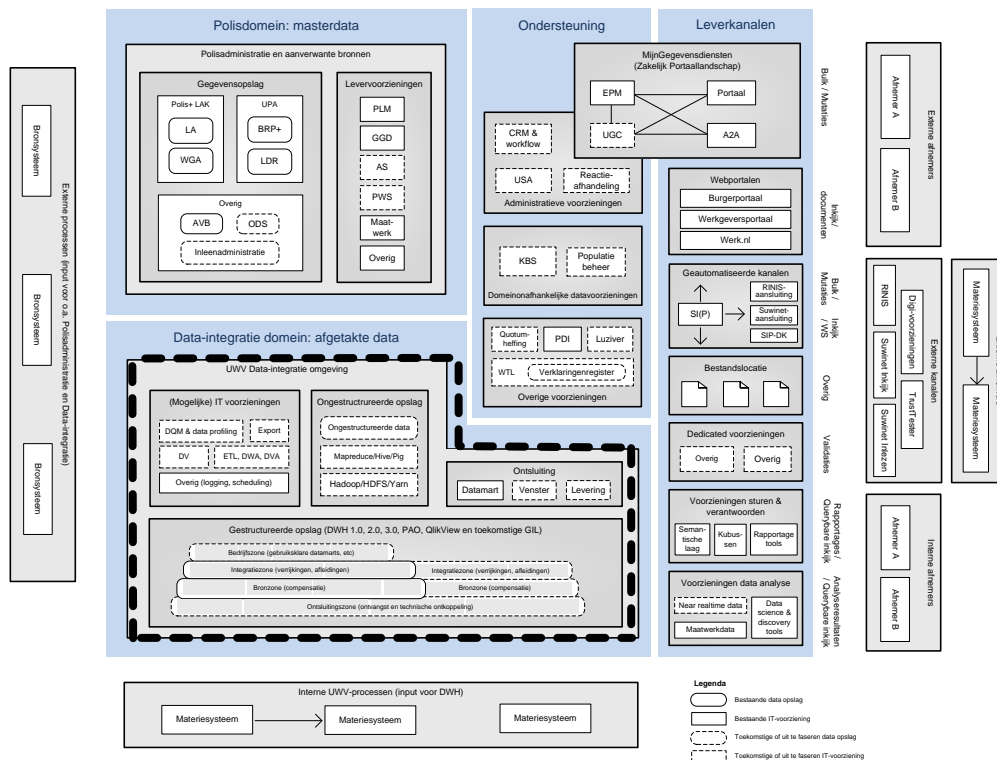
Generiek architectuurprincipe	Aard van de afwijking	Reden voor de afwijking	Besluitvorming
P4 - UWV zet gepaste outsourcing in voor applicatieontwikkeling en -beheer en volledige outsourcing voor exploitatie van infrastructurele voorzieningen	Voor de exploitatie van infrastructurele voorzieningen zal géén sprake zijn van volledige outsourcing: m.n. DBA-activiteiten zullen binnen de eigen verantwoordelijkheid van de IV-organisatie DIM/DWH vallen.	<p>Voor een DWH is "wijziging de regel". Dat maakt dat m.n. bij outsourcing van de DBA-activiteiten de wendbaarheid onacceptabel daalt.</p> <p>Daarnaast geldt dat voor DWH's specifieke DBA-kennis benodigd is. Outsourcing biedt dus ook geen "economy of scale".</p> <p>Vandaar dat de IV-organisatie DIM/DWH, net als de afdeling DWH nu, zelf zorg zal dragen voor de DBA-activiteiten.</p>	Met ICT-Services is afgesproken dat het DWH ook in eerste instantie 'as is' over zal gaan naar het nieuwe HRC, en dat te zijner tijd wordt bekeken (als meer bekend is over het HRC en bijbehorende inrichting) hoe DWH/DF zo goed mogelijk kan aansluiten bij de standaard situatie. ⁴⁵

4.2.4 Hoofdstructuur van de oplossing

Het DIM staat in onderstaande figuur, afkomstig uit DA GEIN v1.0, afgebakend middels de zwarte omlijning. Het DIM bevindt zich daarmee volledig in het DIA-domein (Data Integratie en Analyse).

⁴⁴ Zie de notitie "Eigenaarschap en verantwoordelijkheden Analyseomgevingen" (Ralph Foorthuis/Astrid Kerssens, v1.1 - 4 december 2018) voor meer detail.

⁴⁵ Dit proces loopt nu, maar valt buiten het project. Zie ook 1.4 Beslis- en discussiepunten.



Het Data Integratie Magazijn is in essentie een "klassiek" enterprise datawarehouse (EDW), uitgebreid met een aantal "moderne" concepten:

- Bron-zone met historisch gestapelde ruwe data⁴⁶, als eerste stap naar een data lake;
- Bredere en flexibelere inzetbaarheid door "multi-schema" integratiezone;
- Ontkoppelde lagen t.b.v. maximale wijzigingsflexibiliteit;
- Pseudonimisering en anonimisering als dagelijkse praktijk;
- Geïntegreerd metadata management en datakwaliteitsbeheer;
- Metadata gedreven ETL-functionaliteit⁴⁷;
- Klaar voor doorgroei naar een "analytical ecosystem" (met daarin ook data virtualisatie en Hadoop- en/of NoSQL-componenten);
- Klaar voor een gefedereerde bronzone, waar gegevens uit "perfecte bronnen"⁴⁸, slechts virtueel onderdeel van uit maken;
- Klaar voor levering van halffabrikaten t.b.v. self service BI en advanced analytics;
- In staat om gegevens van "service enabled" bronnen⁴⁹ te ontvangen als "gegevens-gebeurtenissen" (en dus niet in bulk) en op te vangen in queues.

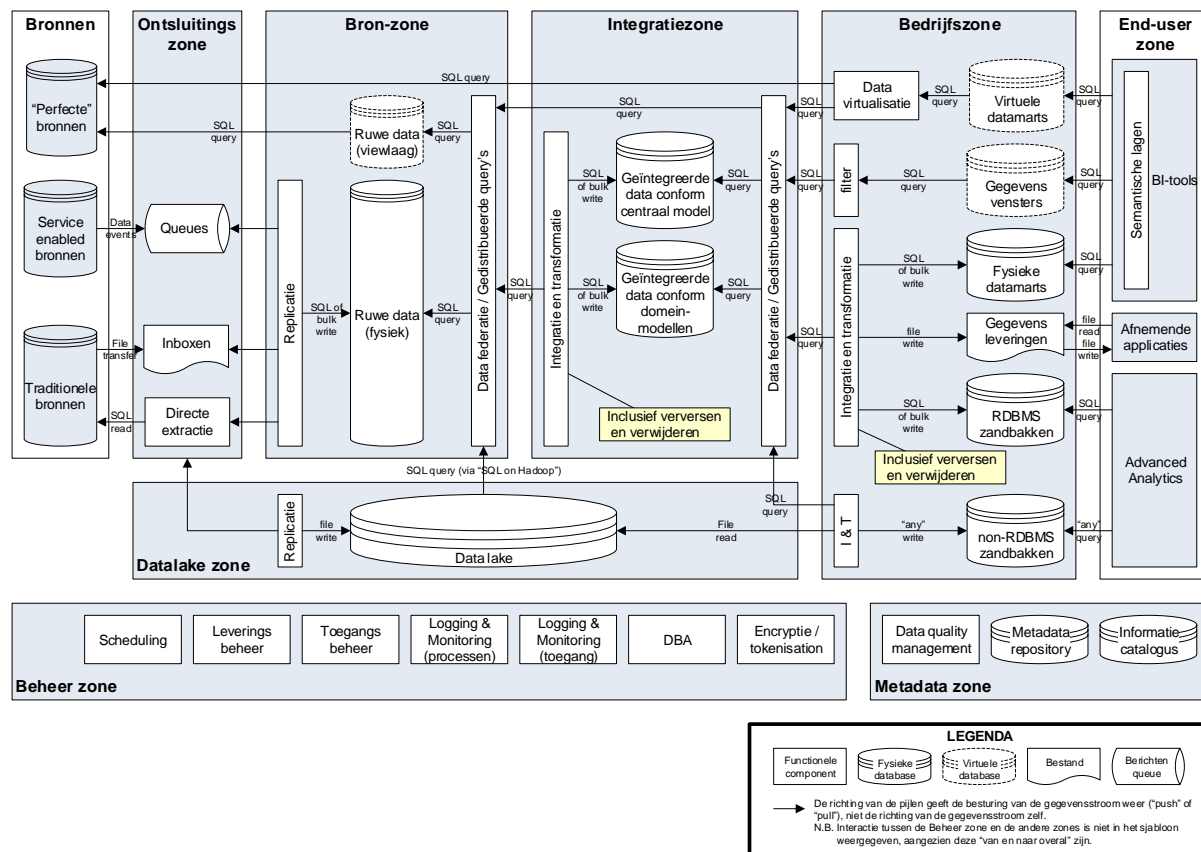
⁴⁶ Met “ruwe data” wordt hier “data zoals aangeleverd door de bron” bedoeld, dus zonder inhoudelijke transformatie of interpretatie binnen het DIM.

⁴⁷ Waar dit waarde biedt; niet elk type functionaliteit leent zich voor (metadata-gedreven) generalisatie

⁴⁸ Bronnen waarvoor geen enkele vorm van compensatie noodzakelijk is.

Kanttekening daarbij: geen enkele van de huidige DWH-bronnen is "perfect".

⁴⁹ Bronnen die in staat zijn om creatie, wijziging en verwijdering van gegevens als "gebeurtenissen" (op basis van publish en subscribe), te leveren aan het DIM (via de ESB en/of messages, ontvangen in queues).



De kern van het Data Integratie Magazijn bestaat uit 4 zones:

- De **Ontsluitingszone** zorgt voor de technische ontkoppeling tussen bronsystemen en DIM;
- De **Bronzone** "compenseert" de eventuele tekortkomingen van de bronsystemen m.b.t. gegevensbeschikbaarheid en historisch besef;
- De **Integratiezone** bevat, t.b.v. bruikbaarheid, consistentie en efficiëntie, veel (en/of verplicht) gebruikte gegevensmodellen en afgeleide gegevens;
- De **Bedrijfszone** stelt de data in het DIM "op maat" beschikbaar aan de diverse eindgebruikers (mensen en machines).

Bovenstaande vier zones vormen het hart van de gegevensketen van de gegevensbronnen naar (de tools en applicaties van) de afnemers in de, buiten het DIM vallende, **End-user zone**. Ze zijn grotendeels gebaseerd op RDBMS-technologie, en daarmee vooral geschikt voor verwerking van relationeel gestructureerde en qua structuur redelijk stabiele data.

Voor minder/anders gestructureerde, of qua structuur sneller wijzigende, gegevens bevat het DIM een **Datalake zone**, die waarschijnlijk gebaseerd zal worden op Open Source oplossingen zoals die uit de Hadoop-familie.

De technische en business metadata betreffende de (verwerking van) gegevens in het DIM zijn terug te vinden in de **Metadata zone**. De metadata repository en de informatiecatalogus in deze zone bevatten alle technische respectievelijk functionele metadata die noodzakelijk is voor gegevensverwerking en -levering door het DIM. Deels wordt deze metadata door het DIM zelf gegenereerd (m.n. horizontale lineage), deels wordt deze overgenomen uit elders gecreëerde metadata (m.n. functionele gegevensdefinities brondata uit de FUGEMs). De koppeling tussen technische velden en hun functionele equivalent (de verticale lineage) zal ook binnen de DIM-metadata worden vastgelegd en ontsloten.

De verticale lineage wordt in IGC beschikbaar gemaakt voor de eindgebruiker.

In IDA wordt deze aangemaakt en gepusht naar IGC.

De functionaliteiten t.b.v. besturing en beheer van het DIM, ten slotte, zijn gegroepeerd in de **Beheer zone**.

N.B. De Datalake zone is niet in scope van het project Datafabriek, aangezien de inrichting ervan sterk afhankelijk is van de technologische en organisatorische keuzes die UWV op dit gebied (buiten het project) gaat maken.

