

Toegang tot data bij Data Integratie Magazijn (DIM)

|  |  |
| --- | --- |
| **UWV Gegevensdiensten** | |
| Auteur | Henk van Tent Becking |
| Business Security Officer (BSO) | Bas Oosterkamp |
| Datum (laatste wijziging) | 2 februari 2023 |
| Versie | 1.0 |



Inhoudsopgave

[**Inleiding** 3](#_Toc126252460)

[Privacy by design by default 3](#_Toc126252461)

[Het Data Integratie Magazijn (DIM) 3](#_Toc126252462)

[OTAP-beleid 4](#_Toc126252463)

[**Toegang tot data bij DIM** 5](#_Toc126252464)

[Bronontsluiting 5](#_Toc126252465)

[Integratie 6](#_Toc126252466)

[Informatieproducten 6](#_Toc126252467)

[Releasematig werken 7](#_Toc126252468)

[Rol gebaseerde toegang 7](#_Toc126252469)

[**Vervolgstappen** 8](#_Toc126252470)

[1. Onderzoek OPTIM 8](#_Toc126252471)

[2. Inrichten gebruikszones en vernieuwen datasets 8](#_Toc126252472)

[3. Releasematig werken 8](#_Toc126252473)

[4. Kaders voor dataminimalisatie 8](#_Toc126252474)

[5. Implementatie verbeteringen 8](#_Toc126252475)

[**Bijlage**: Citaten uit GegevensbeschermingsEffectBeoordeling (GEB) 9](#_Toc126252476)

[Verwerkingen en de klasse van persoonsgegevens 9](#_Toc126252477)

[Maskering 10](#_Toc126252478)

[Toetsing op doelbinding, rechtsgrond, proportionaliteit en subsidiariteit 10](#_Toc126252479)

[Toegangsbeveiliging 10](#_Toc126252480)

[Medewerkers 11](#_Toc126252481)

[**Bijlage**: Grafische weergave “Privacy by Design by Default & Data Lifecycle Management” in DIM 11](#_Toc126252482)

# **Inleiding**

Het project Datafabriek voert verbeteringen door in de centrale datawarehouse voorziening en realiseert een nieuw Enterprise Data Warehouse met de naam Data Integratie Magazijn (DIM).

In het DIM worden veel gegevens van burgers verwerkt en gecombineerd. Deze informatie rijkdom heeft tot gevolg dat er privacy risico’s zijn.

Het DIM is daarom ontworpen conform het “privacy-by-design-by-default” principe. Daarbij wordt voldaan aan geldende UWV-richtlijnen en kaders op het gebied van privacy en informatiebeveiliging.

## Privacy by design by default

Sinds 2016 is de AVG van kracht en wordt privacy by design by default voorgestaan bij vernieuwing en aanpassing van processen en systemen. De bedoeling hiervan is om al in de ontwerpfase privacybescherming van de betrokkene mee te nemen (‘by design’), en tevens standaard de meest privacybeschermende instellingen toe te passen (‘by default’).

De wettelijke tekst wordt door de Autoriteit Persoonsgegevens (AP) als volgt uitgelegd:

* “Privacy by design houdt in dat u als organisatie al **tijdens de ontwikkeling** van producten en diensten (zoals informatiesystemen) ten eerste aandacht besteedt aan **privacy verhogende maatregelen**, ook wel privacy enhancing technologies (PET) genoemd. Ten tweede houdt u rekening met **dataminimalisatie**: u verwerkt zo min mogelijk persoonsgegevens, dat wil zeggen alleen de gegevens die noodzakelijk zijn voor het doel van de verwerking.
* Privacy by default houdt in dat u **technische en organisatorische maatregelen** moet nemen om ervoor te zorgen dat u, als **standaard**, alléén persoonsgegevens verwerkt die noodzakelijk zijn voor het specifieke doel dat u wilt bereiken.

