



UITVOERING WERKNEMERSVERZEKERINGEN

Interface-standaarden (bron-DIM)

DataFabriek

Versie 2.0

Datum: 19 maart 2021

VERSIEBEHEER

Versie	Datum	Status	Korte beschrijving aanleiding / wijziging
0.1	29-11-2018	Concept	Eerste concept, H1-5 ter review aan DA/DBA DWH
0.2	11-12-2018	Concept	Opmerkingen DWH m.b.t. H1-5 verwerkt (is nu H1-6, plus H12 en H13). Structuur aangepast. Eerste versie technische eisen db-links toegevoegd (H9)
0.9a	3-01-2019	Concept	Opmerkingen tweede review DWH verwerkt.
0.9b	10-1-2019	Ter afstemming	Ter bespreking binnen stuurgroep.
0.9c	31-3-2019	Project intern	Tussenversie t.b.v. werkverdeling architectuurteam
0.9d	30-4-2019	Voor akkoord SG	Opmerkingen op 0.9b verwerkt, technische details, voorzover onafhankelijk van ETL-tool, toegevoegd. Daarmee afgerond conform DoD van P43 voor DataFabriek Fase 2.
1.0	10-5-2019	Definitief	Ongewijzigd akkoord
1.1	30-9-2019	Voor team review	Technische details toegevoegd (zie § 1.4)
1.2	2-10-2019	Definitief	Na team review
1.3	3-14-2020	Update	Update pakbon
1.4	6-23-2020	Update	Update pakbon, update par file, combinatie dataleveringen met verschillende leveringstypes
1.5	28-8-2020	Voor team review	Technische details toegevoegd (zie § 1.4)
1.6		Definitief	Na team review
1.9	8-2-2021	Update	Dakpan levering, Beslisbomen, meerdere tekstuele verduidelijkingen/verbeteringen, code tabellen, GLO vervangen door GIA
1.91	17-2-2021	Update	Commentaar Gert-Jan Kooren verwerkt
1.92	11-03-2021	Update	Verwerken commentaar DWH-TDA
2.0	19-03-2021	Definitief	Na goedkeuring TDA DWH

Versiebeheer: *Gert-Jan Kooren*Opgesteld door: *Gert-Jan Kooren, i.s.m. Lex Groot en Richard Hogenberg*

Afgestemd met:	Status	Datum	Naam / Contactpersoon
Projectleider DataFabriek	Afgestemd	19-12-2018	Maarten-Jan de Groot
Domeinarchitect GD	Goedgekeurd	9-1-2019	Ralph Foorthuis
DA DWH (en DBA DWH)	Goedgekeurd	9-1-2019	Bas Marsman
AB+	Afgestemd	5-2-2019	Ralph Foorthuis
BICC Werk	Afgestemd	13-2-2019	Tineke Kras / Michel Verhaaf
SI	Afgestemd	6-3-2019	Mark Verlinden
Werk	Afgestemd	23-4-2019	André Schrande
TDA DWH	Goedgekeurd	18-3-2021	Gerald van Leeuwen

Besproken door:	Status	Datum
Stuurgroep	Versie 0.9b aangehouden voor overleg met achterland. De daaruit volgende opmerkingen (van Handhaven en Werk) zijn verwerkt in versie 0.9d.	31-1-2019

Architectuur Board	Geen opmerkingen ontvangen	5-2-2019
CIO Office		

Distributie: Leden Architectuur Board en AB+
Leden MT DWH
Leden Stuurgroep

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	4
1 INLEIDING	8
1.1 Doel van dit document	8
1.2 Scope van deze standaarden.....	8
1.2.1 Applicatie-perspectief	8
1.2.2 Aandachtsgebieden	8
1.3 Structuur van dit document	8
1.4 Opmerkingen bij deze versie.....	9
2 EISEN AAN GEGEVENS EN GEGEVENSLEVERINGEN.....	10
2.1 G1 - De gegevens zijn volledig	11
2.2 G2 - De gegevens zijn gedefinieerd.....	11
2.3 G3 - De gegevens zijn herleidbaar	12
2.4 L1 - De levering is compliant.....	12
2.5 L2 - De levering is robuust.....	13
2.6 L3 - De levering is toekomstvast	13
3 STANDAARDVEREISTEN M.B.T. TE LEVEREN GEGEVENS	14
3.1 Geleverde objecten zijn volledig	14
3.1.1 Gegevensobjecten o.b.v. behoefte, gegevensattributen o.b.v. toekomstige behoefte	14
3.1.2 Geen afnemer-specifieke filters.....	14
3.1.3 Geen kwaliteitsfilters	14
3.1.4 Vertrouwelijkheidsfilters alleen op stabiele regels.....	15
3.1.5 Ook afgeleide waarden	15
3.1.6 Geen aggregatie	15
3.1.7 Diepste business detail	15
3.1.8 Inclusief vertrouwelijkheidsdetails	15
3.2 Geleverde historie is volledig.....	15
3.2.1 Toestand, mutatie én versie	15
3.2.2 Ook verwijderingen	15
3.2.3 Datum én tijd	15
3.2.4 Alle audit-velden.....	16
3.2.5 Alle peilmomenten	16
3.3 Gegevens zijn identificeerbaar.....	16
3.3.1 Primaire sleutels beschikbaar	16
3.3.2 Business sleutels beschikbaar	16
3.3.3 Alleen gegevens met productie-status	16
4 STANDAARDVEREISTEN M.B.T. LEVERINGEN.....	17
4.1 De levering levert herleidbare gegevens	17
4.1.1 Geautomatiseerde levering.....	17
4.1.2 Logische push op een veilig moment	17
4.1.3 Volledige tracking-gegevens	17
4.2 De levering is veilig	17