Toelichtende tabel:

Systeem	Omschrijving	Effect
DIM	Enterprise Data Warehouse met doorgroei mogelijkheid naar "analytical ecosystem"	Nieuw, vervangt drie bestaande DWH's

Te leveren diensten

Het Data Integratie Magazijn levert:

- Toegang tot de in het DIM opgeslagen gegevens, bv. via een datamart of gegevensvenster, als input voor rapportage/analyse door tools buiten het DIM;
- Uitgaande interfaces / gegevensleveringen: kopieën van in het DIM opgeslagen gegevens t.b.v. opslag/verwerking buiten het DIM;
- Metadata over al deze gegevens;
- Gegevensdefinities (technisch: data-elementen en functioneel: bedrijfstermen);
- Horizontale lineage (het pad dat een gegeven vanaf de bron tot aan de DIM-afnemer heeft doorlopen);
- Verticale lineage (de relatie tussen bedrijfsterm en de data-elementen waarin deze zijn terug te vinden);
- Datakwaliteitsinformatie.

Maar geen:

- Rapporten of dashboards.⁵⁰

Rol binnen UWV

Het Data Integratie Magazijn is:

- De verplichte leverancier (voor heel UWV) van gegevens t.b.v. verslaglegging waarvoor UWV-brede consistentie vereist is;
- De verplichte leverancier (voor heel UWV) van divisie-overstijgende analysegegevens;
- De voorkeursleverancier van geïntegreerde/afgeleide gegevens (in bulk) met lage actualiteitseisen;
- De tweede voorkeursleverancier⁵¹ van ruwe brongegevens (in bulk) met lage actualiteitseisen;
- Géén leverancier van "strikt vertrouwelijke" gegevens (met name medische en strafrechtelijke gegevens), tenzij er geen andere optie is (maar zie gerelateerd beslispunt in paragraaf 1.4);

⁵⁰ De IV-organisatie DIM/DWH biedt wel een rapportagedienst o.b.v. van gegevens in het DIM, maar de daarvoor gebruikte BI/rapportage-tools vallen in de End-user zone, en daarmee buiten het DIM.

⁵¹ De eerste voorkeur blijft de bronapplicatie zelf (als die daartoe in staat is)

- Géén leverancier van gegevens met hoge actualiteitseisen ("near realtime"). Daar ligt het primaat bij service/bus-oplossingen zoals die binnen het KIA/KOA-concept;
- Een mogelijke leverancier van alle overige gegevens. Er is geen beperking m.b.t. het type gegeven (anders dan de vertrouwelijkheid) of het type afnemend proces (anders dan indirect, via de actualiteitseisen).

N.B. Bovenstaande is gebaseerd op de Doelarchitectuur Gegevensintegratie (DA GEIN). Een verdere verbijzondering van de rol van het DIM binnen de architectuur van UWV zal (onder aansturing van de Divisiearchitect Gegevensdiensten) worden vastgelegd in de nieuwe Enterprise Architectuur Gegevenshuishouding (EA GH), en de Doelarchitectuur Gegevensdiensten (DA GD).

Inrichtingsprincipes

De vereiste functies en de prioriteit van de te leveren verbeteringen vertalen zich in een vijftal leidende principes voor de inrichting van de applicatiearchitectuur van het Data Integratie Magazijn:

1. Bewezen concepten

Gezien de risico's op korte termijn m.b.t. stabiliteit en continuïteit van de bestaande centrale data warehouses moet de uitrol van het DIM zo voorspelbaar en beheersbaar mogelijk zijn. Opkomende gegevensintegratietechnologieën ("big data", data virtualisatie) zijn daarom geen onderdeel van de "ruggengraat" van het DIM.⁵² Om ook stabiliteit en continuïteit op de langere termijn te kunnen garanderen voorziet het DIM overigens wel in de ruimte om deze technologieën op een later tijdstip te incorporeren.

2. Geen realtime ambitie

Ook op de langere termijn wordt niet verwacht dat het DIM een rol gaat spelen in het leveren van actuele ("near realtime") gegevens. Het DIM wordt daarom niet ingericht om (al dan niet op de langere termijn) gegevens met een hoge actualiteit (recenter dan één dag oud) te kunnen leveren, aangezien dit de oplossing nodeloos zou compliceren. Wel wordt, binnen het DIM, het aantal replica's geminimaliseerd (kortste keten principe) om zo de vertraging binnen het DIM te minimaliseren.

Daarnaast voorzien we dat een aantal innovatieve nieuwe technologieën op termijn een rol zal spelen binnen het DIA-domein. Op dat moment zal beoordeeld worden welke requirements en beheeraspecten daarbij dan een rol gaan spelen.

N.B. Er is geen beperking qua afnemend proces, mits dat proces kan leven met de beperkingen van het DIM (geen gegarandeerde performance/beschikbaarheid, beperkte actualiteit; "geen realtime ambitie" mag dus niet gelezen worden als "geen rol in de dagelijkse operatie").

3. Maximale ontkoppeling

Om de flexibiliteit van de oplossing te maximaliseren worden de verschillende schakels in de door het DIM ondersteunde gegevensketen(s)⁵³ zoveel mogelijk ontkoppeld. Dit minimaliseert:

- de impact van wijzigingen in een bronsysteem op het DIM;
- de impact van wijzigingen binnen een component van het DIM op de overige componenten van dat DIM;
- de impact van wijzigingen binnen het DIM op de afnemers van dat DIM.

⁵² Dit is ook conform UWV IT-strategie en -beleid

⁵³ Een "gegevensketen" is het geheel aan proces- en applicatie-stappen van gegevensbronnen naar (de tools en applicaties van) de afnemers

4. **Specifiek waar mogelijk, gemeenschappelijk waar noodzakelijk of waardevol**

Zowel in de positie van het DIM binnen de UWV enterprise architectuur (de "scope") als in de opzet van gegevensketens binnen het DIM zelf wordt, om de flexibiliteit van de oplossing te maximaliseren, géén "one size fits all" benadering toegepast.

In het ontwerp van het DIM vertaalt zich dat bijvoorbeeld in de "multi-schema" opzet van de integratiezone.

5. **Kortste keten**

De time-to-market van de via het DIM geleverde gegevensproducten wordt verder verkort door toepassing van het "kortste keten" principe: voor een gegevensproduct worden nooit meer transformatie/replicatie-stappen uitgevoerd (of DIM-zones doorlopen) dan noodzakelijk om de functionele en non-functionele vereisten van een gegevensproduct te dekken.

5 TECHNISCHE ARCHITECTUUR

In dit hoofdstuk worden de onderdelen en componenten van de technische architectuur van UWV belicht die binnen de scope van het project vallen, alsmede de technische kenmerken van het applicatielandschap.

Het hoofd architectuurdomein Technische architectuur bestaat uit de architectuur sub-domeinen

- Software
- Technische infrastructuur

5.1 Software

5.1.1 Afbakening Softwarecomponenten

Het DIM wordt voor het overgrote deel geïmplementeerd o.b.v. metadata-gedreven standaardsoftware. Zie daarvoor paragraaf 5.1.4.

Component	Omschrijving	Effect
Ontsluitingszone	Ontvangst van binnenkomende gegevens, ter verwerking in bronzone	Inboxen onderdeel van ETL-shared storage (hot). Queues in deze fase nog niet ingericht. Directe extractie via database links ⁵⁴
Hot archive leveringen	Opslag van recent verwerkte gegevensleveringen (potentieel nog nodig bij Disaster Recovery)	Onderdeel van ETL-shared storage (hot).
Cold archive leveringen	Archief van eerder verwerkte gegevensleveringen (niet meer nodig bij Disaster Recovery)	Onderdeel van ETL-shared storage (cold).
Bronzone	Opslag van ruwe brondata + historie	Onderdeel van DIM database
Integratiezone	Opslag van voorberewte data met brede relevantie voor afnemers.	Onderdeel van DIM database
Bedrijfszone	Ontsluiting van gegevens t.b.v. afnemers.	Onderdeel van DIM database, in de toekomst eventueel uit te breiden met andere oplossingen (data virtualisatie, specifieke zandbak-infrastructuur t.b.v. advanced analytics)
ETL-mappings	Definities van uit te voeren gegevenstransformaties, vastgelegd als metadata in een repository. Deels uitgevoerd door de ETL-engine, deels door de target-database ("ELT-patroon"). In dat laatste geval genereert de ETL-engine de daarvoor benodigde PL/SQL (voor Oracle) of equivalent. ⁵⁵	Opgenomen in ETL-repository
DIM statische metadata	Data definities (functioneel en technisch), lineage (horizontaal en verticaal)	Opgenomen in IBM Information Governance Catalog

⁵⁴ Aan deze database links worden hoge eisen gesteld. Zie document Interface-standaarden voor details.

⁵⁵ Ook ELT-generatie voor NoSQL-databases wordt (beperkt) ondersteund.

DIM stuurmetadata	Besturing van de generieke ETL-processen (de "harnassen")	Maatwerktabellen in DIM database.
DIM logmetadata	Logs van de aanlevering van gegevens aan het DIM, en de verwerking ervan binnen het DIM.	Deels standaard geleverd door ETL-engine, deels maatwerktabellen in DIM database.
DIM actualiteitsmetadata	Metadata die de actualiteit van de in het Dim beschikbare gegevens beschrijft.	Maatwerktabellen in DIM database.

De door het project Datafabriek verworven ETL+ tooling voldoet niet alleen aan de eisen van lijnorganisatie Datafabriek en DIM, maar biedt ook voldoende brede functionaliteit om inzetbaar te zijn als UWV-brede doeltechnologie voor ETL (inclusief ETL-gerelateerde metadata) en datakwaliteitsbeheer⁵⁶. Deze UWV-brede inzet is niet in scope van het project Datafabriek, maar is wel meegenomen in de aanbestedingscontracten.

5.1.2 Architectuurprincipes Softwarecomponenten

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
(zie paragraaf 4.2.2 Architectuurprincipes Informatievoorziening) Daarnaast zal bij inrichting van de omgeving waar mogelijk voldaan worden aan SSD-eisen en -wensen, en aan andere vormen van "hardening"	

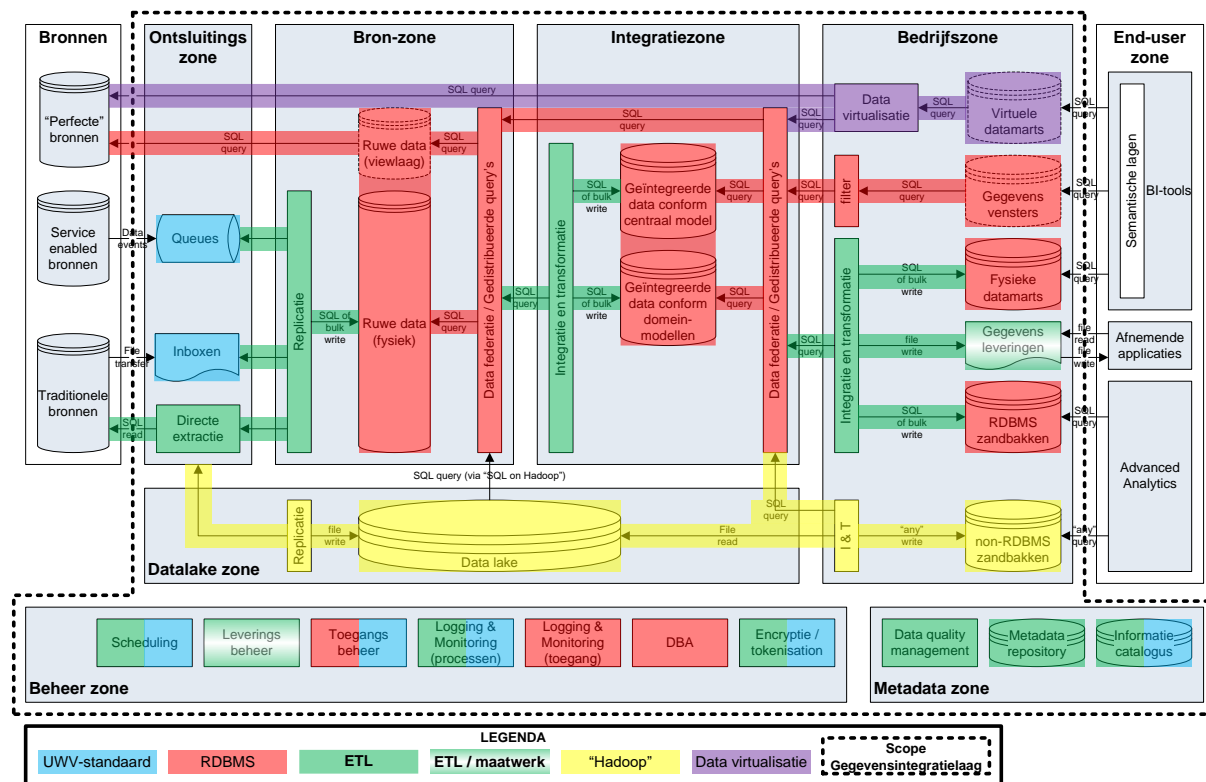
5.1.3 Afwijkingen van architectuurprincipes