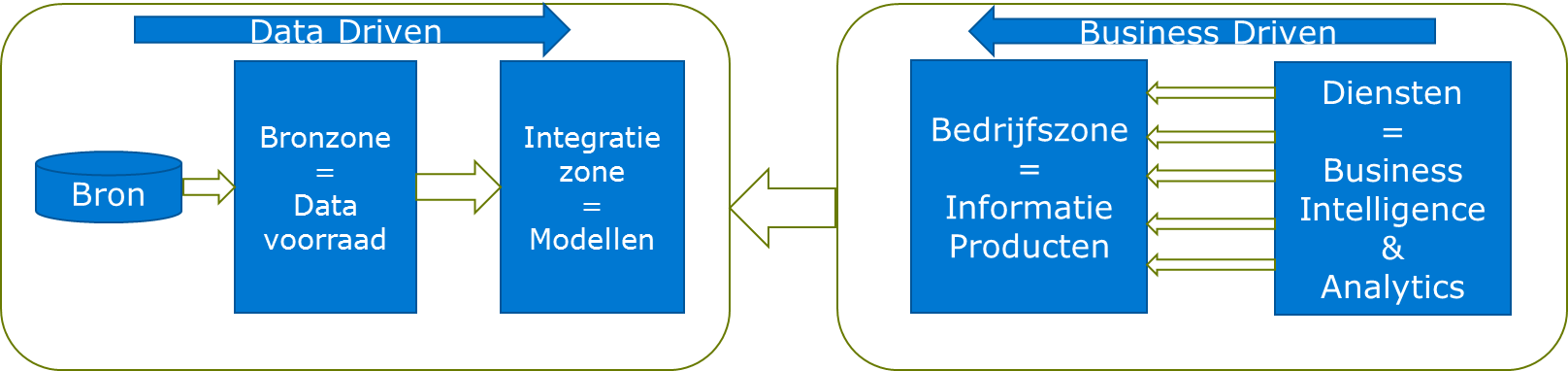
Er zijn specifieke maatregelen, zowel technisch, organisatorisch als juridisch om risico’s te voorkomen of te verminderen. De belangrijkste (klasse van) maatregelen die kunnen worden toegepast zijn:

1. Maskering
2. Toetsing op doelbinding, rechtsgrond, noodzaak, proportionaliteit en subsidiariteit
3. Data Lifecycle Management vanuit de bronbestanden
4. Toegangsbeveiliging van data die beschikbaar wordt gesteld en/of geleverd
5. Screening van nieuwe medewerkers
6. Diverse trainingen en workshops op het gebied van awareness en integriteit

In dit document wordt vooral ingegaan op de klasse “4 ”. In de op pagina 13 worden de andere klassen globaal beschreven.

## Het Data Integratie Magazijn (DIM)

Het DIM omvat de “Data Delivery” functionaliteit. In het Project Start Architectuur (PSA) document kunnen we de volgende samenvatting vinden:



Voor meer informatie over het DIM en de maatregelen wordt verwezen naar de PSA Datafabriek.

## OTAP-beleid

De uitgangspunten van privacy by design by default zijn, op basis van geclassificeerde dataverwerking, verwerkt in het OTAP-beleid (zie OTAP-beleid).



De toelichting van dit OTAP-beleid ten behoeve van het ontwikkelen en onderhouden van data-logistiek volgt hierna;

Het O-regime en T-regime zijn bedoeld voor functionele ontwikkeling en testen. Zij bevatten subsets van niet herleidbare persoonsgegevens. De identificerende attributen zijn op een eenduidige wijze “gesynthetiseerd” waardoor het mogelijk is om verschillende bronnen te kunnen integreren.

Dit wordt bereikt door deze attributen te maskeren.

Bij deze maskering worden identificeerbare persoonsgegevens zodanig vervormd dat het niet mogelijk is de maskering ongedaan te maken om originele waarde terug te krijgen. Er wordt gebruik gemaakt van de standaard maskeringstechnieken die door IBM DataStage OPTIM geboden worden. Deze technologie is onderdeel van de verworven ETL+ tooling.