4.2.1	Afscherming aan de bron	17
4.2.2	Veilig transport.....	17
4.3	De levering is fout-tolerant	17
4.3.1	Geen type casts.....	18
4.3.2	Controleerbaar tegen specificaties.....	18
4.3.3	Controleerbaar op volledigheid en volgordelijkheid	18
4.3.4	Reproduceerbaar	18
4.4	De levering is bestand tegen wijzigingen	18
4.4.1	Ontkoppeld model.....	18
4.4.2	Brede ontsluiting.....	18
4.4.3	Meer data	19
4.4.4	Hogere frequentie	19
4.4.5	Code tabellen	19
5	STANDAARDEN M.B.T. LEVERINGSWIJZEN HISTORIE	21
5.1	Historisch gedrag brongegevens	21
5.2	Vergelijking leveringswijzen historie.....	22
5.3	Inzet leveringswijzen	22
5.3.1	Event-gedreven levering	26
5.3.2	Stapelbare levering	26
5.3.3	Incrementele levering.....	27
5.3.4	Volledige levering.....	28
6	STANDAARDEN VOOR DE PAKBON.....	30
6.1	Tracking-gegevens	30
6.2	Controle-gegevens.....	30
7	STANDAARDEN M.B.T. LEVERINGSTECHNIEK	32
7.1	Vergelijking leveringstechnieken.....	32
7.1.1	Afhankelijkheid tussen technische implementaties bron en DIM	33
7.1.2	Eisen aan beschikbaarheid bron ("uptime")	33
7.1.3	Ontwikkeltijd interface (bronzijde)	33
7.1.4	Ontwikkeltijd interface (DIM-zijde).....	34
7.1.5	Archivering geleverde gegevens (bij de bron)	34
7.1.6	Ondersteuning "logische push"	34
7.1.7	Impact op de bron performance.....	34
7.2	Inzet leveringstechnieken voor het DIM.....	35
7.2.1	Levering via berichten	36
7.2.2	Levering via bestanden.....	37
7.2.3	Levering via database-link.....	38
7.2.4	Levering via database-link vanuit bron-kopie.....	39
7.2.5	Levering via export/import	39
7.2.6	Levering via CDC	40
7.3	Terugkoppeling vanuit het DIM.....	40
8	STANDAARDEN VOOR BERICHT-GEBASEERDE INTERFACES	42
9	STANDAARDEN VOOR BESTANDSLEVERINGEN	43
9.1	Basis-opzet.....	43
9.2	Data-bestand	43
9.2.1	Technisch formaat.....	43
9.2.2	Eén recordtype per bestand.....	43

9.2.3	<i>Grote objecten</i>	43
9.2.4	<i>Overige punten:</i>	44
9.3	Pakbon.....	44
9.3.1	<i>Afwijkend pakbon-formaat</i>	45
9.3.2	<i>Voorbeeld pakbon-formaat</i>	45
9.4	Volgorde levering bestanden	46
9.5	Bestandsnamen.....	47
10	STANDAARDEN VOOR DATABASE-LINKS	49
10.1	Basis-opzet.....	49
10.2	Ontkoppel-views.....	49
10.3	Technische inrichting	50
10.3.1	<i>Bron-zijde</i>	50
10.3.2	<i>DIM-zijde</i>	50
10.4	Pakbon info.....	51
10.5	Beschikbaarheid en vrijgave voor leesactie	51
10.6	Ondersteuning herlevering	51
10.7	Inzet bron-kopie.....	51
11	STANDAARDEN VOOR EXPORT/IMPORT	52
11.1	Basis-opzet.....	52
11.2	Pakbon info.....	52
11.2.1	<i>Technische inrichting</i>	53
11.3	Volgorde levering bestanden	55
11.4	Bestandsnamen.....	55
11.5	Ondersteuning herlevering	56
11.6	Inzet bron-kopie.....	56
12	STANDAARDEN VOOR CDC.....	57
13	STANDAARDEN M.B.T. BEHEERSPROCESSEN	58
13.1	Bewaking keten-integriteit.....	58
13.1.1	<i>Bewaking keten-integriteit bij event-gedreven levering</i>	58
13.1.2	<i>Bewaking keten-integriteit bij stapelbare levering</i>	58
13.1.3	<i>Bewaking keten-integriteit bij incrementele levering</i>	59
13.2	Missende levering	59
13.3	Herlevering.....	60
14	STANDAARDEN VOOR TE LEVEREN GEGEVENSDEFINITIES.....	62
14.1	Definities.....	62
14.1.1	<i>Gegevensdefinities eenduidig vastgelegd</i>	62
14.1.2	<i>Formaat vastgelegd</i>	62
14.1.3	<i>Sleutels gedefinieerd</i>	62
14.1.4	<i>Historisch gedrag gedefinieerd</i>	62
14.2	Gebruiksgrenzen	63
14.2.1	<i>Vertrouwelijkheid gedefinieerd</i>	63
14.2.2	<i>Maskerings-eisen gespecificeerd</i>	63
14.2.3	<i>Gebruiksbeperkingen gedefinieerd</i>	63
15	STANDAARDEN VOOR SPECIFICATIES.....	64
15.1	Gegevenslijst ("RLO")	64
15.2	FO en TO.....	64

16	STANDAARDEN VOOR GIA'S	65
	BIJLAGE A: AFKORTINGEN	66
	BIJLAGE B: BEKENDE INTERFACE UITZONDERINGEN.....	67
	16.1 Dakpan Levering	67

1 INLEIDING

1.1 Doel van dit document

De rol van het Data Integratie Magazijn (DIM) binnen het applicatielandschap van UWV, en de wet- en regelgeving (intern en extern) die daarop van toepassing is, stellen eisen aan het DIM zelf, maar ook aan de interfaces vanuit de bronnen naar dat DIM.

Dit document beschrijft de standaarden die uit deze eisen aan de bron-interfaces zijn afgeleid.

1.2 Scope van deze standaarden

1.2.1 Applicatie-perspectief

Deze standaarden zijn primair van toepassing op interfaces van UWV-bronsystemen naar het DIM.

Waar mogelijk¹ worden ook nieuwe interfaces naar de bestaande DWH's conform de standaarden ingericht, zodat deze bij latere migratie niet aangepast hoeven te worden.

Voor interfaces tussen bronsystemen buiten UWV en het DIM zijn de standaarden slechts bruikbaar als "ideaalbeeld". De achterliggende eisen blijven echter onverminderd van toepassing.

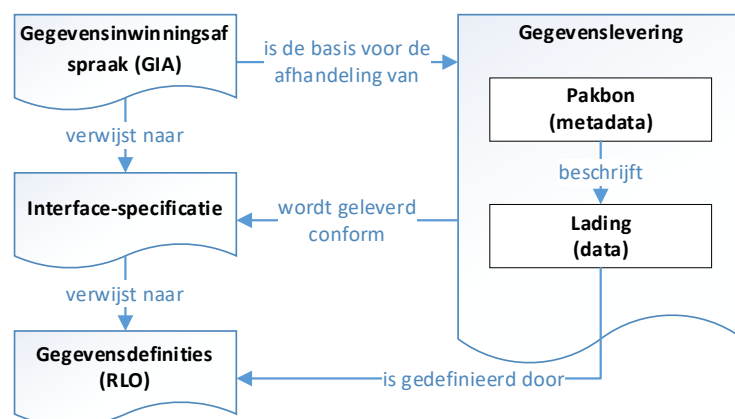
De standaarden zijn **niet** van toepassing op interfaces tussen andere UWV-systemen, maar kunnen wel als beginpunt gebruikt worden bij het, in de toekomst, definiëren van een dergelijke meer algemene standaard.