Generiek architectuurprincipe	Aard van de afwijking	Reden voor de afwijking	Besluitvorming
Geen			

⁵⁶ De bredere inzet zelf (het gebruik in andere omgevingen dan het DIM) valt echter buiten de scope van het project DataFabriek, en is in deze PSA dan ook niet beschreven

5.1.4 Hoofdstructuur van de oplossing

Onderstaand plaatje beschrijft de inzet van (standaard)software voor het DIM:



Direct geraakte componenten

Component	Technologie	Effect
DIM database	Oracle RDBMS	Nieuwe instances op bestaande infrastructuur.
ETL-engine tier	IBM DataStage (ETL-tooling)	Nieuwe installatie op nieuwe infrastructuur
ETL-services tier	IBM DataStage (ETL-tooling)	Nieuwe installatie op nieuwe infrastructuur
ETL-repository	Oracle RDBMS	Onderdeel van DataStage. Nieuwe instance op bestaande infrastructuur.
ETL-shared storage (hot)	Gedeelde bestandsstructuren (bereikbaar vanuit zowel ETL-engine tier als vanuit DIM database, dit om ook ELT-patronen aan te kunnen).	Nieuwe opslagallocaties, toegekend aan ETL-engine tier

Component	Technologie	Effect
ETL-shared storage (cold)	Gedeelde bestandsstructuren (bereikbaar vanuit zowel ETL-engine tier als vanuit DIM database. Er is (nog) geen technische verschil tussen "hot" en "cold"; de scheiding is er vooral om bij herstart na Disaster Recovery niet te hoeven wachten op het terugrollen van het (in "cold" opgeslagen) archief. ⁵⁷	Nieuwe opslagallocaties, toegekend aan ETL-engine tier
Maskeringsengine	IBM InfoSphere Optim Data Privacy (maskeringsroutines)	Geïnstalleerd als library's binnen de nieuwe Oracle instances (zie: DIM database)
Toegangsbeheer	ABS / Active Directory	Aansluiting op bestaande UWV-oplossingen en -infrastructuur
Logging / monitoring	Standaard-inrichting (levering gegevens aan QRadar), bij DXC waar nodig uitgebreid met extra monitoring van beheer-acties m.b.v. PAM.	Aansluiting op bestaande UWV-oplossingen en -infrastructuur
Scheduler (keten)	Applicatie-overstijgende scheduling: IBM Workload Scheduler	Aansluiting op bestaande UWV-oplossingen en -infrastructuur
Scheduler (applicatie)	Interne DIM-scheduling en dependency management: IBM Workload Scheduler + IBM InfoSphere DataStage	Onderdeel van de mappings
Glossary	IBM Information Governance Catalog	Nieuwe installatie op nieuwe infrastructuur
Metadata-Bridge	InfoSphere Metadata Asset Manager	Nieuwe installatie op nieuwe infrastructuur
DK-engine	IBM Information Analyzer / QualityStage	Nieuwe installatie op nieuwe infrastructuur
Workflow-engine DK-processen	IBM Business Process Manager	Nieuwe installatie op nieuwe infrastructuur
Rapportage-engine meegeleverde rapportages	IBM Cognos ⁵⁸	Nieuwe installatie op nieuwe infrastructuur

⁵⁷ Mocht het DIM in de toekomst qua infrastructuur "zilver" moeten worden, dan hoeft dat niet voor de "cold" storage

⁵⁸ Cognos is integraal onderdeel van de verworven ETL+-software. Het zal alleen ingezet worden voor standaard-rapportages, zodat (a) de tool gezien kan worden als embedded, en (b) opbouw van tool-expertise binnen UWV niet vereist is.

Indirect geraakte componenten

Onderdeel / Component	Omschrijving	Effect
DWH 1, 2, 3	Legacy DWH's	Beperkte wijzigingen om voor het DIM verbeterde bronleveringen te kunnen blijven verwerken in de bestaande DWH's. Dit voorkomt onnodig dubbel draaien en dubbel beheer bij de bronnen. ⁵⁹
DMAP	Analyse-omgeving met DWH-lagen	Beperkte wijzigingen om (via een gegevensvenster) gegevens te kunnen inlezen uit de bronzone en (voor ResaFasa) potentieel ook de integratiezone van het DIM. Dit als vervanging van de huidige directe koppelingen tussen DMAP en de DTL-en MVL-lagen van DWH3.0 Verdere wijzigingen in DMAP vallen buiten scope van het project
Diverse decentrale analyse omgevingen	Analyse-omgevingen met deels eigen integratie en/of ETL-oplossingen	Beperkte wijzigingen om (via een gegevensvenster) gegevens te kunnen inlezen uit het DIM, of te verkrijgen middels een gegevenslevering. Dit als vervanging van de huidige leveringen en/of database links. Verdere wijzigingen in deze omgevingen (bijvoorbeeld vanwege standaardisatie en/of AVG-compliance) vallen buiten scope van het project

(Veranderende) Koppelingen:

Koppeling	Omschrijving	Effect
Bronlevering Excasso	Brede bronontsluiting van Excasso	Levering aan bronzijde robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface zowel gebruikt worden voor het DIM als voor de legacy DWH's.
Bronlevering GINA	Brede bronontsluiting van GINA	Levering aan bronzijde robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface zowel gebruikt worden voor het DIM als voor de legacy DWH's

⁵⁹ Zoals beschreven in Hergebruik DIM-interfaces in legacy DWH's (v1.0), en afgestemd met TDA DWH

Koppeling	Omschrijving	Effect
Bronlevering BAS	Brede bronontsluiting van BAS	Levering aan bronzijde robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface zowel gebruikt worden voor het DIM als voor de legacy DWH's
Bronlevering UZS	Brede bronontsluiting van UZS	Levering aan bronzijde robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface zowel gebruikt worden voor het DIM als voor de legacy DWH's
Bronlevering IMF	Brede bronontsluiting van IMF	Levering aan bronzijde robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface zowel gebruikt worden voor het DIM als voor de legacy DWH's
Bronlevering CBBS	Brede bronontsluiting van CBBS	Levering aan bronzijde robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface zowel gebruikt worden voor het DIM als voor de legacy DWH's
Bronlevering PEERCODE	Brede bronontsluiting van PEERCODE	Levering aan bronzijde robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface zowel gebruikt worden voor het DIM als voor de legacy DWH's
Bronlevering Peoplesoft HCM	Brede bronontsluiting van Peoplesoft HCM	Levering aan bronzijde robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface zowel gebruikt worden voor het DIM als voor de legacy DWH's
Bronlevering ResaFasa	Brede bronontsluiting van ResaFasa	Levering aan bronzijde robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface slechts gedeeltelijk gebruikt worden voor zowel het DIM als voor de legacy DWH's. De bestaande interface blijft dus (deels) in stand tot na uitfasering DWH2.
Bronlevering WWO	Brede bronontsluiting van WWO (via KDB WWO)	Nieuwe levering; wordt voortaan via de kopiedatabase van WWO geleverd. Bestaande levering blijft actief voor legacy.

Koppeling	Omschrijving	Effect
Bronlevering SMF	Brede bronontsluiting van het SMF-systeem-complex (via MIS DB ⁶⁰)	Levering aan bronzijde (w.o. ook de interfacing tussen SMF en MIS DB) robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface zowel gebruikt worden voor het DIM als voor de legacy DWH's.
Bronlevering Winter	Brede bronontsluiting van Winter	Levering aan bronzijde robuuster en traceerbaarder gemaakt. Inhoud beter gedefinieerd. Naar verwachting kan de gewijzigde interface zowel gebruikt worden voor het DIM als voor de legacy DWH's
Door DWH beheerde referentiegegevens	Reference data die nu, door de afdeling DWH, direct in DWH3 wordt ingevoerd of ingeladen.	Als de hierboven genoemde bronnen die reference data ook zelf kunnen leveren, dan wordt deze opgenomen in de "brede bronontsluiting" van die bron. In de overige gevallen wordt eigenaarschap en beheer-proces van de reference data formeel vastgelegd en wordt de resulterende dataset via standaard-programmatuur in het DIM ingelezen.
VIP-gegevens UPA	Levering van VIP-gegevens aan het DIM, t.b.v. filtering	Nieuwe interface, beperkt tot gegevens die noodzakelijk zijn voor VIP-filters. ⁶¹
SUAG-parameters UGC (service UGC01)	Service-interface t.b.v. ophalen levercontracten SUAG (als parameter voor de SUAG-generatie)	Wijziging nog te bepalen (lopende discussie over SOLL-oplossing voor SUAG)
SUAG-parameters POLIS LEVEREN (service WS GGD 21)	Service-interface t.b.v. ophalen contractdetails SUAG (deelnemers pensioenfondsen) (als parameter voor de SUAG-generatie)	Wijziging nog te bepalen (lopende discussie over SOLL-oplossing voor SUAG)
<i>Bovenstaande koppelingen zijn in scope voor fase 4/5 van het project (vervanging DWH 3.0); koppelingen voor latere fasen zullen worden opgenomen in een volgende versie van de PSA.</i>		

(meer informatie over over de verschillende typen koppelingen in Bijlage B : *Typen gegevensgebruik door DIM*)

In de context van het project Datafabriek zijn de bronsystemen van het DIM dezelfde als die voor de gezamenlijke uit te faseren DWH's. Alle koppelingen tussen deze bronsystemen en het DIM (de "brede bronontsluitingen"), zullen, t.o.v. de bestaande koppelingen met de uit te faseren DWH's, zowel technisch als functioneel, veranderen.

⁶⁰ Deels worden de SMF-gegevens door MIS DB alleen maar "doorgelust" naar het DIM.

⁶¹ Deze interface beperkt zich tot VIP-gegevens, en is randvoorwaardelijk voor VIP-filtering bij doorlevering van DIM-gegevens aan afnemers. In een latere fase van het project wordt UPA ook via een brede bronontsluiting aan het DIM gekoppeld.

Voor koppelingen van de bronsystemen naar de DWH's zijn de voornaamste veranderingen:

- Aanpassing aan nieuwe interface-standaarden:
 - Voldoende rijke interface-metadata⁶² en voldoende rigide besturing bij het aanmaken van die interface (bij de bron) om automatische bewaking van volledigheid en volgordeelijkheid mogelijk te maken;
 - Verwijdering eventuele afnemer-specifieke filters en/of afleidingen;
 - Waar relevant opname van additionele reference data;
 - Brede ontsluiting t.b.v. voorraadvorming⁶³.
- Herrotering naar DIM (i.p.v. naar de legacy DWH's), via SIP-FT, over SFTP
- Bij huidige non-compliance: aanpassen interface om te voldoen aan de standaards van UWV op het gebied van gegevenstransport⁶⁴

Waar mogelijk worden deze veranderingen geïmplementeerd als een wijziging in de bestaande koppeling, en wordt de via de koppeling geleverde bestanden via SIP-FT naar zowel de legacy DWH's als naar het DIM gestuurd. Na uitfasering van het ontvangende legacy DWH hoeft dan slechts de SIP-FT routing te worden aangepast; wijzigingen aan de bronzijde zijn niet nodig.

Waar de aanpassingen aan de legacy DWH's om een gewijzigde broninterface te kunnen verwerken niet mogelijk of niet zinlijk zijn is een nieuwe koppeling (overigens wel gebaseerd op de bestaande) noodzakelijk. De oude broninterface blijft dan ook draaien tot het ontvangende legacy DWH is uitgefaseerd.

Aanpassingen aan het technisch formaat van de koppelingen (m.u.v. het meeleveren van een pakbon) zijn i.h.a. niet nodig. Dit omdat de ETL-tooling van het DIM een grote verscheidenheid aan formaten ondersteunt.