In de Acceptatie en Productie omgevingen vallen onder het productie-regime (P-regime) en hier zijn naast gemaskeerde gegevens ook de ongemaskeerde gegevens beschikbaar zolang ze ook in de bronbestanden zijn opgenomen. Hier vinden de performance testen en controle van aantallen plaats. De toegang tot ongemaskeerde gegevens is beperkt tot alleen medewerkers die hiervoor aangewezen zijn. De toegang tot de gemaskeerde gegevens (welke dus ook integreerbaar zijn) is breder toegestaan.

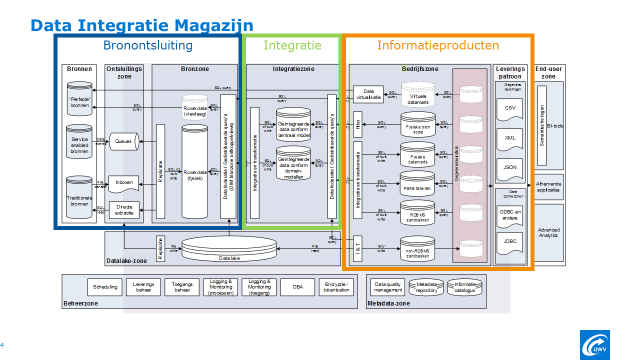
Voor toegang is “nut en noodzaak” conform de AVG een voorwaarde.

Alle bronnen (breder dan alleen applicatie bronnen) die noodzakelijk zijn voor integratie dienen in het DIM gemaskeerd te worden opgenomen. Ongemaskeerde opname is afhankelijk van de bron en gebruik.

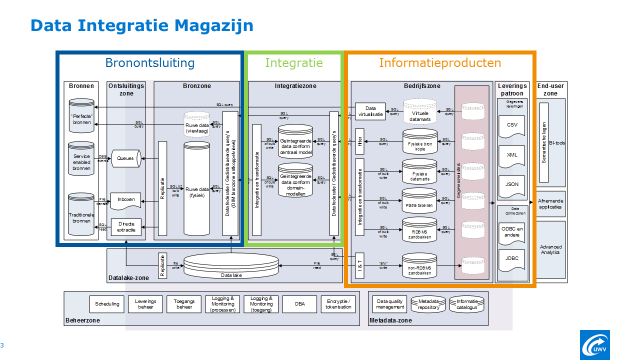
De maskeringswijze voor iedere OTAP-omgeving is verschillend. Er worden per omgeving andere “Salt keys” gebruikt. Deze sleutels zijn de cryptografische elementen.

In de O-omgeving van de bronontsluiting heeft 2 verschillende Salt-keys. Zie verder hieronder.

# **Toegang tot data bij DIM**

Het DIM bestaat uit een bronzone, integratiezone en bedrijfszone. Deze zones worden gebruikt voor Bronontsluiting, Gegevens integratie uit meerdere bronnen en het maken van Informatieproducten:

De toegang tot data en de vulling in de OTAP-omgevingen zijn ook per gebruikszone verschillend.



## Bronontsluiting

Bij een nieuwe applicatie/ bron dient een analist de bron en de record layout (RLO) te analyseren. Indien noodzakelijk kan dit met toestemming in de productieomgeving of in een gemaskeerde levering. Deze levering dient gemaskeerd te zijn zodat hier geen aan een persoon herleidbare informatie in zit.

Vanwege privacy redenen zijn de ongemaskeerde bestanden in de O- en T-omgevingen ook gemaskeerd. Hiervoor kan een eerste bron levering met behulp van OPTIM worden geanonimiseerd en als subset worden aangeleverd. Op dit moment wordt dit nog via TCS gerealiseerd.