1.2.2 Aandachtsgebieden

Deze standaarden dekken functionele, technische, documentaire en procedurele eisen aan gegevensleveringen van bron naar DIM.

Ze betreffen:

- De via de interface te leveren gegevens (de "interface-lading")
- De via de interface te leveren metadata (de "interface-pakbon")
- De fysieke implementatie van de interface
- De diepgang waarmee en wijze waarop de functionaliteit van de interface vastgelegd.
- De diepgang waarmee en de wijze waarop definities van de via de interface geleverde gegevens worden vastgelegd.
- De noodzakelijke procesafspraken rondom de interface en de diepgang waarmee en wijze waarop deze in gegevensleveringsovereenkomsten (GIA's) worden vastgelegd.



1.3 Structuur van dit document

Dit document beschrijft allereerst de globale eisen die het DIM stelt aan interfaces en de erin geleverde (meta)data (hoofdstuk 2 - Eisen aan gegevens en gegevensleveringen).

De hieruit volgende functionele eisen voor de in de interface te leveren (meta)data worden beschreven in de hoofdstukken 3 t/m 6 (respectievelijk Standaardvereisten m.b.t. te leveren gegevens, Standaardvereisten m.b.t. leveringen, Standaarden m.b.t. leveringswijzen historie en Standaarden voor de pakbon).

In de hoofdstukken 7 t/m 12 worden, per leveringstechniek (berichten, bestanden, database-links, export/import, CDC), de technische eisen beschreven.

Hoofdstuk 13 beschrijft de standaarden m.b.t. beheersprocessen.

De hoofdstukken 14 t/m 16, tenslotte, beschrijven de eisen betreffende de vastlegging van gegevens- en interface-specificaties en van proces-afspraken.

1.4 Opmerkingen bij deze versie

Mocht er tijdens het (bouw)proces geconstateerd worden dat er verdere detaillering nodig is, dan wordt dit document daarmee aangevuld in een nieuwe versie. De aanvulling is grotendeels afhankelijk van de technische (on)mogelijkheden van zowel de bronnen als de ETL-tooling.

Geel-gemarkeerde tekst is nieuw ten opzichte van de voorgaande versie.

Open punten:

- Verwijder tabellen t.b.v. volgen DLM bron. Hoe aan te leveren (invulling dient uitgezocht te worden)
- De beslisboom in paragraaf 7.2 heeft een nieuwe update. Preferentie van leveringsmethode dient actueel te worden gemaakt. DB-link alleen als er géén (of een voor iedereen, dus ook rekencentrum, acceptabele) performance impact op de bron is. Waarschijnlijk betekent dit dat er alleen DB-links gebruikt worden op replica's (kopie-databases, andere DWH's etc.). Levering van NIEUWE bronnen bij voorkeur via bestanden, mits deze voorkeursformaat hebben / door harnas kunnen worden verwerkt (extra actie": voorkeursformaat in wel een bijlage van deze IFS opnemen)
- Paragraaf 7.3 zal verder worden uitgedetailleerd zodra het helder is hoe dit vormgegeven gaat worden.

¹ Ter beoordeling van "lijn"-organisatie DWH

2 EISEN AAN GEGEVENS EN GEGEVENSLEVERINGEN

Het Data Integratie Magazijn (DIM) dient o.a. om de eventuele beperkingen van de bronsystemen m.b.t. historisch besef en de beschikbaarheid van gegevens voor het gebruik binnen het gegevensverwerkingsdomein "Data Integratie en Analyse" (DIA) te "compenseren".

Dit wordt gerealiseerd door gegevens uit de bronsystemen in kopie (en zonder verdere bewerkingen) in de "bron-zone" van het DIM op te slaan. Als een bronsysteem geen gegevenshistorie bijhoudt (m.a.w. bij wijziging van gegevens de oude waarden overschrijft), dan wordt deze gegevenshistorie in de bron-zone alsnog opgebouwd.

De gegevens in de bron-zone moeten, voor DIA-gebruik, kunnen worden beschouwd als "identiek" aan die in bron. Hieruit volgt een aantal eisen m.b.t. de aan het DIM geleverde gegevens en de wijze waarop deze geleverd worden.

Daarnaast stellen ook wet- en regelgeving (m.n. de AVG) en het eigen UUV-beleid t.a.v. gegevensbeheer eisen aan replicatie naar en opslag in het DIM.

Al deze vereisten kunnen in zes "hoofdeisen" worden samengevat; drie eisen aan de door de bronsystemen te leveren gegevens, en drie eisen aan de wijze waarop die gegevens geleverd worden:

G1 - De gegevens zijn volledig

De aan het DIM geleverde gegevens moeten de inhoud van het bronsysteem in volledig detail weergeven; er mag geen sprake zijn van verlies van informatie tussen bron en DIM.

G2 - De gegevens zijn gedefinieerd

Voor alle aan het DIM geleverde gegevens moet duidelijk zijn wat ze betekenen en op welke wijze ze uit de bron zijn geëxtraheerd.

G3 - De gegevens zijn herleidbaar

De aan het DIM geleverde gegevens moeten te herleiden zijn op (een momentopname van) de gegevens in het bronsysteem, en deze herleidbaarheid moet, in data en definities, zijn vastgelegd.

L1 - De levering is compliant

De gegevensleveringen aan het DIM moeten voldoen aan wet- en regelgeving, en moeten daarnaast voldoende (meta)data bevatten om ook de compliance van het DIM mogelijk te maken.

L2 - De levering is robuust

De wijze van levering moet zo zijn ingericht dat foutsituaties (a) zo min mogelijk optreden, en (b) bij eventueel optreden óf kunnen worden verwerkt zonder dat ze schade aanrichten aan de gegevens in het DIM, óf kunnen worden herkend voordat ze schade aanrichten aan die gegevens.

L3 - De levering is toekomstvast

De levering moet zo zijn ingericht dat de kans op wijzigingen erop (als gevolg van een gewijzigde situatie bij de bron of gewijzigde vereisten vanuit het DIM of haar afnemers) zo gering mogelijk is.

2.1 G1 - De gegevens zijn volledig

De aan het DIM geleverde gegevens moeten de inhoud van het bronsysteem in volledig detail weergeven; er mag geen sprake zijn van verlies van informatie tussen bron en DIM.

Rationale:

De in het DIM opgeslagen gegevens moeten, binnen het DIA-gegevensverwerkingsdomein, gebruikt kunnen worden als een betrouwbare weergave van de brongegevens en hun wijzigingsgeschiedenis.

Impact:

Geen afnemer-specifieke filters; de geleverde gegevens mogen geen afnemer-specifieke subset van de gegevens in het bronsysteem zijn.