Voor koppelingen met door DWH beheerde referentiegegevens geldt:

- Als de "reguliere" bronnen die reference data ook zelf kunnen leveren, dan wordt deze opgenomen in de "brede bronontsluiting" van die bron.
- In de overige gevallen wordt eigenaarschap en beheer-proces van de reference data formeel vastgelegd en wordt de resulterende dataset (waarschijnlijk een spreadsheet) via standaard-programmatuur in het DIM ingelezen. In geen geval wordt reference data direct (bijvoorbeeld door scherm invoer en/of scripts) in het DIM gewijzigd.
Meer informatie hierover in Bijlage B : *Typen gegevensgebruik door DIM*

Voor koppelingen van de legacy DWH's naar gegevensafnemers is de voornaamste verandering:

- Herrotering; nu geleverd door DIM (i.p.v. door de legacy DWH's)

⁶² Voorbeelden hiervan zijn: gebruikte filterparameterwaarden (bij de bron) bij het creëren van de interface, begin- en einddatum van de selectie (bij de bron), volgnummer van de interface, extractiemoment, bestandsstructuur, aantal regels in interface. Deze typen metadata worden i.h.a. middels een "pakbon" meegeleverd met de interface.

Daarnaast de gegevensdefinities van geleverde data en de eventuele tijdens extractie uit de bron uitgevoerde business rules. Deze typen metadata veranderen slechts langzaam, en kunnen dus middels documentatie met het DIM gedeeld worden.

⁶³ Het DIM volgt, voor interfaces en bronzone, een "beter mee verleggen dan om verleggen" benadering. Dit betekent dat een bron-entiteit bij voorkeur volledig ontsloten wordt, ook als er initieel maar een paar attributen vereist zijn. Dit concept heet in het huidige DWH "voorraadbeheer". De impact hiervan blijkt overigens beperkt; de meeste bestaande bronkoppelingen zijn reeds voldoende "breed" om voorraadvorming te ondersteunen.

⁶⁴ Beperkt zich i.h.a. tot transport via SI-FT (i.p.v. direct), en overgang van FTP naar SFTP..

Functionele karakteristieken integratievoorziening						
<i>Bron</i>	<i>Doel / context</i>	<i>Uitwissel patroon</i>	<i>Berichtgrootte</i>	<i>Frequentie</i>	<i>Bericht Samenstelling / Transformatie</i>	<i>Integratie voorziening</i>
Excasso	Brede bronontsluiting	<i>Bestand</i>	10 GB	1/week	NVT	SIP-FT
GINA	Brede bronontsluiting	<i>Bestand</i>	6 GB	1/week	NVT	SIP-FT
BAS	Brede bronontsluiting	<i>Bestand</i>	240 MB	1/week	NVT	SIP-FT
UZS	Brede bronontsluiting	<i>Bestand</i>	43 GB	1/week + 1/maand	NVT	SIP-FT
IMF	Brede bronontsluiting	<i>Bestand</i>	2 GB	1/week + 1/maand	NVT	SIP-FT
CBBS	Brede bronontsluiting	<i>Bestand</i>	40 GB	1/maand	NVT	SIP-FT
PEERCODE	Brede bronontsluiting	<i>Bestand</i>	5 MB	1/dag	NVT	SIP-FT
Peoplesoft HCM	Brede bronontsluiting	<i>Bestand</i>	2 MB	1/maand	NVT	SIP-FT
ResaFasa	Brede bronontsluiting	<i>Bestand</i>	10 MB (dag) 233 MB (week) 300 MB (maand)	1/dag + 1/week + 1/maand	NVT	SIP-FT
WWO	Brede bronontsluiting	<i>Bestand</i>	700 MB (max)	1/week	NVT	SIP-FT
SMF	Brede bronontsluiting	<i>Bestand</i>	2,5 GB	1/dag	NVT	SIP-FT
Winter	Brede bronontsluiting	<i>Bestand</i>	NTB	NTB	NVT	SIP-FT
GD/DWH	DWH reference data	NTB	NTB (maar klein)	1/jaar of lager	NVT	Geen ⁶⁵
UPA	VIP-gegevens	<i>Bestand</i>	< 1MB	1/dag	NVT	SIP-FT
UGC	SUAG-parameters	<i>Service</i>	Bestaande service UGC01			Geen ⁶⁶
POLIS LEVEREN	SUAG-parameters	<i>Service</i>	Bestaande service WS GGD 21			Geen ⁶⁷

Voor alle brede bronontsluitingen worden de legacy DWH's waar mogelijk/eenvoudig ook aangesloten op de DIM-versie van de interface. In een dergelijk geval wordt het bestand door de bron slechts één keer geleverd aan SIP-FT, en dan door SIP-FT doorgezet naar zowel het DIM als de legacy DWH's (ook wel de "Y-constructie" genoemd).

5.1.5 Check Licentie Positie (CLP)

De voor het project benodigde software is verworven middels de Europese Aanbesteding voor ETL+ tooling. Deze aanbesteding heeft geresulteerd in een raamcontract met IBM⁶⁸ voor de levering van ETL+ tooling en gerelateerde adviesdiensten. Binnen dat raamcontract zijn de voor het DIM benodigde licenties expliciet door de leverancier begroot, en door UWV afgenomen. Voor de meerderheid van de tooling zijn de licentiekosten gebaseerd op de voor de tools gebruikte rekenkracht (PVU's). De voor het DIM in het IBM-rekencentrum bestelde infrastructuur valt binnen de afgenomen licenties.⁶⁹ Voor twee tools (IBM Business Process Manager en IBM Cognos) geldt dat gebruik binnen de aangeboden licentie alleen is toegestaan voor expliciete use cases (BPM: datakwaliteitsprocessen, Cognos: beheer-rapportages)

5.2 Technische infrastructuur

5.2.1 Afbakening Technische infrastructuur

Duiding van de impact op de technologie

⁶⁵ Koppeling binnen afdeling DWH

⁶⁶ Servicekoppeling binnen Divisie Gegevensdiensten.

⁶⁷ Servicekoppeling binnen Divisie Gegevensdiensten.

⁶⁸ Hier als softwareleverancier, staat dus los van het aflopende rekencentrum-contract.

⁶⁹ Deze licenties kunnen worden "meegenomen" naar DXC, maar herberekening van de PVU's van de nieuwe infrastructuur is daarbij wel noodzakelijk.

Onderdeel / Component	Omschrijving	Effect
Besturingssysteem	Linux	OS voor de diverse ETL+-engines. Nieuwe installaties.
Besturingssysteem	Windows	OS voor de Metadata bridge server. Nieuwe installatie OS van de UCRA terminal server. Geen wijzingen vereist.
Besturingssysteem	AIX	OS van de bestaande DWH-servers.- Deze worden ook gebruikt voor de DIM-databases.
Hypervisor	<i>Conform rekencentrum standaards</i>	
Storage	<i>Conform rekencentrum standaards</i>	
Netwerk	<i>Conform rekencentrum standaards</i>	

Zie voor verdere details het document "UWV HLD DIM" (v1.0, 11 Maart 2020)

Illustratie van de netwerkzoning in de huidige situatie

Onderdeel / Component	Effect
Aparte zone(s) voor O, T, A en P van het DIM	<p>Het IBM Rekencentrum eist (in afwijking van de goedgekeurde HLD) dat de nieuwe O, T, A en P omgevingen van het DIM worden geïnstalleerd in separate ("fenced") VLAN's.</p> <p>Voor het DIM (of voor de toekomstige IV-organisatie DIM/DWH) is dit niet onoverkomelijk, maar voegt het ook geen waarde toe.</p> <p>ICT Services heeft aangegeven dat dergelijke fenced VLAN's niet conform beleid zijn, maar dat ze voor het DIM tijdelijk gedoogd worden om de project-voortgang niet te frustreren.</p>

5.2.2 Architectuurprincipes Technische infrastructuur

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
REQ-01 - Gebruik Standaard Technische Stacks	De in fase 2 van het project verworven ETL-tooling wordt op Standaard technische stacks geïnstalleerd.
REQ-02 - Gebruik de Standaard Technische Stacks voor Linux en Windows bij nieuw te verwerven producten	De in fase 2 van het project verworven ETL-tooling wordt op Standaard technische stacks geïnstalleerd.
REQ-03 - De applicatie is Web-based	Voor de in fase 2 van het project verworven ETL-tooling was een harde eis: "De ontwikkel- en beheerssoftware van de tool(set) kan, als package, via software distribution programmatuur worden uitgerold naar de clients, of vereist geen uitrol ("thin client")."
REQ-04 - De applicatie heeft geen afhankelijkheden van specifieke plug-ins (zoals ActiveX, Java o.i.d.)	<i>Ontwikkel- en beheertools hebben geen afhankelijkheden met specifieke plug-ins</i>

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
REQ-05 - De applicatieleverancier ondersteunt actieve ontwikkeling van front-end zodat toekomstige browsers ondersteund blijven/worden	De in fase 2 van het project verworven ETL-tooling is aangeschaft inclusief meerjarige support.
REQ-06 - Applicaties die bedrijfslogica bevatten worden in het HRC geplaatst en ontsloten via het KWN domein	Het DIM wordt in het HRC gehost. Enige, potentiële, uitzondering is de metadata-tooling, maar deze bevat geen bedrijfsgegevens of -logica (slechts de definities daarvan).
REQ-07 - Bij gelijke geschiktheid heeft open source software de voorkeur	De aanbesteding voor ETL-tooling stond open voor zowel proprietary als open source leveranciers. Helaas heeft geen enkele aanbieder van "enterprise ready" open source distributies (ook wel bekend als "distro's") ingeschreven op de Europese Aanbesteding.
REQ-08 - De applicatie kan gevirtualiseerd worden	Voor de in fase 2 van het project verworven ETL-tooling was dit een harde eis. ⁷⁰
REQ-09 - Monitoring en reporting is ingericht	<i>Wordt in fase 4/5 ingericht; dekking van dit principe zal in de volgende versie van de PSA worden beschreven.</i>
REQ-10 - Service Level monitoring is flexibel ingericht	<i>Wordt in fase 4/5 ingericht; dekking van dit principe zal in de volgende versie van de PSA worden beschreven.</i>
REQ-11 - Architectuur en samenhang is inzichtelijk	Het DIM wordt "on premise" gehost conform UWV-standaards
REQ-12 - De applicatie is Cluster aware	Voor de in fase 2 van het project verworven ETL-tooling was dit een harde eis.
REQ-13 - De Deployment van de applicatie is geautomatiseerd	Deployment van nieuwe/gewijzigde DIM-functionaliteit gebeurt grotendeels door promotie van metadata (en dus niet van, al dan niet gegenereerde, code). Dit promotieproces is geautomatiseerd.
REQ-14 - Splicing en recharging is mogelijk vanuit de applicatie	<i>Niet van toepassing</i>
REQ-15 - Onderhoudskaders en inzichtelijkheid	Het DIM wordt "on premise" gehost conform UWV-standaards
REQ-16 - Dataopslag binnen de EU	Het DIM wordt "on premise" gehost conform UWV-standaards
REQ-17 - Domain Name Service (DNS) wordt door UWV ingevuld	<i>Configuratie maakt gebruik van, door UWV gedefinieerde, logische namen.</i>
REQ-18 - De applicatie wordt ontsloten op de gangbare versie van het werkplek OS	Voor de in fase 2 van het project verworven ETL-tooling was uitrolbaarheid van de ontwikkel- en beheer-software op Windows 10 een harde eis.
REQ-19 - De applicatie wordt aangesloten op de directory services van UWV	Voor de in fase 2 van het project verworven ETL-tooling was toegangsbeheer m.b.v. OAM een harde eis.
REQ-20 - De applicatie maakt gebruik van de horizontale schaalbaarheid die het (Cloud) management platform biedt	Voor de in fase 2 van het project verworven ETL-tooling was dit een harde eis.

⁷⁰ Om de software te kunnen licenseren o.b.v. van virtuele PVU's (in plaats van o.b.v. totale grootte onderliggende infrastructuur) dient de IBM License Metric Tool geïnstalleerd te zijn.

5.2.3 Afwijkingen van architectuurprincipes

Generiek architectuurprincipe	Aard van de afwijking	Reden voor de afwijking	Besluitvorming
Geen (voor zover nu bekend). Zie echter wel opmerkingen over fenced VLAN in paragraaf 5.2.1.			