Deze gemaskeerde subset wordt in de O-omgeving geplaats als “ongemaskeerde” subset, waarna de bronontsluiting teams deze kunnen gebruiken voor de bron ontsluiting activiteiten.

Wanneer de in de T-omgeving wordt getest dan wordt van de gemaskeerde subset een opnieuw gemaskeerde subset gerealiseerd. Dit dubbel maskeren heeft geen gevolgen voor de test en integratie.

Onder P-regime worden de A-omgeving en de P-omgeving gevuld met productiedata. Hiervan wordt een gemaskeerde variant opgebouwd. Deze kan worden gebruik voor het testen en analyse. Analyse in productie-omgeving is voorbehouden toegang door middel van de Yellow-Enveloppe Procedure (YEP).

De aantallen en performance dient in A en P-omgeving gelijk te zijn.

## Integratie

Wanneer de bronontsluiting overgaat naar de A-omgeving dan kunnen de opgebouwde resultaten gemaskeerd en ongemaskeerd beschikbaar worden gesteld in de O-omgeving voor de Integratie activiteiten. Doordat deze gemaskeerd en ongemaskeerd niet herleidbaar zijn, is ook hier van privacy maatregelen sprake.

Onder P-regime worden de A-omgeving en de P-omgeving gevuld met productiedata. Hiervan wordt een gemaskeerde variant opgebouwd. Deze kan worden gebruik voor het testen en analyse. Analyse in productie-omgeving is voorbehouden toegang door middel van de Yellow-Enveloppe Procedure (YEP).

De aantallen en performance dient in A en P-omgeving gelijk te zijn.

## Informatieproducten

Wanneer de integratie overgaat naar de A-omgeving dan kunnen de opgebouwde resultaten gemaskeerd en ongemaskeerd beschikbaar worden gesteld in de O-omgeving voor de Informatieproducten activiteiten.

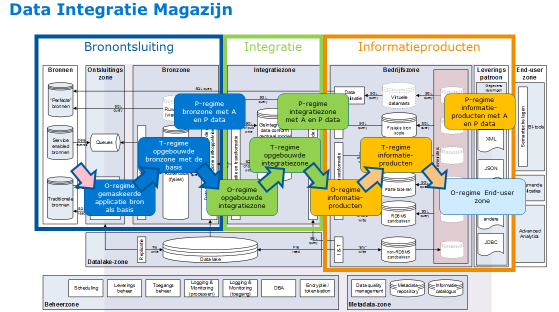
Onder P-regime worden de A-omgeving en de P-omgeving gevuld met productiedata. Hiervan wordt een gemaskeerde variant opgebouwd. Deze kan worden gebruik voor het testen en analyse. Analyse in productie-omgeving is voorbehouden toegang door middel van de Yellow-Enveloppe Procedure.

De aantallen en performance dient in A en P-omgeving gelijk te zijn.

Wanneer de Informatieproducten overgaat naar de A-omgeving dan kunnen de opgebouwde resultaten gemaskeerd en ongemaskeerd beschikbaar worden gesteld in de O-omgeving voor de End-user zone.

Wanneer we gebruik maken van “end-2-end” en/of geïntegreerde gebruikszone teams dan zijn ook de OTAP-omgevingen van de 3 gebruikszones geïntegreerd. Het is echter verstandig om de workflow tussen de gebruikszones te blijven volgen.

Het geheel ziet er dan zo uit:



Bij het “verlaten van het T-regime naar de A-omgeving worden ook opvolgende O-omgevingen beïnvloed. Dit betekend dat er een volgordelijkheid is tussen de beschikbaarheid van data. Dit dient te worden begeleid met releasematig werken.