De extractie van de gegevens uit het bronsysteem kent geen uitval; ook incorrecte en/of onverwachte data moet aan het DIM geleverd worden.

Geen verlies van granulariteit; gegevens moeten ongeaggregeerd aan het DIM geleverd worden.

Geen verlies van historie; alle in het bronsysteem beschikbare wijzigingsinformatie (transactiemomenten, begin- en einddata, etc.) moeten aan het DIM worden meegeleverd.

2.2 G2 - De gegevens zijn gedefinieerd

Voor alle aan het DIM geleverde gegevens moet duidelijk zijn wat ze betekenen en op welke wijze ze uit de bron zijn geëxtraheerd.

Rationale:

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.

Daarnaast zijn eenduidige gegevensdefinities 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.

Impact:

Van elk gegevenselement in een broninterface moeten de technische én de functionele gegevensdefinitie bekend zijn.

Van elk gegevenselement in een broninterface moet het historisch gedrag bekend zijn.

Van elk gegevenselement in een broninterface moet bekend zijn op welk moment het uit de bron is geëxtraheerd, of tenminste tot op welk moment wijzigingen erop in de extractie zijn verwerkt.

2.3 G3 - De gegevens zijn herleidbaar

De aan het DIM geleverde gegevens moeten te herleiden zijn op (een momentopname van) de gegevens in het bronsysteem, en deze herleidbaarheid moet, in data en definities, zijn vastgelegd.

Rationale:

De in het DIM opgeslagen gegevens moeten kunnen voldoen aan de hoogste afnemerseisen m.b.t. consistentie met de bron.

Daarnaast is herleidbaarheid noodzakelijk om datakwaliteits- en leveringsproblemen op correcte wijze terug te kunnen melden aan het bronsysteem.

Impact:

De extractie van de gegevens moet gedocumenteerd en volledig geautomatiseerd zijn.

Extractie moet gebeuren op een, door de bron bepaald, "veilig moment" gebeuren; een moment waarop de databron logisch consistent² is.

Het moment van extractie en de eventueel bij de extractie gebruikte parameters moeten met de extractie worden meegeleverd.

Bijzondere vormen van extractie (bv. een gecorrigeerde levering) moeten als zodanig herkenbaar zijn.

2.4 L1 - De levering is compliant

De gegevensleveringen aan het DIM moeten voldoen aan wet- en regelgeving, en moeten daarnaast voldoende (meta)data bevatten om ook de compliance van het DIM mogelijk te maken.

Rationale:

Een groot deel van de gegevens in het DIM (en daarmee ook in de gegevensleveringen aan het DIM) kan gecategoriseerd worden als persoonsinformatie. Deze vallen daarmee onder de AVG, en dienen daarom te voldoen aan het "privacy by design" principe.

Daarnaast stellen zowel AVG als andere wet- en regelgeving (bv. Wet SUWI) grenzen aan het gebruik van data in (en het leveren van data uit) het DIM.

Impact:

De levering moet zo zijn ingericht dat deze geen onaanvaardbare beveiligings-risico's oplevert, noch aan de bron-kant, noch aan de DIM-kant, noch tijdens transport van bron naar DIM.

De aan de levering gerelateerde metadata moeten voldoende "rijk" zijn om de geleverde gegevens automatisch (dus zonder menselijke tussenkomst en daaruit volgende niet-noodzakelijke gegevenstoegang) te kunnen controleren, verwerken en, waar nodig, te maskeren³.

Voor alle gegevens in de levering moeten de gebruiksbeperkingen, de gevoeligheidskwalificatie en de eventuele noodzaak tot maskeren bekend zijn.

Om het data lifecycle management (DLM) van de bron te kunnen repliceren in het DIM, moet ook het verwijderen van gegevens uit de bron herkenbaar zijn in de gegevenslevering.

² Afhankelijk van het type gegevens in de bron betreft dit consistentie op het niveau van de geleverde objecten of op het niveau van de volledige geleverde gegevensset.

³ De term "maskeren" wordt in document gebruikt voor zowel anonimiseren als pseudonimiseren. technisch zijn deze processen vaak gelijk; de status (anoniem of pseudoniem) van het gemaskeerde resultaat is vooral afhankelijk van de aanwezigheid van persoonsgegevens elders in de organisatie.

2.5 L2 - De levering is robuust

De wijze van levering moet zo zijn ingericht dat foutsituaties (a) zo min mogelijk optreden, en (b) bij eventueel optreden óf kunnen worden verwerkt zonder dat ze schade aanrichten aan de gegevens in het DIM, óf kunnen worden herkend voordat ze schade aanrichten aan die gegevens.

Rationale:

De in het DIM opgeslagen gegevens moeten kunnen worden gebruikt als een betrouwbare weergave van de brongegevens en hun wijzigingsgeschiedenis. Dat betekent dat de juistheid van deze gegevens en hun historie niet mag worden bedreigd door incorrecte leveringen en/of de reparatie daarvan.

Impact:

Gegevensleveringen moeten ook in staat zijn om gegevens met incorrecte en/of onverwachte waarden aan het DIM te leveren

De consistentie (t.o.v. specificatie), volledigheid en volgorde van gegevensleveringen moet automatisch kunnen worden gevalideerd.

Opnieuw samenstellen van een levering en herlevering van deze gecorrigeerde levering moeten, in geval van leveringsfouten, mogelijk zijn.

2.6 L3 - De levering is toekomstvast

De levering moet zo zijn ingericht dat de kans op wijzigingen erop (als gevolg van een gewijzigde situatie bij de bron of gewijzigde vereisten vanuit het DIM of haar afnemers) zo gering mogelijk is.

Rationale

Een zo kort mogelijke "time-to-market"⁴ van nieuwe informatieproducten is een belangrijke eis aan de Datafabriek.

Wijzigingen in leveringen vanuit de bronsystemen hebben impact op zowel de bron als het DIM, en verlengen daarmee, door de afhankelijkheid van de releasekalenders van beide systemen, de time-to-market voor van die wijzigingen afhankelijke informatieproducten aanzienlijk.

Impact:

Leveringen moeten schaalbaar zijn; ze moeten ook bij groeiende gegevenshoeveelheden correct en performant blijven.

Leveringen moeten breed zijn; ze moeten alle gegevenselementen bevatten die potentieel relevant zijn voor afnemers van het DIM, en niet alleen de gegevenselementen waarvoor reeds een benoemde informatiebehoefte bestaat.

De leveringsfrequentie moet omhoog kunnen (bijvoorbeeld van maandelijks naar dagelijks⁵), zonder dat dit wijzigingen in de leveringslogica vereist.

⁴ Met deze term wordt, in deze context, de tijd bedoeld die verloopt tussen het benoemen van een informatievraag en het (kunnen) beantwoorden ervan.