5.2.4 Hoofdstructuur van de oplossing

Beschrijving van de gewenste situatie :

Zie document "UWV HLD DIM, versie 1.0"⁷¹

Beschrijving in termen van technologie/platform gebruikt in de bouw en productie van de applicaties:

Onderdeel / Component	Technologie	Effect
Zie 5.1.4		

Bij nieuwbouw: Duiding van de impact op de technologie

Onderdeel / Component	Omschrijving	Effect
Zie 5.2.1 en "UWV HLD DIM"		

Illustratie van de netwerkzoning in de gewenste situatie

Onderdeel / Component	Effect
<i>Zal worden beschreven in nieuwe, door IBM bij te werken, versie van "UWV HLD DIM"</i>	

Verwachte requirements:

Onderdeel	Omschrijving	Metriecken
Aantal gebruikers	Voorzien totaal aantal gebruikers van de applicatie	800 eindgebruikers ⁷² N.B. Dit zijn de gebruikers van het DIM (al dan niet via BI-tools) en/of de metadata-tooling. De ETL-tooling zal alleen gebruikt worden door ontwikkelaars en beheerders. Dat zullen er ca. 30 zijn. ⁷³
Aantal gebruikers per locatie	Voorzien aantal gebruikers uitgesplitst naar (netwerktechnische) locatie	850 gebruikers UWV kantoren ⁷⁴ 0 gebruikers internet 0 gebruikers andere locaties
Aard gebruik	Verwachte aard van het gebruik van de applicatie	Interactief (lezen) & Batch (lezen schrijven)

⁷¹ In een, door IBM nog op te leveren, bijgewerkte versie van deze HLD zullen ook de fenced VLAN's (zie 5.2.1) worden beschreven. Deze worden beschouwd als een "noodzakelijk kwaad" voor hosting bij IBM, en zijn bij hosting door DXC dus hoogstwaarschijnlijk niet langer noodzakelijk.

⁷² Vergelijkbaar met aantal gebruikers huidige DWH's/analyseomgevingen

⁷³ Vergelijkbaar met aantal ontwikkelaars/beheerders huidige DWH's/analyseomgevingen

⁷⁴ Inclusief thuiswerkende UWV-medewerkers.

Onderdeel	Omschrijving	Metrieken
Wijze van ontsluiten	Applicatie ontsluiting en compatibiliteit met KWN werkplek	SQL naar bestaande BI- en analyse-tooling
Basislaag compatibiliteit	Compatibiliteit met KWN werkplek basislaag	<i>Nog te bepalen (afhankelijk van geselecteerde tooling)</i>
Hoeveelheid gebruik	Verwachte hoeveelheid van het gebruik van de applicatie	Inschatting is 100 concurrent eindgebruikers. ⁷⁵
Authenticatie	Wijze van authenticatie	<i>Nog te bepalen (afhankelijk van geselecteerde tooling)</i>
Plaatsing	Specificatie van de plaatsing van de functionele componenten in UWV infrastructuur	<i>Grotendeels IBM Data Center. Ontwikkel- en beheer-tools op UCRA.</i>
Netwerkbelasting	Voor zover bekend: de verwachte belasting van het netwerk	De netwerkbelasting piekt buiten kantooruren (21:00u - 07:00 uur). Throughput dan 2 TB/dag, met maandelijkse pieken tot 5 TB/dag. ⁷⁶
Openingstijden	Gewenste gegarandeerde openingstijden	7x24
Beschikbaarheid	Gewenste gegarandeerde beschikbaarheid binnen de openingstijden	98% ⁷⁷
Ontwikkel en testvoorzieningen	Specificatie van ontwikkel- en testomgevingen	Volledige OTAP-straat; infra-topologie identiek voor alle omgevingen ⁷⁸ Infrastructuur-sizing identiek voor A en P. Zie "UWV HLD DIM" voor details.

⁷⁵ Vergelijkbaar met aantal concurrent gebruikers huidige DWH's/analyseomgevingen

⁷⁶ De netwerkbelasting wordt grotendeels veroorzaakt door het transporteren van brongegevens naar het DIM. De throughput is vergelijkbaar met die van de huidige DWH's/analyseomgevingen. Wel wordt een korter batch window aangehouden. Dit vanwege hogere de hogere tijdigheidsambities van het DIM.

⁷⁷ Het betreft hier de beschikbaarheid van de productieomgeving. Geplande downtime (mits buiten kantoor tijden, en zonder impact op het dagelijks batch window) geen probleem.

⁷⁸ M.u.v. de databases; daar draaien de O- en T-omgevingen op dezelfde infrastructuur, omdat dit voor de legacy DWH's ook zo is. Bij migratie naar DXC kan dit worden aangepast, zodat de O- en T-omgevingen van de legacy DWH's als die van het DIM op dedicated O- en T-infrastructuur draaien.

Onderdeel	Omschrijving	Metrieken
Overige niet functionele requirements	Eventuele andere gewenste niet-functionele requirements	<p>Continuïteit is één van de voornaamste drivers van het project.</p> <p>Schaalbaarheid is ook een vereiste. Initieel, qua virtualisatiepatroon, verticaal, daarna horizontaal.</p> <p>Tenslotte worden de eisen op het gebied van tijdigheid/actualiteit steeds hoger. Voor het DIM wordt ingezet op "einde dag in de bron is de volgende dag in het DIM". Dit heeft zowel impact op de sizing als op de inrichting van de laadprocessen (optimale parallelisering).</p>

6 INFORMATIEBEVEILIGING EN -BEHEER

6.1 Informatiebeveiliging

6.1.1 Principes Informatiebeveiliging

In deze paragraaf worden de binnen het verandergebied geldende principes gepresenteerd die betrekking hebben op informatiebeveiliging. In de tabel is voor alle geldende principes binnen dit aandachtsgebied gespecificeerd of deze voor het project relevant zijn.

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
P1 - Beschikbaarheid : Bij het doorvoeren van veranderingen in het IT landschap mogen er geen beveiligingsrisico's ontstaan die de continuïteit van de dienstverlening bedreigen.	<p>1) Het DIM heeft, inherent aan zijn rol, een beperkt voorspelbare belasting, en daarmee ook een beperkt voorspelbare responstijd.</p> <p>2) De infrastructuur zal zo worden ingericht dat (piek)belastingen op het DIM geen invloed hebben op de responstijd en/of beschikbaarheid van de overige UWV-applicaties.</p> <p>3) Het DIM zal geen onderdeel zijn van ketens waarbinnen de onvoorspelbare responstijd van dat DIM onacceptabel is.</p>
P2 - Integriteit : Er kan worden aangetoond dat acties of gebeurtenissen in het IT landschap hebben plaatsgevonden, dat data/informatie blijvend aan betrouwbaarheidsnormen voldoet.	<p>Integriteit tussen brongegevens en de replica daarvan in de bron-zone van het DIM:</p> <ul style="list-style-type: none"> • Wordt deels bewaakt door zelf-reparerende processen ("target delta"), deels door dataredundantie (checksums en vergelijkbaar) binnen een gegevenstransport ("interface") van bron naar DIM, en deels door periodieke reconciliatie van bron en DIM. • Wordt daarnaast gegarandeerd door gegevens in het DIM alleen te wijzigen op instigatie van de bron. • Additioneel wordt de volledigheid en volgorde van de interface bewaakt door controles op de metadata (bv. gebruikt fileformaat, gebruikte filter-parameters) tijdens het laden. <p>Integriteit tussen de zones van het DIM onderling:</p> <ul style="list-style-type: none"> • Wordt grotendeels gegarandeerd door rigide testprocessen, en door gegevens in een DIM-zone alleen te wijzigen op instigatie van een eerder in de DIM-keten liggende zone. • Kan bij "critical data elements" extra worden bewaakt middels zone-naar-zone vergelijkingen o.b.v. Datakwaliteitsmetingen

Generiek Architectuurprincipe	Specifieke betekenis of toepassingswijze in dit project.
P3 - Vertrouwelijkheid : de toegang tot informatie in systemen alleen mogelijk is met de juiste autorisatie.	Het DIM voldoet aan de UWV-standaards op het gebied van autorisatie en authenticatie.
P4 - Controleerbaarheid : de mate van beschikbaarheid, integriteit en vertrouwelijkheid van de informatievoorziening moet objectief vastgesteld kunnen worden.	Zie onder P2. Daarnaast worden alle gegevensverwerkende processen gelogd, en wordt de data zelf ook gemarkeerd met verwerkings-timestamps.

Onderstaande principes zijn afkomstig uit een set concept-aanpassingen aan het PSA-sjabloon gerelateerd aan AVG-compliance. Hoewel formeel nog zonder status toch in deze PSA opgenomen, gezien het cruciale belang van AVG-compliance bij de ontwikkeling van de lijnorganisatie Datafabriek en het DIM.

Generiek Architectuurprincipe (Privacy aspecten)	Specifieke betekenis of toepassingswijze in dit project.
<p>P1: Doelbinding Uitdrukkelijk omschreven en Gerechtvaardigd</p> <p>Persoonsgegevens worden slechts verzameld en verwerkt voor vooraf welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doeleinden en worden niet verder verwerkt voor andere doelen die hiermee onverenigbaar zijn.</p>	<p>Het DIM bevat, in principe, alleen gegevens die óf ook in de bron aanwezig zijn (en waarvoor die bron doelbinding en rechtsgrond dient te hebben), óf die in de bron aanwezig waren, maar waarvoor UWV een wettelijke grondslag heeft voor een langere bewaartermijn. In dit tweede geval zijn de gegevens gepseudonimiseerd.</p> <p>N.B. In beide gevallen geldt dat incompliance bij de bron (bijvoorbeeld te veel of te lang bewaarde gegevens) automatisch leidt tot incompliance in het DIM. Het DIM geeft immers de administratieve werkelijkheid van UWV weer, óók als die werkelijkheid incompliant is.</p> <p>De IV-organisatie DIM/DWH levert vanuit het DIM alleen gegevens waarvoor de afnemer rechtsgrond, proportionaliteit en subsidiariteit kan aantonen.</p> <p>Alle directe toegang tot het DIM loopt via gegevensvensters, die toegang bieden tot een nauw omschreven subset van de gegevens in het DIM. Voor deze toegang is weer doelbinding vereist.</p>
<p>P2: Doelbinding Wettelijke grondslag</p> <p>De verzameling en verwerking van persoonsgegevens is toegespitst op een gespecificeerd doel met een wettelijke grondslag. De persoonsgegevens zijn daartoe toereikend, ter zake dienend en beperkt tot wat noodzakelijk is voor de doeleinden waarvoor zij worden verwerkt.</p>	Idem

Generiek Architectuurprincipe (Privacy aspecten)	Specifieke betekenis of toepassingswijze in dit project.
<p>P3: Bewaren Persoonsgegevens worden bewaard in een vorm die het mogelijk maakt de betrokkenen niet langer te identificeren dan voor de doeleinden waarvoor de persoonsgegevens worden verwerkt noodzakelijk is.</p>	<p>Zodra de operationele bewaartermijn verstreken is (en de gegevens dus uit de bron verwijderd worden) bestaat in het DIM alleen nog een gemaskeerde replica t.b.v. rapportage/analyse. Deze gemaskeerde replica wordt verwijderd zodra de bewaartermijn voor rapportage- en analysegegevens is verstreken.⁷⁹</p>
<p>P4: Beveiliging Passende technische en organisatorische maatregelen, zoals anonimisering of pseudonimisering van persoonsgegevens, zijn op een dusdanige manier genomen dat een passende beveiliging van de verwerking en de persoonsgegevens gewaarborgd is en dat zij onder meer beschermd zijn tegen ongeoorloofde of onrechtmatige verwerking en tegen onopzettelijk verlies, vernietiging of beschadiging.</p>	<p>Anonimisering/pseudonimisering gebeurt in het DIM op database-niveau (en niet in eindgebruikers-tools), zodat alternatieve toegangsmethoden (b.v. maatwerk-SQL) geen beveiligingsgat opleveren.</p> <p>Ongemaskeerde persoonsgegevens worden in een separaat deel van het DIM opgeslagen, zodat de toegang ertoe zo eenvoudig mogelijk kan worden afgeschermd.</p> <p>Autorisatie op de gegevens in het DIM wordt, binnen de beperkingen van de bij UWV gebruikte BI-omgevingen, zoveel mogelijk op database-niveau beheerd. Waar dit niet het geval is zijn de autorisatieprocessen en –oplossingen binnen de BI-omgevingen integraal onderdeel van het totale autorisatieproces.</p>
<p>P5: Doorgifte naar derden Persoonsgegevens worden slechts doorgegeven wanneer formeel afdoende garanties zijn vastgelegd, zodat aangetoond kan worden dat ook bij de doorgifte aan de AVG wordt voldaan.</p>	<p>De lijnorganisatie Datafabriek levert alleen gegevens door (intern of extern) waarvoor de afnemer⁸⁰ rechtsgrond, proportionaliteit en subsidiariteit kan aantonen. Dit geldt ook bij levering aan derden.</p>
<p>P6: Kwaliteit De persoonsgegevens en de verwerking ervan voldoen aan vooraf vastgestelde kwaliteitseisen, zodat ze juist zijn en zo nodig worden geactualiseerd en waarbij alle redelijke maatregelen moeten zijn genomen om de persoonsgegevens, die gelet op de doeleinden waarvoor zij worden verwerkt onjuist zijn, onverwijld te wissen of te rectificeren.</p>	<p>Het DIM bevat een replica van de bron. Eventuele gegevensreparaties in die bron zullen dus (typisch met een dag vertraging) ook in het DIM worden doorgevoerd.</p> <p>N.B. Het eerder bestaan van foutieve informatie is een “administratieve waarheid”, en zal dus in het DIM beschikbaar blijven totdat de bewaartermijn voor die informatie verstreken is.</p>

⁷⁹ Geanonimiseerde en anderszins afgeleide data zal, conform de Archiefwet, periodiek geschoond worden.