## Releasematig werken

We willen ook sneller en flexibelere werken door vaker (kleinere) werkende onderdelen in producten te nemen. Dit wordt ook wel CI/CD genoemd[[1]](#footnote-1). Hiervoor is ook een gecontroleerde samenstelling van de onderdelen in de hele OTAP-cyclus noodzakelijk. Wanneer een onderdeel niet werkt dan is het mogelijk om dit onderdeel geïsoleerd te herstellen en aan een nieuwe productversie toe te voegen. Ook is het noodzakelijk om meerdere versies van functionaliteit in productie te hebben. Hierdoor kan er een nieuwe versie in productie worden genomen zonder dat de vorige versie ook direct aangepast moeten worden. Hierbij zullen architectuurkaders worden ingezet die data-minimalisatie ondersteunen.

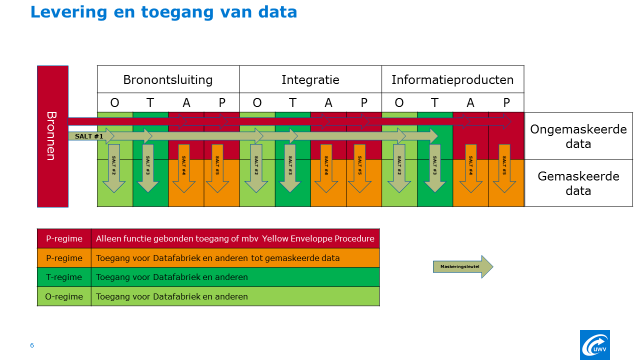
VOORBEELD MINIMALISATIE EN MEER GEMASKEERDE DATA DAN ONGEMASKEERDE DATA UITWERKEN

Dit releasematig werken dient bij Datafabriek nog verder vorm te worden gegeven.

## Rol gebaseerde toegang

Door de O en T-omgeving van niet herleidbare data te voorzien is het mogelijk om functionele toegang te verlenen. Voor de A en P-omgeving blijft het P-regime van toepassing en hiermee is toegang beperkt tot de gemaskeerde data. Voor ongemaskeerde data is aanvullende toestemming nodig gebaseerd op “nut en noodzaak”.

Visueel ziet dat er zo uit:



De ongemaskeerde data in de O- en T-omgevingen zijn wel gemaskeerd ten behoeve de toegang.

Alle O-omgevingen (en T-, A-, P-omgevingen) van de gebruikszones zijn onderling integreerbaar door gebruik van dezelfde “Salt”-key. De Data “loopt” door de keten van gebruikszones heen en wordt hergebruikt en/of bewerkt. Dit ondersteund ook GAT en keten testen.

De O- en T-omgevingen hebben ten opzichte van de A- en P-omgevingen een beperkte subset. Voor alle O- en T-omgevingen zijn verschillende versies van de datasets c.q. bronnen beschikbaar.

# **Vervolgstappen**

Om de hiervoor genoemde onderdelen goed te implementeren of te verbeteren zijn de volgende activiteiten noodzakelijk:

1. **Onderzoek OPTIM**

OPTIM heeft mogelijkheden met betrekking tot data filtering naar subsets. Deze functie is nu belegd bij TCS. Wanneer we deze functie sneller en flexibeler willen inzetten dan moeten we onderzoeken hoe dit met OPTIM kan worden bereikt.

Wanneer OPTIM een alternatief is dient ook onderzocht te worden of dit past in het huidige gebruik en licenties.

1. **Inrichten gebruikszones en vernieuwen datasets**

In de O- en T-omgevingen wordt op dit moment nog geen verschil gemaakt tussen de verschillende gebruikszones. Hierdoor zitten verschillende teams elkaar regelmatig “in de weg”. De gebruikszones dienen te worden vormgegeven en geïmplementeerd.

Per gebruikszone kunnen we de O- en T-omgeving datasets vernieuwen. Hiermee is het probleem met niet integreerbare datasets ook verholpen.

1. **Releasematig werken**

Op dit moment is het maken van verschillende releases niet goed mogelijk. Kleine aanpassingen hebben grote vertragingen tot gevolg. Het releasematig werken dient te worden uitgewerkt in de huidige tooling en geïmplementeerd. Het is goed mogelijk dat hierbij opleidingen en/of trainingen noodzakelijk zijn.