⁵ Niet elke levering wordt hierdoor geraakt: voor sommige brongegevens is een dagelijkse frequentie niet relevant, en voor anderen is dit juist al vanaf de eerste versie van de levering een eis.

3 STANDAARDVEREISTEN M.B.T. TE LEVEREN GEGEVENS

De meeste van onderstaande standaardvereisten volgen direct uit G1 (De gegevens zijn volledig). Waar dit niet (uitsluitend) zo is staat zulks bij het betreffende vereiste vermeld.

De aan het DIM te leveren gegevens verschillen uiteraard per bronsysteem, en een lijst van te leveren gegevens dient dan ook onderdeel te zijn te van de interface-specificatie⁶.

Dit hoofdstuk beschrijft de algemene (bronsysteem-onafhankelijke) vereisten m.b.t. te leveren gegevens. Deze vereisten zijn de basis voor de specifieke gegevenslijst per interface.

In onderstaande worden de volgende termen beschouwd als functioneel equivalente termen binnen verschillende technologieën (met **vet** de standaard gebruikte term):

- **Gegevensobject**, entiteit, tabel en record-type
- Gegevenseigenschap, attribuut, kolom en **veld**
- Gegevensvoorkomen, entiteit-voorkomen, rij en **record**

3.1 Geleverde objecten zijn volledig

3.1.1 Gegevensobjecten o.b.v. behoefte, gegevensattributen o.b.v. toekomstige behoefte

Dit vereiste volgt zowel uit G1 als uit L3 (De levering is toekomstvast).

De bron levert **tenminste** alle gegevensobjecten waaraan de afnemers van het DIM behoefte hebben, maar voor die gegevensobjecten alle velden waaraan de afnemers van het DIM **potentieel** behoefte hebben.

Aansluiting van een bron op het DIM betekent dus **niet** automatisch dat meteen alle gegevensobjecten van die bron moeten worden geleverd, maar **wel** dat voor de geleverde gegevensobjecten (bijna) alle velden moeten worden geleverd.

3.1.2 Geen afnemer-specifieke filters

De geleverde gegevens mogen geen afnemer-specifieke subset van de gegevens in het bronsysteem zijn. Dit geldt zowel voor records als voor velden.

Bron-specifieke filters, bijvoorbeeld om zuiver technische informatie te verwijderen of vanwege incrementele levering, zijn uiteraard wel toegestaan.

3.1.3 Geen kwaliteitsfilters

Ook velden met incorrecte en/of onverwachte waarden (of objecten met daarin velden met incorrecte en/of onverwachte waarden) moeten aan het DIM geleverd worden.

Denk hierbij, bijvoorbeeld, aan velden die niet voldoen aan het ervoor gedefinieerde waardebereik, of velden die niet voldoen aan een afgesproken formaat.

⁶ Zie hoofdstuk 14 (Standaarden voor te leveren gegevensdefinities) en 15 (Standaarden voor specificaties)

3.1.4 Vertrouwelijkheidsfilters alleen op stabiele regels

Het uitfilteren van zeer gevoelige gegevens mag alleen als deze filter-regel over tijd niet kan wijzigen. VIP-data mag dus, bijvoorbeeld, niet worden uitgefilterd, maar medische informatie wel.

3.1.5 Ook afgeleide waarden

Als de bron een eigenschap van een gegevensobject niet fysiek opslaat, maar wél verantwoordelijk is voor de afleiding ervan, dan dient de bron ook (momentopnames van) dit afgeleide gegeven te leveren.

3.1.6 Geen aggregatie

Er mag tussen bron en DIM geen verlies van granulariteit zijn; alle gegevens moeten ongeaggregeerd aan het DIM geleverd worden.

3.1.7 Diepste business detail

Er mag tussen bron en DIM geen verlies van detail zijn; als gegevensobjecten zelf weer opgebouwd zijn uit detailobjecten (bv. de processtappen binnen een proces), dan moeten ook die detailobjecten aan het DIM geleverd worden.

N.B. Deze verplichting geldt **niet** als de detailobjecten alleen technische gegevens bevatten.

3.1.8 Inclusief vertrouwelijkheidsdetails

Als in de bron vertrouwelijkheidsdetails voor de gegevens worden bijgehouden of afgeleid (EP, BGB, VIP, etc), dan dienen ook deze aan het DIM geleverd te worden.

3.2 Geleverde historie is volledig

3.2.1 Toestand, mutatie én versie

Als een bron van een gegeven zowel de toestand op een gegeven moment als de mutaties die daartoe hebben geleid bevat, dan dienen beiden aan het DIM geleverd te worden.

Als een bron zowel de huidige als eerdere versies van gegevenswaarden bijhoudt, dan dienen beiden aan het DIM geleverd te worden.

3.2.2 Ook verwijderingen

Dit vereiste volgt zowel uit G1 als uit L2 (De levering is compliant).

Dit om te kunnen garanderen dat het DIM verwijderingen in de bron correct volgt, en daarmee voldoet aan het in de AVG opgenomen "recht op vergetelheid".

Ook de verwijdering van een gegeven dient aan het DIM gemeld te worden.

3.2.3 Datum én tijd

Alle datum/tijd velden dienen in volledig detail (dus inclusief de tijd) aan het DIM geleverd te worden, ook als de bron voor de tijd altijd een default-waarde gebruikt.

3.2.4 Alle audit-velden

Alle velden die informatie bevatten over wijzigingen in een gegeven (timestamp laatste wijziging, medewerker laatste wijziging, verwijderd, etc.) en geplande wijzigingen in een gegeven (bijvoorbeeld "houdbaarheidsdatums") dienen aan het DIM geleverd te worden.

3.2.5 Alle peilmomenten

Dit vereiste volgt zowel uit G1 als uit G3 (De gegevens zijn herleidbaar).

Van elk gegevenselement in een broninterface moet bekend zijn op welk moment het uit de bron is geëxtraheerd (en/of afgeleid).

Van elk afgeleid gegevenselement in een broninterface moet daarnaast bekend zijn tot op welk moment de input ervoor in de extractie (en/of afleiding) zijn verwerkt.

N.B. Als een peilmoment voor alle records in een interface gelijk is, dan wordt het gegeven bij voorkeur opgenomen in de pakbon (de interface-metadata). Kan het per record verschillen, dan dient het peilmoment als veld binnen dat record geleverd te worden.

3.3 Gegevens zijn identificeerbaar

Deze vereisten volgen zowel uit G1 als uit G2 (De gegevens zijn gedefinieerd) en G3 (De gegevens zijn herleidbaar).