⁸⁰ Vanuit de IV-organisatie DIM/DWH gezien is de afnemer overigens altijd een interne partij (bv. Gegevensdiensten Leveren, FEZ). Deze interne partij draagt weer zorg voor interactie (ook qua regelgeving) met de externe partij.

Generiek Architectuurprincipe (Privacy aspecten)	Specifieke betekenis of toepassingswijze in dit project.
<p>P7: Verantwoording (accountability)</p> <p>Verantwoordelijken hebben aantoonbaar maatregelen genomen, zodat de wijze van verwerken ten aanzien van de betrokkene behoorlijk is, waarbij de risico's zijn geëlimineerd of gemitigeerd door het toepassen van Privacy by Design, het uitvoeren van gegevensbeschermingseffect beoordelingen (GEB's) en het gebruik van standaardinstellingen.</p>	<p>Binnen het project Datafabriek is reeds een GEB uitgevoerd, en zullen er, per projectfase en/of nieuw aangesloten groep bronnen, nog een aantal worden uitgevoerd.</p> <p>Verder geldt voor het gehele DIM-ontwerp "Privacy by Design".</p> <p>In het DIM:</p> <ul style="list-style-type: none"> • Worden alle niet-identificerende gegevens gescheiden opgeslagen van, al dan niet geanonimiseerde, persoonsgegevens. • Kunnen deze gegevens zowel worden gekoppeld aan persoonsgegevens als aan geanonimiseerde persoonsgegevens. Dit laatste is de default-optie ("privacy by design"), en bij verwijdering van de ongemaskeerde gegevens (zie hierboven) zelfs de enige optie.
<p>P8: Recht op transparantie</p> <p>De wijze waarop de persoonsgegevens worden verwerkt is voor het publiek en de betrokkene transparant en maakt het de betrokkene mogelijk zijn rechten uit te oefenen. Hierbij is specifiek aandacht voor de bescherming van kinderen. Bij een inbreuk in verband met persoonsgegevens (datalek, 'personal data breach') worden de betrokkenen en de AP geïnformeerd als deze inbreuk waarschijnlijk een risico inhoudt voor de rechten en / of vrijheden van natuurlijke personen.</p>	<p>Opslag binnen het DIM heeft altijd een wettelijke grondslag.</p> <p>Alle doorlevering (door de IIV-organisatie DIM/DWH) van gegevens uit het DIM moet binnen de grenzen vallen die de bron daaraan stelt. Het DIM bevat voldoende metadata om de bron van een gegeven én de eisen van die bron altijd te kunnen bepalen.</p> <p><i>Nog te bepalen (uitstaande vraag bij IM A&A Gegevensdiensten, raakt m.n. het lijn-implementatietraject "GD Uitvoering en IV"):</i></p> <p><i>Moet elke losse levering/gegevensvenster worden opgenomen in het verwerkingsregister, of voldoet een algemenere wijze van documenteren?</i></p>

Naast bovenstaande worden SSD-beveiligingseisen bij de inrichting van de ETL-tooling en de DIM-databases zoveel mogelijk meegenomen.

6.1.2 Afwijkingen van principes

Generiek architectuurprincipe	Aard van de afwijking	Reden voor de afwijking	Restrisico	Besluitvorming
-- geen --				

6.2 Informatiebeheer

Geen verdere principes

7 ARCHITECTUURCONFLICTEN EN ONZEKERHEDEN

7.1 Architectuurconflicten

<Naam architectuurconflict>	
Typering conflict	<Welk afbakeningsconflict is geconstateerd? Wat is er aan de hand?>
Betrokken partij	<Geef aan met welke partijen afstemming nodig is over het conflict.>
Advies	<Geef een advies hoe binnen het project omgegaan moet worden met het geconstateerde conflict.>
Status	<Vermeld de naam van beslissingsbevoegde gremium/instantie (Stuurgroep, Architectuurplatform, IMU of SIVU) en actuele stand van zaken besluitvorming (datum bespreking, datum goedkeuring/afwijzing).>

7.2 Architectuur aannames

Opsomming aannames:

- 01 < beschrijving aanname">
- 02 < beschrijving aanname">
- 03

7.3 Openstaande punten

Openstaande punten:

- 01 < beschrijving openstaand punt">
- 02 < beschrijving openstaand punt">
- 03

8 METHODEN, TECHNIEKEN EN HULPMIDDELEN

Binnen de organisatie gelden de kaderstellingen voor methoden, technieken en hulpmiddelen zoals vastgelegd op de intranetpagina's onder 'Home > Service > Standaarden bedrijfs- en ICT-verandertrajecten'.

Vanuit deze UWV kaderstelling worden voor dit project de volgende specifieke keuzes gemaakt voor toe te passen methoden, technieken en hulpmiddelen.

Toepassingsgebied: Architectuur

Methode	:	Dynamische Architectuur (DYA), Infrastructurele benadering
Techniek	:	ArchiMate, Modellenreeks UWV, UWV Classificatiemodel
Hulpmiddel	:	BiZZdesign Architect

Toepassingsgebied: Projectmanagement

Methode	:	UWV Projectmanagement Methode (UPM)
Techniek	:	nvt
Hulpmiddel	:	MS Office

Toepassingsgebied: Releasemanagement

Methode	:	Releasematig werken (cf. procesplaten bedrijfs- en IV-veranderproces)
Techniek	:	nvt
Hulpmiddel	:	MS Office, UWV Servicekalender

Toepassingsgebied: Integraal business ontwerp

Methode	:	Handboek UWV Business Consultancy (UBC)
Techniek	:	nvt
Hulpmiddel	:	MS Office

Toepassingsgebied: Procesontwerp

Methode	:	Handboek Business Process Engineering (BPE)
Techniek	:	UWV Proces Modelleer Conventies (UPMC)
Hulpmiddel	:	BiZZdesigner

Toepassingsgebied: Functioneel ontwerp

Methode	:	Functioneel Ontwerp voor nieuwbouwprojecten (VBP BS)
Techniek	:	Use case modellering, Entiteit Relatie Diagram modellering
Hulpmiddel	:	MS Office, MS Visio, Powerdesigner, IBM Data Architect (IDA), IBM Information Governance Catalog (IGC) ⁸¹

Toepassingsgebied: Functioneel beheer

Methode	:	Business Information Services Library (BiSL®)
Techniek	:	Best practices voor functioneel beheer
Hulpmiddel	:	MS Office, MS Visio, Powerdesigner IBM Data Architect (IDA), IBM Information Governance Catalog (IGC)

Toepassingsgebied: Technisch beheer

Methode	:	Information Technology Infrastructure Library (ITIL®)
Techniek	:	Best practices voor technisch beheer
Hulpmiddel	:	MS Office, MS Visio

⁸¹ De inzet van hulpmiddelen wordt in fase 4 van het project DataFabriek verder verbijzonderd

BIJLAGE A : RATIONALE BREDE BRONONTSLUITING

Bij het definiëren van de gegevens-scope van een bronontsluiting (de “breedte”) zijn drie benaderingen mogelijk:

	Scope	Aanpassingen als
Smal	De bronontsluiting bevat alleen de brongegevens waaraan een expliciete behoefte is (bij de gegevensafnemers van het DIM).	De bronontsluiting wordt aangepast als nieuwe informatiebehoeften additionele brongegevens vereisen.
Breed	De bronontsluiting bevat alleen de brongegevens waaraan een verwachte behoefte is (bij de gegevens afnemers van het DIM).	De bronontsluiting wordt aangepast als nieuwe informatiebehoeften onverwacht additionele brongegevens vereisen, óf als wijzigingen in de bron resulteren in extra gegevens met een verwachte behoefte.
Volledig	De bronontsluiting bevat de volledige inhoud van het bronsysteem.	De bronontsluiting wordt bij elke wijziging van de bron aangepast.

Deze drie benaderingen hebben, vanuit beheer-perspectief, de volgende voor- en nadelen:

	Voordelen	Nadelen
Smal	<ul style="list-style-type: none"> ➤ Minimale dubbele opslag ➤ Geen impact op DIM als bron wijzigt voor (nog) niet door afnemers gebruikte gegevens ➤ Expliciete rechtsgrond voor van de replicatie naar (en de opslag in) het DIM 	<ul style="list-style-type: none"> ➤ Lange time-to-market voor informatieproducten die nieuwe brongegevens vereisen ➤ Met terugwerkende kracht laden van de historie van nieuw vereiste gegevens meestal niet mogelijk ➤ Veel wijzigingsverzoeken bij bronnen, vaak met hoge urgentie
Breed	<ul style="list-style-type: none"> ➤ Alleen lange time-to-market voor informatieproducten die onverwacht nieuwe brongegevens vereisen. ➤ Historie van nieuw vereiste gegevens vaak al beschikbaar 	<ul style="list-style-type: none"> ➤ Brede rechtsgrond voor de replicatie naar (en de opslag in) het DIM noodzakelijk.
Volledig	<ul style="list-style-type: none"> ➤ Brongegevens reeds beschikbaar voor alle nieuwe informatieproducten 	<ul style="list-style-type: none"> ➤ Volledige dubbele opslag ➤ Doelbinding replicatie complex ➤ Grote investering vooraf, zowel bij bron als bij DIM ➤ Elke bronwijziging heeft impact op het DIM.

Vanuit project-perspectief gelden daarnaast de volgende voor- en nadelen:

	Voordelen	Nadelen
Smal	Alleen migratie van historische gegevens naar het DIM als deze gegevens nu al in een informatieproduct worden gebruikt.	Scope bestaande bronleveringen (naar de legacy DWH's) moet herijkt worden; niet in een informatieproduct gebruikte gegevens moeten uit de interface worden verwijderd.
Breed	Geen herijking van bronleveringen; de scope van de bestaande leveringen blijft in grote lijnen ongewijzigd. Waar verbreding van een levering toch noodzakelijk blijkt is de impact hiervan beperkt; kan relatief eenvoudig worden meegenomen met de overige vereiste wijzigingen aan die levering.	Vereist aanpassingen aan een deel van de bestaande interfaces.
Volledig	geen	Vereist grote aanpassingen aan alle bronleveringen. Vertraagd daarmee de voortgang van het project, zonder dat hier duidelijke bedrijfswaarde tegenover staat.

Voor het DIM is daarom, zowel binnen het project als voor de langere termijn, gekozen voor brede bronontsluiting:

- Brede bronontsluiting is een "market best practice" die in DWH-omgevingen (zowel binnen als buiten UWV) algemeen wordt toegepast;
- Impact ervan op het Datafabriek-project (qua doorlooptijd en risico) is beperkt, aangezien de bronontsluitingen toch al moeten worden aangepast t.b.v. robuustheid/compliance.
- Om deze impact nog verder te beperken is in de interface-standaarden gekozen voor de benadering "Gegevensobjecten o.b.v. behoefte, gegevensattributen o.b.v. toekomstige behoefte".
Daarmee wordt bedoeld dat **binnen het project** brede ontsluiting alleen geldt voor nu reeds (aan de "legacy" DWH's) geleverde gegevensobjecten; nu nog niet ontsloten gegevensobjecten worden dus pas opgenomen in een bron-interface als een nieuw DIM-informatieproduct deze, buiten het project om, vereist.
- In het algemeen zal het ontwikkelen van een dergelijk nieuw DIM-informatieproduct, en dus ook de eventuele daarvoor noodzakelijke verbreding van een bronontsluiting buiten scope vallen van het project Datafabriek, en ook niet randvoorwaardelijk zijn voor de voortgang van dat project. Het zal echter wel "business as usual" zijn voor de nieuwe IV-organisatie DIM/DWH.

BIJLAGE B : TYPEN GEGEVENSGEBRUIK DOOR DIM

Deze bijlage is gebaseerd op de presentatie [Typen gegevensgebruik door DIM \(1.1\)](#).