1. **Kaders voor dataminimalisatie**

Om te voorkomen dat door gebruik van verschillende versies er veel extra data wordt bewaard zullen architectuurkaders voor het minimaliseren van data verder worden uitgewerkt.

1. **Implementatie verbeteringen**

Het implementeren kan gebeuren terwijl “de winkel openblijft”. Het zal versnellingen, flexibeler werken en verbeteringen in de kwaliteit tot gevolg hebben.

Het releasematig werken en bewaken van de datasets workflow kan worden ondergebracht bij de releasemanager.

# **Bijlage**: Citaten uit GegevensbeschermingsEffectBeoordeling (GEB)

**De informatie in het GEB-rapport bevat bedrijfskritische informatie over de gegevensverwerkingen van UWV. Deze informatie dient daarmee vertrouwelijk te worden behandeld.**

Het DIM is ontworpen conform het “privacy-by-design” en “privacy-by-default” (PbDD) principe. Daarbij wordt tevens voldaan aan alle geldende UWV-richtlijnen op het gebied van informatiebeveiliging en privacy (IB&P). Een opsomming hiervan is terug te vinden in de Project Start Architectuur (PSA).

## Verwerkingen en de klasse van persoonsgegevens

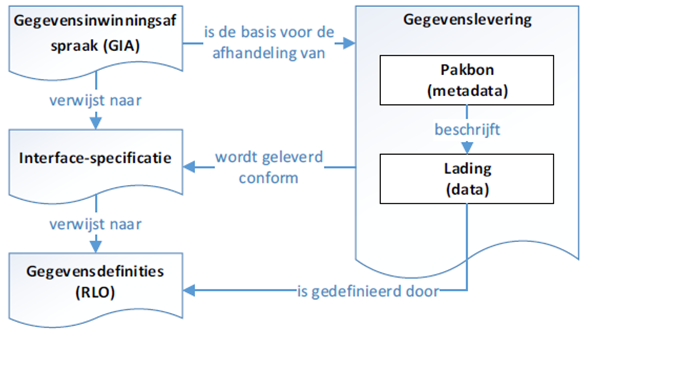
Het DIM zal gegevens bevatten uit de volgende vertrouwelijkheidsklassen:

* Vertrouwelijkheidsklasse 3: Strikt vertrouwelijk
* Vertrouwelijkheidsklasse 2+: Extra vertrouwelijk
* Vertrouwelijkheidsklasse 2: Vertrouwelijk
* Vertrouwelijkheidsklasse 1: Basisniveau
* Vertrouwelijkheidsklasse 0: Publiek niveau

Het DIM mag altijd vertrouwelijkheidsklasse 3 (VK3) gegevens laden in de bronzone (= datakluis). Eventuele gebruiksbeperkingen (voor zowel VK3 als ook andersoortige gegevens) kunnen vanuit de bron worden opgelegd door deze op te nemen in de GegevensInwinningsAfspraak[[2]](#footnote-2) (GIA) tussen de bron en het DIM.

Voor het beschikbaar stellen van gegevens met vertrouwelijkheidsklasse 3 (VK3) gelden een aantal extra regels:

* In principe stelt het DIM géén VK3-gegevens beschikbaar aan afnemers.
* Als er een zéér zwaarwegend belang is om zulks toch te doen (dit ter beoordeling van bron én afnemer, en vergezeld van een GegevensLeveringsAfspraak[[3]](#footnote-3) (GLA) en vastlegging van doelbinding, subsidiariteit en proportionaliteit), dan kunnen VK3-gegevens toch binnen het DIM beschikbaar worden gesteld
* In de GLA zullen de eventuele (tussen de bron en afnemer) overeengekomen gebruiksbeperkingen met betrekking tot VK3 worden vastgelegd
* Andere gebruiksbeperkingen op gegevens geleverd door een bron worden vanuit de GIA overgenomen naar de GLA indien ze van toepassing zijn voor het informatiegebied