3.3.1 Primaire sleutels beschikbaar

Voor elk geleverd gegevensobject dient een tijds-invariante primaire sleutel gedefinieerd te zijn. Deze sleutel kan, per record, als separaat veld geleverd worden, of gedefinieerd zijn als een combinatie van een aantal geleverde velden.

Bij voorkeur is de geleverde primaire sleutel gelijk aan die in de bron.

N.B. Als de bron geversioneerde gegevens levert, dan is de primaire sleutel van een geleverde object-versie de combinatie van de tijds-invariante primaire sleutel en het startmoment van die versie.

3.3.2 Business sleutels beschikbaar

Als de primaire sleutel een technische is (en daarmee alleen binnen de bron van betekenis), dan dienen de geleverde gegevens voldoende rijk te zijn om, binnen het DIM ook integratie met andere bronnen te ondersteunen.

3.3.3 Alleen gegevens met productie-status

Als de bron gegevens zonder productie-status bevat (bijvoorbeeld dummy-gegevens voor testdoeleinden), dan moeten deze vóór levering (dus aan de bronzijde) zijn uitgefilterd.

4 STANDAARDVEREISTEN M.B.T. LEVERINGEN

De implementatie van de meeste van onderstaande vereisten is afhankelijk van de gebruikte leveringstechniek. Zie hiervoor hoofdstuk 7 en verder.

4.1 De levering levert herleidbare gegevens

Deze vereisten volgen uit G3 (De gegevens zijn herleidbaar).

4.1.1 Geautomatiseerde levering

De extractie van de gegevens moet gebeuren middels een gedocumenteerd en volledig geautomatiseerd proces.

N.B. Dit geldt ook bij het, ingeval van leveringsfouten, opnieuw samenstellen van een levering.

4.1.2 Logische push op een veilig moment

Extractie moet gebeuren op een, door de bron bepaald, "veilig moment"; een moment waarop de databron logisch consistent⁷ is.

N.B. Dit geldt zowel in het geval dat de bron de gegevens naar het DIM verstuurd ("technische push") als in het geval dat het DIM de gegevens bij de bron ophaalt ("technische pull").

4.1.3 Volledige tracking-gegevens

Alles wat de levering uniek maakt (de eventueel bij de extractie gebruikte parameters, het moment van extractie, etc.) moeten met de levering worden meegeleverd.

Bijzondere vormen van extractie (bv. een gecorrigeerde levering) moeten ook als zodanig herkenbaar zijn.

Tracking-gegevens zijn onderdeel van de "pakbon". Zie Hoofdstuk 6 voor meer details.

4.2 De levering is veilig

Deze vereisten volgen uit L1 (De levering is compliant).

4.2.1 Afscherming aan de bron

De bron is verantwoordelijk voor het afschermen van niet voor het DIM relevante gegevens.

4.2.2 Veilig transport

Gegevens worden geleverd via een "veilig kanaal", e.e.a. conform de eisen van IB&P en/of CISO.

4.3 De levering is fout-tolerant

Deze vereisten volgen uit L2 (De levering is robuust), en indirect (vanwege het minimaliseren van niet-noodzakelijke gegevenstoegang) ook uit L1 (De levering is compliant).

⁷ Afhankelijk van het type gegevens in de bron betreft dit consistentie op het niveau van de geleverde objecten of op het niveau van de volledige geleverde gegevensset.

4.3.1 Geen type casts

De levering mag geen type casts (geforceerde formaat-conversies) bevatten, aangezien dit resulteert in impliciete kwaliteitsfilters (zie ook 3.1.3 - Geen kwaliteitsfilters)

4.3.2 Controleerbaar tegen specificaties

De met de levering meegeleverde metadata moeten voldoende "rijk" zijn om de geleverde gegevens automatisch (dus zonder menselijke tussenkomst en daaruit volgende niet-noodzakelijke gegevenstoegang) te kunnen controleren op consistentie met de voor de levering vastgelegde specificaties.

Controle-gegevens zijn onderdeel van de "pakbon". Zie Hoofdstuk 6 voor meer details.

4.3.3 Controleerbaar op volledigheid en volgordelijkheid

De met de levering meegeleverde metadata moeten voldoende "rijk" zijn om de geleverde gegevens automatisch (dus zonder menselijke tussenkomst en daaruit volgende niet-noodzakelijke gegevenstoegang) te kunnen controleren op volledigheid en volgordelijkheid.

Controle-gegevens zijn onderdeel van de "pakbon". Zie Hoofdstuk 6 voor meer details.

4.3.4 Reproduceerbaar

Opnieuw samenstellen van een levering en herlevering van deze gecorrigeerde levering moeten, in geval van leveringsfouten, mogelijk zijn.

N.B. Opnieuw samenstellen is alleen mogelijk indien de bron (een vorm van) transactie-historie opslaat. Is dat niet zo, dan kan de herlevering (doordat deze op een ander peilmoment wordt geëxtraheerd) afwijken van de oorspronkelijke levering. Dit dient in de interface-specificaties te zijn vastgelegd,

4.4 De levering is bestand tegen wijzigingen

Deze vereisten volgen uit L3 (De levering is toekomstvast).

4.4.1 Ontkoppeld model

Een wijziging in het fysieke datamodel van de bron heeft geen directe impact op de structuur van de interface.

N.B. Deze ontkoppeling is bedoeld als "noodrem", en voorkomt slechts dat de release-kalenders van bron en DIM hard gekoppeld zijn, met de daaruit volgende negatieve impact op de agility van de bron. Uiteraard heeft elke wijziging in de bron (fysiek of functioneel) potentieel impact op de erin opgeslagen data, en daarmee op de interface naar het DIM.

4.4.2 Brede ontsluiting

Leveringen moeten alle gegevenselementen bevatten die potentieel relevant zijn voor afnemers van het DIM, en niet alleen de gegevenselementen waarvoor reeds een benoemde informatiebehoefte bestaat.

(Zie ook vereiste 3.1.1 - Gegevensobjecten o.b.v. behoefte, gegevensattributen o.b.v. toekomstige behoefte)

4.4.3 Meer data

Leveringen moeten schaalbaar zijn; ze moeten ook bij groeiende gegevenshoeveelheden correct en performant blijven.

4.4.4 Hogere frequentie

De leveringsfrequentie moet omhoog kunnen (bijvoorbeeld van maandelijks naar dagelijks), zonder dat dit wijzigingen in de leveringslogica vereist.

4.4.5 Code tabellen

Indien er gegevensattributen gebruikt worden welke codes zijn moeten de bijbehorende code tabellen ook onderdeel zijn van de levering. Dit in het kader van brede ontsluiting.

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

Het **Beleid Reference Data UWV**⁸ 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 het DIM zal een administratieve historie van de code tabellen worden opgebouwd.