Het DIM gebruikt gegevens uit andere systemen voor drie verschillende doelen:

- als **basis** voor de informatiegebieden (combinaties van, al dan niet voorbewerkte, gegevens) waarop de informatieproducten (datamarts, leveringen, etc) zijn gebaseerd
- t.b.v. **verdere structurering** binnen die geleverde informatieproducten
- als **parameters** bij het creëren van die informatieproducten
- als input voor de **beveiliging** van informatieproducten

Elk van deze gebruiksdoelen stelt eigen eisen aan de wijze waarop de gegevens aan het DIM ter beschikking worden gesteld.

Gegevens als basis – brede bronontsluiting

Het overgrote deel van de gegevens in het DIM wordt gebruikt als basis voor informatieproducten. Deze gegevens worden door de bron aangeleverd middels een gestandaardiseerde bronlevering.

De regels hiervoor zijn beschreven in het document **Interface-standaarden (bron-DIM)**.

De meeste van deze bronontsluitingen zijn “brede bronontsluitingen”, ze dekken alle brongegevens waarvoor een bestaande óf een verwachte behoefte bestaat bij één of meer van de DIM-afnemers (zie **Fout! Verwijzingsbron niet gevonden.**).

Smal	Breed	Volledig
De bronlevering bevat alleen de brongegevens waaraan een expliciete behoefte is (bij de gegevensafnemers).	De bronlevering bevat alleen de brongegevens waaraan een verwachte behoefte is (bij de gegevensafnemers).	De bronlevering bevat de volledige inhoud van het bronsysteem.

De term “breed” kan voor de meeste bronnen daarbij gelezen worden als:

“alle objecten met een bestaande behoefte, en daarbinnen alle attributen (ook die zonder bestaande behoefte), tenzij deze attributen geen relevantie hebben buiten het bronsysteem”.

Voor een beperkt aantal bronnen wordt, al dan niet tijdelijk, slechts een smalle bronontsluiting geïmplementeerd.

Dit is alleen toegestaan als brede bronuitsluiting op compliance-problemen stuit (bijvoorbeeld omdat de gegevens eigendom zijn van een partij buiten UWV ligt, of omdat het HRM-data betreft), of als het direct implementeren van een brede bronontsluiting leidt tot onoverkomelijke planningsissues bij de bron.

N.B. Het doorleveren van persoonsgegevens uit de bron aan het DIM (om het daar ook op te slaan) is an sich geen AVG-issue; het enige dat er immers gebeurt is dat de dezelfde data nu op twee plekken staat. Voor de AVG hoeft de levering dus niet versmald te worden. Daarnaast is levering van een gegeven aan het DIM géén vrijbrief voor het doorleveren van die gegevens aan de afnemers van dat DIM: die moeten daarvoor nog steeds aantonen dat het gebruik van de gegevens rechtsgrond heeft, en voldoet aan de AVG-principes van proportionaliteit en subsidiariteit.

Belang en rol RLO

Voor gegevens die als basis worden gebruikt, is nauwkeurige metadata van groot belang:

- De gegevens in het DIM moeten gebruikt kunnen worden bij het maken van informatieproducten (door de Datafabriek en/of, bij self service, door de afnemers van het DIM) zonder dat hiervoor navraag nodig is bij de bronnen, en, vooral, zonder dat hiervoor aannames t.a.v. de betekenis moeten worden gedaan.
- Eenduidige gegevensdefinities en helder gedefinieerde vertrouwelijkheid (en andere gebruiksbeperkingen) zijn een randvoorwaarde voor compliance; als de betekenis van een gegeven niet duidelijk is kan het vereiste beveiligings- en gebruiksregime (doelbinding, proportionaliteit, etc.) niet bepaald worden. Ook het maskeren van identificerende attributen vereist eenduidige gegevensdefinities.
- Historieopbouw van de brongegevens (in de DIM-bronzone) kan alleen betrouwbaar worden uitgevoerd als het historisch gedrag van die gegevens (in de bron) bekend is.

Al deze metadata wordt (door de bron-beheerders en/of eigenaars) vastgelegd in een RLO (Record LayOut). Deze RLO is de basis voor de inrichting van de verwerking van de brongegevens binnen het DIM. De RLO is daarmee ook integraal onderdeel van de leverafspraken tussen bron en DIM (zie verder **Interface-standaarden (bron-DIM)**).

N.B. Als de bron beschikt over een actueel FUGEM, dan vereenvoudigt dit het opstellen van een RLO aanzienlijk.

Geen weigering aan de poort

Bronleveringen worden alleen, in hun geheel, afgewezen (en dus niet verwerkt in het DIM) als de levering (of de inhoud ervan) technisch niet verwerkbaar is.

Gegevens met kwaliteitsissues (lege verplichte velden, loshangende verwijssleutels, etc) worden gewoon geladen. Die lage kwaliteit is immers de administratieve werkelijkheid van de bron.

N.B. Kwaliteitsissues die de verwerkbaarheid raken (lege sleutelvelden of dubbele primaire sleutels, bijvoorbeeld), resulteren het afwijzen van de volledige levering.

Zeer gevoelige gegevens

Binnen het DIM en de daaraan gerelateerde vraagsturings- en beheerprocessen bestaat expliciete aandacht voor privacy-risico's.

Voor gegevens met vertrouwelijkheidsklasse 3 (VK3) gelden een aantal extra regels:

- In principe laadt het DIM géén VK3-gegevens.
- Als er een zéér zwaarwegend belang is om zulks toch te doen (dit ter beoordeling van bron én afnemer, en vergezeld van een GEB⁸²), dan kunnen VK3-gegevens toch binnen het DIM worden verwerkt. Verwerking zelf is hierbij "DIM-standaard", maar opslag van de gegevens, en autorisatie voor de toegang daartoe, is gescheiden van die voor de overige gegevens.
- Een bijzonderheid zijn gegevens die alleen VK3 hebben voor bijzondere personen/zaken, of bijzondere waarden. Deze worden in het algemeen wél ingelezen in het DIM, maar alleen voor zeer specifiek toegestaan gebruik ter beschikking gesteld aan afnemers. Zie volgende twee paragrafen voor meer details.

Zeer gevoelige personen/zaken

In een aantal gevallen worden gegevens pas gevoelig als ze betrekking hebben op een specifieke persoon of zaak:

⁸² GegevensbeschermingsEffectBeoordeling (Engels: Privacy Impact Assessment; PIA)

- **VIP's**

Als iemand de VIP-status krijgt (of zijn VIP-status verliest), dan geldt dat met terugwerkende kracht.

Daarnaast moeten gegevens van VIP's wel worden opgenomen in diverse totaalstellingen t.b.v. Sturen en Verantwoorden (S&V).

Vandaar dat VIP-gegevens wél aan het DIM moeten worden geleverd, ook al zijn ze VK3. Binnen het DIM worden deze gegevens vervolgens afgeschermd van de overige gegevens door toegang ertoe altijd te valideren tegen de laatste versie van de "VIP-lijst" uit UPA.⁸³

- **Bijzondere Zaken (BZ), Bijzondere GevalsBehandeling (BGB), EP+ (Eigen Personeel + familie)**

UWV kent op dit moment geen centrale registratie van BZ, BGB en EP+; classificatie als zodanig verschilt per bron.⁸⁴ Daarnaast moeten gegevens betreffende dergelijke zaken/personen wel worden opgenomen in diverse totaalstellingen t.b.v. S&V.

Vandaar dat deze gegevens wél aan het DIM moeten worden geleverd, ook al zijn ze, deels, VK3.

Binnen het DIM worden deze gegevens vervolgens, waar nodig en mogelijk, afgeschermd van de overige gegevens middels bron-specifieke filters op de informatieproducten. De functionaliteit van dergelijke filters moet door de bron worden gevalideerd.

Velden met zeer gevoelige waarden

Soms zijn alleen bepaalde waarden in een veld "zeer gevoelig", en is dat voor de meerderheid van de waarden niet het geval:

- **Velden met vrije tekst**

In vrije tekstvelden kunnen allerlei gegevens staan, dus ook, in theorie, zeer gevoelige (VK3) gegevens.

In de RLO dient te worden aangegeven hoe "vrij" een dergelijk tekstveld is, en of de kans op zeer gevoelige gegevens in dat veld reëel is. Is zulks het geval, dan wordt het veld behandeld als VK3.

Velden met vrije tekst moeten in het algemeen wél aan het DIM geleverd worden. Dit, bijvoorbeeld, voor datakwaliteitsanalyse. Uitzonderingen zijn velden die geen enkele waarde hebben buiten de bron zelf, en velden waarvan het belang voor rapportage/analyse te klein is om het risico op VK3-gegevens te rechtvaardigen.

- **Gevoelige waarden in gestructureerde velden (velden met een vastgesteld waardebereik)**

Als bepaalde waarden in een gestructureerd veld VK3 zijn (bv. beroep=sekswerker, reden stopzetting uitkering = detentie), dan worden deze waarden in de informatieproducten (ook de ongemaskeerde) onherkenbaar gemaakt, bijvoorbeeld door deze te vervangen door de waarde "overig" (of equivalent).

In het DIM zelf zijn ze wel herkenbaar opgeslagen, omdat het mogelijk moet blijven om, bij voorbeeld, te kunnen berekenen hoe groot het aantal stopgezette uitkeringen is dat "detentie" als reden heeft.

Het is ook hierom dat gestructureerde velden met mogelijke VK3-waarden vrijwel altijd wél aan het DIM geleverd moeten worden.

⁸³ Zie "Gegevens t.b.v. beveiliging" later in deze bijlage

⁸⁴ Er bestaan plannen om binnen UWV een centrale registratie voor BZ/BGB/EP+ op te zetten.

Als er in de toekomst een centrale registratie voor BZ/BGB/EP beschikbaar komt, dan zal het DIM erop gaan aansluiten.

Gegevens t.b.v. verdere structurering – Reference data

De door de bronnen geleverde basisgegevens moeten in een aantal gevallen binnen het DIM verrijkt worden met reference data. Dit om ze beter geschikt te maken voor analyse/rapportage.

Het Beleid Reference Data UWV omschrijft reference data als:

Gegevens die worden gebruikt om bedrijfsgegevens te classificeren.

Reference data hebben altijd een vast waardebereik gedurende een bepaalde periode in de tijd.

In de context van het DIM komt dat neer op:

- a) Lijsten van de in een bron gebruikte codes met hun omschrijvingen
- b) Cross-referencelijsten om code-stelsels uit verschillende bronnen op elkaar te kunnen herleiden.
- c) Groeperingshierarchyën t.b.v. analyse en rapportage

In de volgende paragrafen wordt het in het DIM inlezen/verwerken van deze drie typen reference data verder uitgewerkt.

N.B. In vrijwel alle gevallen zijn de daar beschreven processen en IT-oplossingen tijdelijk: verwachte looptijd enige jaren.

Ze zullen overbodig worden zodra de (nog te bepalen) gemeenschappelijke UWV-processen voor reference data management zijn ingericht, en de daarbij behorende (nog te ontwikkelen) IT-oplossingen als bron op het DIM zijn aangesloten.

Codes + Omschrijvingen

In het overgrote deel van de gevallen bevat het bronsysteem waarin een code gebruikt wordt óók een tabel met de toegestane waarden voor die code, inclusief omschrijving.

- Deze tabel dient dan, als onderdeel van de “brede bronontsluiting” met de overige gegevens te worden meegeleverd.
- Verwerking ervan binnen het DIM is vervolgens onderdeel van de verwerking van die bronontsluiting.
- De relatie tussen deze codetabel en de objecten waarin de code gebruikt wordt dient in de RLO te zijn vastgelegd.

In sommige gevallen bevat het bronsysteem geen codetabel, en is de lijst met toegestane codes en hun betekenis buiten de applicatie vastgelegd, bijvoorbeeld in de werkinstructies voor de medewerkers die gegevens in het bronsysteem invoeren.

- De broneigenaar dient (wijzigingen in) de lijst met codes en omschrijvingen (inclusief begin- en einddatum geldigheid) dan apart aan Gegevensdiensten (als eigenaar van het DIM) op te leveren. E.e.a. conform in het levercontract vastgelegde afspraken.
- De relatie tussen deze externe codelijst en de objecten waarin de code gebruikt wordt dient in de RLO te zijn vastgelegd.
- DIM-beheer legt (de wijzigingen op) de lijst vast in een “reference data spreadsheet” (zie paragraaf “Verwerking van niet via bronontsluiting geleverde reference data”), zodat de gegevens op een standaardwijze in het DIM kunnen worden ingelezen.