## Maskering

Er wordt gebruik gemaakt van de standaard maskeringstechnieken die door IBM DataStage OPTIM geboden worden. Deze technologie is onderdeel van de verworven ETL+ tooling. Bij deze maskering worden identificeerbare persoonsgegevens zodanig vervormd dat het niet mogelijk is de maskering ongedaan te maken om originele waarde terug te krijgen.

Als in dit document (*GEB DIM*) wordt gesproken over een gemaskeerde waarde of de gemaskeerde vorm van een gegeven, gaat dit om een in de database opgeslagen waarde en niet een techniek waarbij het originele persoonsgegeven dynamisch wordt gemaskeerd (het is voor het DIM dus een ander opgeslagen gegeven).

Alle persoonsgegevens in het DIM zijn initieel zowel in ongemaskeerde als gemaskeerde vorm beschikbaar. De twee vormen worden gescheiden van elkaar opgeslagen in aparte gedeelten van het DIM. Deze delen zijn niet aan elkaar te koppelen. De gemaskeerde vorm kan beschouwd worden als pseudonieme gegevens. Dit omdat er in de bron en in het DIM een ongemaskeerde versie bestaat, waarmee gepseudonimiseerde persoonsgegevens terug herleid kunnen worden tot de betrokkene.

De maximale maskering wordt bepaald door analyse van herkenbaarheidsbehoefte. Er wordt maximaal gemaskeerd, maar de gemaskeerde persoonsgegevens moeten nog wel bruikbaar zijn door de afnemende partij. Verdere maskering bij levering is mogelijk, maar het ongedaan maken van maskering veelal niet. De herkenbaarheidseis bepaalt de “ondergrens” voor het maskeren. Verder maskeren dan deze ondergrens maakt het onmogelijk voor het afnemende proces om deze gegevens nog te gebruiken.

Wanneer de ongemaskeerde versie verwijderd wordt, uit zowel de bron als het DIM, blijven de gemaskeerde gegevens nog steeds pseudoniem. Dit is een gevolg van de rijkdom van informatie waar het UWV over beschikt. Juridisch gezien is er dan nog steeds sprake van op organisatieniveau gepseudonimiseerde gegevens en niet geanonimiseerd. De AVG is nog steeds van toepassing op deze gegevens.

Pseudonimisering gebeurt waar nodig tot het hoogste niveau zoals dat is beschreven in de UWV ICT Richtlijn Data Anonimisering en Pseudonimisering. Andere gegevens dan persoonsgegevens worden in het DIM gescheiden opgeslagen van – al dan niet gemaskeerde – persoonsgegevens. Deze gegevens kunnen nog wel gekoppeld worden met zowel de ongemaskeerde als gemaskeerde gegevens. Hier zal gelden dat koppeling aan de gemaskeerde gegevens de default-optie is (privacy-by-design). Koppeling aan ongemaskeerde gegevens is alleen mogelijk indien hier specifiek doelbinding, rechtsgrond, proportionaliteit en subsidiariteit voor is aangetoond.

## Toetsing op doelbinding, rechtsgrond, proportionaliteit en subsidiariteit

Voor het verzamelen en opslaan van de brongegevens volgen de Datafabriek en het DIM de bron. De Datafabriek betrekt alleen gegevens uit een bron als daarvoor een gegevensleveringsafspraak (GLA) De GLA is ene resultaat van het doorlopen van het GegevensLeveringsVerzoek (GLV) proces. Dit proces toetst op doelbinding, rechtsgrond, proportionaliteit en subsidiariteit bij de afnemer en controleert tevens of levering van het Informatieproduct past binnen eventuele beperkingen die door de bron zijn opgelegd. De Datafabriek levert vanuit het DIM-persoonsgegevens alleen wanneer specifiek rechtsgrond, proportionaliteit en subsidiariteit is aangetoond voor gebruik ervan binnen het afnemend proces is vastgesteld.