Codetabel beschikbaar in de bron

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 en periode van geldigheid.

- 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.

Codetabel niet beschikbaar in de bron

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 of hardcode in de applicatie.

⁸ 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.

- 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", zodat de gegevens op een standaardwijze in het DIM kunnen worden ingelezen.

Cross reference lijsten

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", zodat de gegevens op een standaardwijze in het DIM kunnen worden ingelezen.

Reference data spreadsheet

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

1. 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
2. Gegevensdiensten geeft de gegevens door DIM-beheer, die ze verwerken in een "reference data spreadsheet". Dit spreadsheet is geversioneerd langs de geldigheidstijdlijn.
3. DIM-beheer voert tijdens deze verwerking ook technische en inhoudelijke controles uit. Eventuele issues worden teruggekoppeld; DIM-beheer repareert niet zelf.
4. 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.
5. De gegevens worden vervolgens middels een database-link op deze database op standaardwijze verder verwerkt in het DIM.

5 STANDAARDEN M.B.T. LEVERINGSWIJZEN HISTORIE

De eisen L2 (De levering is robuust) en L3 (De levering is toekomstvast) staan, bij het leveren van gegevenshistorie uit de bronsystemen, op gespannen voet met elkaar; de meest robuuste leveringswijze (volledige afslagen van het bronsysteem, en daar in het DIM de historie uit afleiden) is het minst schaalbaar, zowel qua gegevenshoeveelheid als qua leverfrequentie.

Dit hoofdstuk beschrijft hoe, voor elk gegevensobject binnen een levering, de optimale wijze van levering bepaald moet worden.

5.1 Historisch gedrag brongegevens

Het DIM volgt, zo nauwkeurig mogelijk, de gegevenswijzigingen (de "delta's") in de bron. Via welke wijze deze delta's door de bron aan het DIM moeten worden geleverd (of hoe deze delta's door het DIM moeten worden afgeleid) is afhankelijk van de manier waarop de bron wijzigingen opslaat in de bron-database.

Vanuit het perspectief van delta-afhandeling kunnen gegevensobjecten in vijf groepen verdeeld worden:

Niet overschrijfbaar

Een record kan, eenmaal aangemaakt, niet worden gewijzigd, maar alleen worden verwijderd.

Overschrijfbaar

De inhoud van een record kan, na aanmaken, wijzigen. Hierbij wordt de oorspronkelijke inhoud overschreven.

Afgeleid bij opvragen

Het gegeven (het kan zowel gaan om een heel record als om een veld binnen een record) wordt niet opgeslagen in de bron, maar pas berekend als het opgevraagd wordt.

Binnen bovenstaande bestaan twee bijzondere vormen:

Beperkt overschrijfbaar

Van een record kan, eenmaal aangemaakt, maar een zeer beperkt aantal velden worden gewijzigd. Meestal zijn dit velden met (proces)status-informatie.

Geversioneerd

Bij wijzigen van een record worden geen velden overschreven, maar wordt een nieuwe "versie" van het record aangemaakt. De oorspronkelijke versie van het record blijft hierbij als "vorige versie" ook bewaard.

Soms wordt bij het aanmaken van een nieuwe versie aan de "vorige versie" een einddatum/tijd toegevoegd om aan te geven dat deze verlopen is. Soms ook niet, en dan moet de "actieve" versie van een record worden bepaald door te zoeken naar de versie met de laatste begindatum/tijd.

In het eerste geval kunnen de object-versies zelf weer beschouwd worden als "beperkt overschrijfbaar" objecten, in het tweede geval zelfs als "niet overschrijfbaar".

N.B. In bovenstaande is "overschrijfbaarheid" gebruikt als bepalende karakteristiek, en niet "tijdsafhankelijkheid", zoals beschreven in het document "Tijdsdimensies" van het Data Office.⁹

⁹

<https://digitalewerkplek.sharepoint.uwv.nl/documentcenter/Documenten/Gegevensdiensten/Beleidsdocument/Tijdsdimensies%20versie%201.4.pdf>

Dit omdat bij het consistent houden van bron en DIM de technische wijzigbaarheid van een gegeven bepalend is, en niet de functionele.

5.2 Vergelijking leveringswijzen historie

Gegevenswijzigingen kunnen met vier verschillende manieren door de bron aan het DIM geleverd worden:

- **Event-gedreven**
Elke wijziging in een gegeven (elke "gegevensgebeurtenis"; het aanmaken, wijziging of verwijderen van een gegeven) resulteert in het leveren van die wijziging aan het DIM.
- **Volledig**
Periodiek levert de bron alle gegevens zoals opgeslagen in de bron, inclusief eventuele in de bron beschikbare administratieve historie. Het DIM vergelijkt deze met de laatst bekende situatie (zoals opgeslagen in de bron-zone), en leidt daaruit af welke gegevens nieuw, gewijzigd of verwijderd zijn.
- **Incrementeel**
Periodiek levert de bron alle gewijzigde gegevens (sinds de laatste levering). Het DIM vergelijkt deze met de laatst bekende situatie (zoals opgeslagen in de bron-zone), en leidt daaruit af welke gegevens nieuw, gewijzigd zijn¹⁰.
- **Stapelbaar**
Periodiek levert de bron alle nieuwe gegevens (sinds de laatste levering). Al deze gegevens zijn voorzien van een "houdbaarheidsdatum".
Voor niet-overschrijfbaar gegevens is een stapelbare levering bijna hetzelfde als een incrementele levering; het enige verschil is de wijze waarop verwijderingen worden doorgegeven.