Cross-referencelijsten

Verschillende bronnen gebruiken vaak verschillende codestelsels voor hetzelfde type gegeven. Om geconsolideerd over deze bronnen te kunnen rapporteren moet een cross-referencelijst worden opgesteld (en in het DIM ingelezen, waarin de verschillende code-stelsels op elkaar worden herleid.

- Gegevensdiensten dient (als gegevenseigenaar van bron- en afnemer-overstijgende afgeleiden het DIM) een overlegproces in te richten om de vereiste kruisverwijzingen te achterhalen en actueel te houden (inclusief begin- en einddatum geldigheid), en om ze formeel vast te leggen
- DIM-beheer legt (de wijzigingen op) de diverse kruisverwijzingen vast in een "reference data spreadsheet" (zie paragraaf "Verwerking van niet via bronontsluiting geleverde reference data"), zodat de gegevens op een standaardwijze in het DIM kunnen worden ingelezen.

Groeperingshiërarchieën

Geaggregeerd rapporten/analyseren vereist vaak criteria om waardebereiken of groepen codes "onder één noemer te vatten". Deze criteria worden gestructureerd vastgelegd in groeperingshiërarchieën.

Deze hiërarchieën zijn vrijwel altijd opgesteld door een afnemende organisatie (bijvoorbeeld FEZ).

- De hiërarchie-eigenaar dient (wijzigingen in) de hiërarchie (inclusief begin- en einddatum geldigheid) aan Gegevensdiensten (als eigenaar van het DIM) op te leveren. Hiervoor wordt een levercontract opgesteld.
- DIM-beheer legt (de wijzigingen op) de diverse hiërarchieën vast in een "reference data spreadsheet" (zie paragraaf "Verwerking van niet via bronontsluiting geleverde reference data"), zodat de gegevens op een standaardwijze in het DIM kunnen worden ingelezen.

N.B. Als een hiërarchie maar een beperkte scope heeft, geen eisen t.a.v. historische traceerbaarheid, en alleen relevant is voor "in selfservice" ontwikkelde/beheerde BusinessObjects universes (of andere typen semantische lagen), dan voldoet opname ervan in die semantische laag/lagen. Inlezen in het DIM is dan niet noodzakelijk.

Verwerking van niet via bronontsluiting geleverde reference data

Alle reference data die niet via een brede bronontsluiting aan het DIM geleverd wordt doorloopt de volgende stappen:

- De Gegevenseigenaar (zie voorgaande slides) vergaart en actualiseert reference data, en geeft wijzigingen door aan Gegevensdiensten (als eigenaar DIM). Welk onderdeel van Gegevensdiensten is nog niet bepaald; onderdeel Implementatie
- Gegevensdiensten geeft de gegevens door aan DIM-beheer, die ze verwerken in een "reference data spreadsheet". Dit spreadsheet is geversioneerd langs de geldigheidstijdlijn.
- DIM-beheer voert tijdens deze verwerking ook technische en inhoudelijke controles uit. Eventuele issues worden teruggekoppeld; DIM-beheer repareert niet zelf.
- DIM-beheer draait, na verwerken van de wijzigingen, een script om de volledige inhoud van het spreadsheet (alle versies) te laden in een hiervoor ingerichte database.
- De gegevens worden vervolgens middels een database-link op deze database op standaardwijze verder verwerkt in het DIM.

Gegevens als parameters

Parameters worden gebruikt om het aanmaken van informatieproducten te sturen, en niet als basis voor analyse/rapportage. Ze worden pas opgehaald als ze nodig zijn (tijdens het aanmaken van

het betreffende informatieproduct) en worden alleen in het DIM opgeslagen als zulks voor de traceerbaarheid of reproduceerbaarheid van het informatieproduct vereist is.

Het DIM onderscheidt twee typen parameters:

- **Abonnementsgegevens**

Soms moeten informatieproducten (m.n. bestandsleveringen) samengesteld worden o.b.v. een “abonnement”, bijvoorbeeld zoals opgeslagen in UGC.

Deze abonnementsgegevens worden pas bij het aanmaken van het informatieproduct door het DIM opgehaald (bv. via een service). Ze worden dus niet, zoals bij de basis-gegevens, middels een bronlevering vergaard. De via de service opgehaalde parameters worden wel in het DIM vastgelegd, maar dat is t.b.v. de traceerbaarheid/reproduceerbaarheid van het erop gebaseerde informatieproduct, en **niet** t.b.v. verdere analyse.⁸⁵

N.B. Levering van abonnementsgegevens als parameters is volledig gescheiden van levering van abonnementsgegevens als basis-gegevens (t.b.v. analyse/rapportage over de abonnementen zelf).

Als er, bijvoorbeeld, in de toekomst een behoefte ontstaat om via het DIM het abonnementsgedrag van abonnees te analyseren, dan zal UGC dan op twee manieren gegevens aan het DIM leveren: als “gewone” bron (via een brede bronontsluiting), en als provider van de services met parameters.

- **Ad hoc parameters**

Gebruik van ad hoc parameters zal in het algemeen vooral voorkomen in de End-userzone, en dus bij het uitlezen van een informatiegebied, en niet bij het aanmaken ervan. Deze parameters vallen daarmee buiten het DIM.

Gegevens t.b.v. beveiliging

Gegevens betreffende gevoelige personen/zaken worden wél in het DIM opgeslagen, maar toegang ertoe is afgeschermd.

Voor deze afscherming zijn “beveiligingsgegevens” nodig.

Op dit moment haalt het DIM maar één type beveiligingsgegevens op:

- **VIP's**

Voor VIP's is deze afscherming gebaseerd op de “VIP-lijst” uit UPA⁸⁶. Levering van deze gegevens aan het DIM volgt technisch dezelfde route als die voor basis-gegevens, maar heeft uitdrukkelijk alleen beveiliging als doel, en is dus géén brede bronontsluiting.

Opslag van de VIP-lijst is dan ook gescheiden van opslag van “gewone” brongegevens.

N.B. Mochten, in de toekomst, UPA-gegevens ook gebruikt gaan worden in rapportages/analyses, dan zal UPA op twee manieren gegevens aan het DIM gaan leveren: als “gewone” bron (via een brede bronontsluiting), en als leverancier van beveiligingsgegevens.

Dit omdat de twee leveringen een volledig ander doel hebben, en dus ook anders beargumenteerd en bestuurd moeten worden.

Beveiligingsgegevens worden binnen het DIM op dezelfde wijze verwerkt als gewone (brede) bronontsluitingen (zie paragraaf 0 (Gegevens als basis – brede bronontsluiting), zij het dat er in de gebruiksbeperkingen (in RLO en/of GLO) expliciet is opgenomen dat de betreffende gegevens alleen voor beveiligingsdoeleinden mogen worden gebruikt (en dus niet voor analyse op,

⁸⁵ Voor fase 4/5 worden alleen voor de SUAG-levering abonnementsgegevens gebruikt. Deze worden, via standaardservices, opgehaald uit UGC en POLIS+

⁸⁶ Een officiële lijst met alle BSN's die als VIP moeten worden beschouwd. Het betreft hier “echte VIP's”, dus niet eigen personeel o.i.d.

bijvoorbeeld, het VIP-zijn zelf).

Er bestaan plannen om binnen UWV een centrale registratie voor BZ/BGB/EP+ op te zetten. Zodra die er is zal het DIM erop gaan aansluiten. Deze aansluiting (en de opslag van de via die aansluiting vergaarde gegevens) zal vergelijkbaar zijn met die voor VIP's.

Tot die tijd zal het afschermen van BZ/BGB/EP+ gebaseerd zijn op, via brede bronlevering verkregen, gegevens uit de bron zelf.

BIJLAGE C : AFKORTINGEN

Lijst van afkortingen (afkortingen van applicaties zijn te vinden in Confipedia)

Afkorting	Verklaring, eventueel met toelichting
BIT	Bureau ICT-Toetsing
CGM	Canoniek Gegevens Model
DA GEIN	DoelArchitectuur GEGEvensINtegratie. Context voor deze PSA
DF	Datafabriek (het project) of Datafabriek (de oorspronkelijk voorziene IV-organisatie)
DIA	Data Integratie en Analyse
DIM	Data Integratie Magazijn
distro	Een door een leverancier "enterprise ready" gemaakte samenhangende set van open source software componenten, meestal geleverd inclusief support.
DTL	DaTa-Laag Onderdeel van DWH 3.0, vergelijkbaar met de bronzone van het DIM.
DWARFS	Officiële naam van DWH 1.0
DWH	Data Warehouse De afkorting wordt ook gebruikt voor de beheer-afdeling van de huidige centrale data warehouses, en voor het toekomstige IV-domein waar dit beheer zal gaan landen.
EA GH	Enterprise Architectuur GegevensHuishouding. Document in wording. Dit document zal een groot aantal "open punten" m.b.t. de precieze rol van het DIM, en ander applicaties in het DIA-domein, adresseren.
ELT-patroon	Een manier van gebruik van ETL-software waarbij alle gegevens eerst vanuit een gegevensbron naar het gegevensdoel (meestal een relationele database) worden getransporteerd (Extract, Load), waarna de rekenkracht van die doel-omgeving wordt gebruikt voor de gegevenstransformaties (Transform). Bij het ELT-patroon voert de ETL-software dus niet zelf de transformaties uit (zoals bij het ETL-patroon). In plaats daarvan genereert de ETL-software code (bv. PL/SQL) die binnen de doelomgeving kan worden uitgevoerd.
ETL	Extraction Transformation Loading Als pakket in de markt beschikbare functionaliteit voor gegevenstransformatie en -replatie, i.h.a. geoptimaliseerd voor gebruik in bulkprocessen.
ETL-patroon	Een manier van gebruik van ETL-software waarbij alle gegevens eerst vanuit een gegevensbron naar de ETL-omgeving worden getransporteerd (Extract), daar worden getransformeerd (Transform) en vervolgens geladen in het gegevensdoel (meestal een relationele database) (Load).
ETL+	ETL-software plus daaraan nauw gerelateerde, ook door UWV-benodigde software t.b.v. gegevensmaskering, datakwaliteitsmanagement en metadata management. (UWV-specifieke term, gebruikt in de ETL-aanbesteding)
FO	Functioneel Ontwerp
FUGEM	FUnctioneel GEGEvens Model
GEB	GegevensbeschermingsEffectBeoordeling (Engels: Privacy Impact Assessment; PIA)
GIA	GegevensInwinningsAfspraak Opvolger van de GLO.
GLA	GegevensLeveringsAfspraak Opvolger van de SNO.
GLO	GegevensLeveringsOvereenkomst. Een "contract" waarin de afspraken rondom een levering zijn vastgelegd.

	In de context van het project Datafabriek betreft het dan de afspraken voor een levering van een bron aan het DIM (of een ander datawarehouse). De GLO wordt, als onderdeel van de invlechting van de afdeling DWH in Gegevensdiensten, vervangen door de GIA.
HRC	HoofdRekenCentrum. Kan zowel verwijzen naar het bestaande rekencentrum (bij IBM) als naar het nieuwe (bij DXC)
HUDSV	Handboek Uniforme Definities Sturen & Verantwoorden
IDA	IBM Data Architect
IGC	IBM Information Governance Catalog
MVL	Materialized View Layer Onderdeel van DWH 3.0 Een op de DTL-laag van datzelfde DWH gebaseerde set Oracle Materialized Views (alleen die objecten die in de integratiezone nodig zijn). Ligt op het snijvlak van bronzon (DTL) en integratiezone. De inhoud van de MVL is op wat schoning na gelijk aan de DTL, maar wordt in plaatjes vaak onderin de integratiezone getekend.
NGP	Nieuw GegevensPakhuis – Officiële naam van DWH 2.0
RDBMS	Relationeel DataBase Management Systeem
RLO	Record LayOut; een gedetailleerde beschrijving van de door een bron geleverde gegevens.
SNO	ServiveNiveauOvereenkomst. Een “contract” waarin de afspraken rondom een levering zijn vastgelegd. In de context van het project Datafabriek betreft het dan de afspraken voor een levering van het DIM (of een ander datawarehouse) aan een afnemer. De SNO wordt, als onderdeel van de invlechting van de afdeling DWH in Gegevensdiensten, vervangen door de GLA.
TEGEM	TEchnisch GEgevens Model
TCO	Total Cost of Ownership
UDS	UWV Data Store – Officiële naam van DWH 3.0