## Toegangsbeveiliging

Maskering gebeurt in het DIM op databaseniveau (en niet in eindgebruikers-tools), zodat alternatieve toegangsmethoden (bijvoorbeeld maatwerk-SQL) geen beveiligingsrisico opleveren. Ongemaskeerde persoonsgegevens worden in een separaat deel van het DIM opgeslagen, zodat de toegang ertoe zo eenvoudig mogelijk kan worden afgeschermd. Alle directe toegang tot de in het DIM opgeslagen gegevens loopt via een datamart of een gegevensvenster.

Gegevensvensters zijn virtuele gegevensverzamelingen die ontsloten worden ten behoeve van selfservice BI en/of selfservice analytics, dat wil zeggen rechtstreeks gebruik door afnemers, bijvoorbeeld ten behoeve van het uitvoeren van query’s en ad-hoc rapportage. Anders dan bij datamart 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 etc. geldt. Deze filtering wordt ook ingezet om te garanderen dat gegevens van extra gevoelige personen (VIP’s, eigen personeel) alleen zichtbaar zijn als er expliciet voor die gegevens rechtsgrond etc. is.

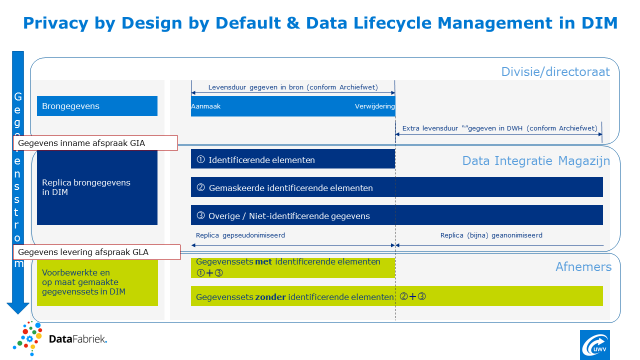
Het DIM wordt ingericht volgens alle geldende UWV-richtlijnen op het gebied van privacy en informatiebeveiliging. Een opsomming hiervan is terug te vinden in de PSA.

## Medewerkers

Voor het bevorderen van awareness op het gebied van verwerking van persoonsgegevens, vinden jaarlijks diverse trainingen en workshops plaats. Bijvoorbeeld de workshop “Integriteit” welke wordt verzorgd door Bureau Integriteit.

Tevens is er de voorgenomen maatregel om nieuwe medewerkers te screenen. Ook wordt van iedereen gevraagd om een VOG (Verklaring Omtrent Gedrag) in te leveren. Veel van deze maatregelen zijn onderdeel van het HRM-proces.

# **Bijlage**: Grafische weergave “Privacy by Design by Default & Data Lifecycle Management” in DIM



1. CI staat voor Continuous Integration: het continu en automatisch integreren en testen van code aanpassingen. CD staat voor Continuous Delivery: het continu en automatisch opleveren van software aanpassingen naar test- en productie omgevingen. [↑](#footnote-ref-1)
2. **Extra toelichting**: In de GIA staan de afspraken voor doorleveren vanuit de broneigenaar. Denk hieraan; wie mag de gegevens wel hebben, wie niet en wanneer moet hierbij de toestemming van de broneigenaar gevraagd worden. [↑](#footnote-ref-2)
3. **Extra toelichting**: Bij de GLA worden privacy aspecten (zoals doelbinding, proportionaliteit en subsidiariteit) en leveringsvoorwaarden getoetst. Bij een bestandslevering wordt door de afnemer aangegeven wanneer de gegevens vernietigd worden. Dit is opgenomen in de GEB van de afnemer. [↑](#footnote-ref-3)