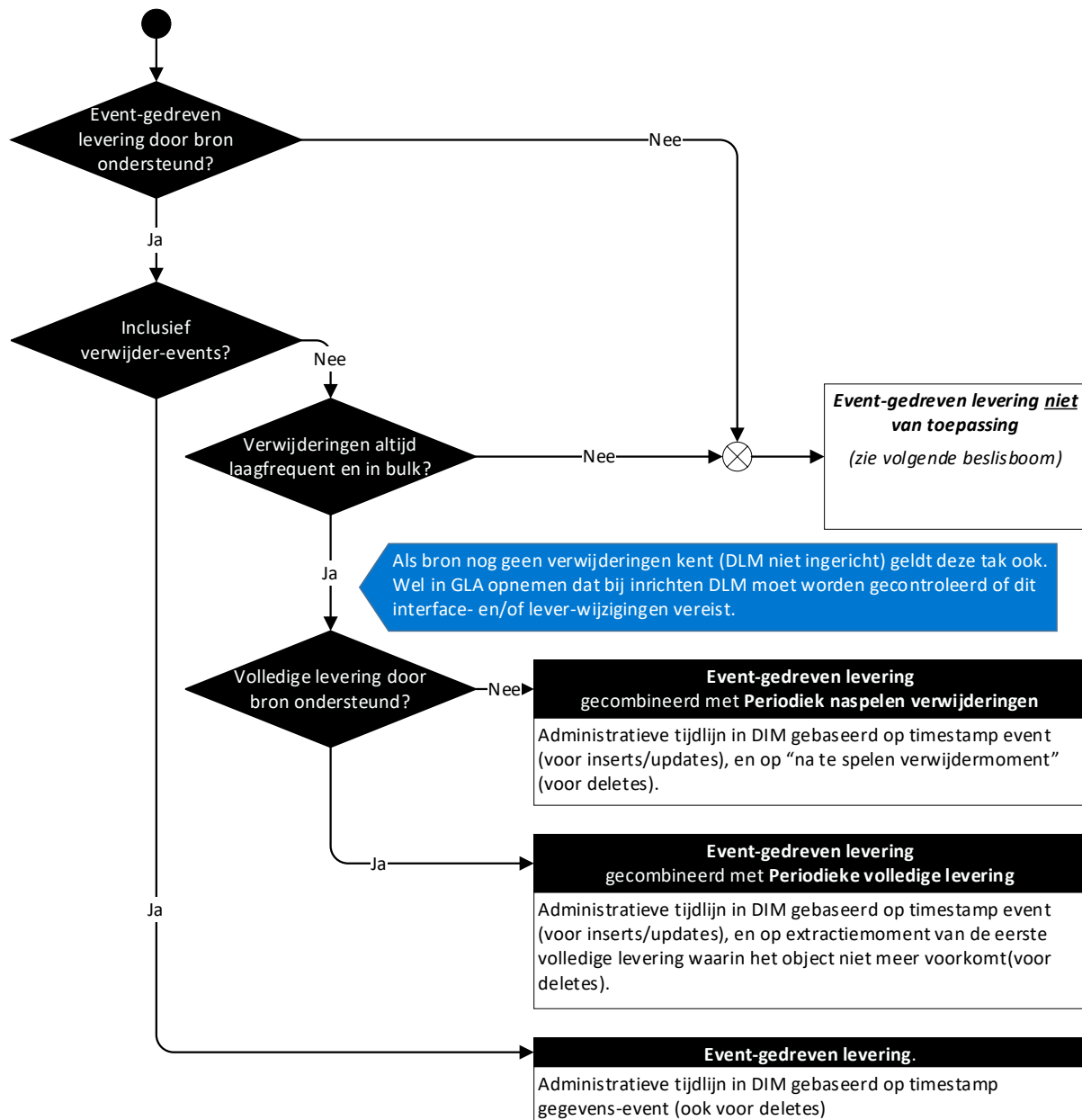
5.3 Inzet leveringswijzen

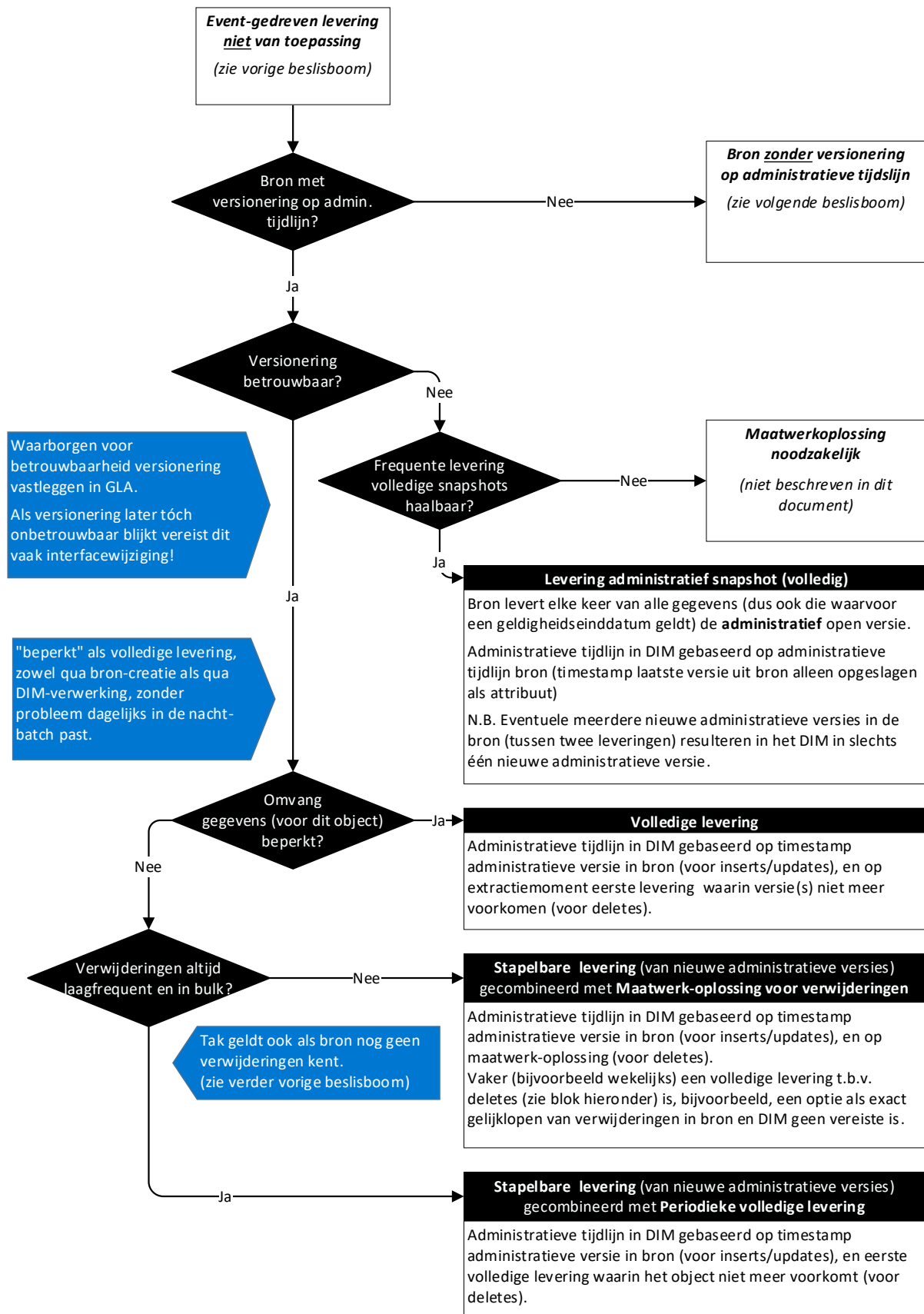
Elke van deze leveringswijzen heeft voor- en nadelen:

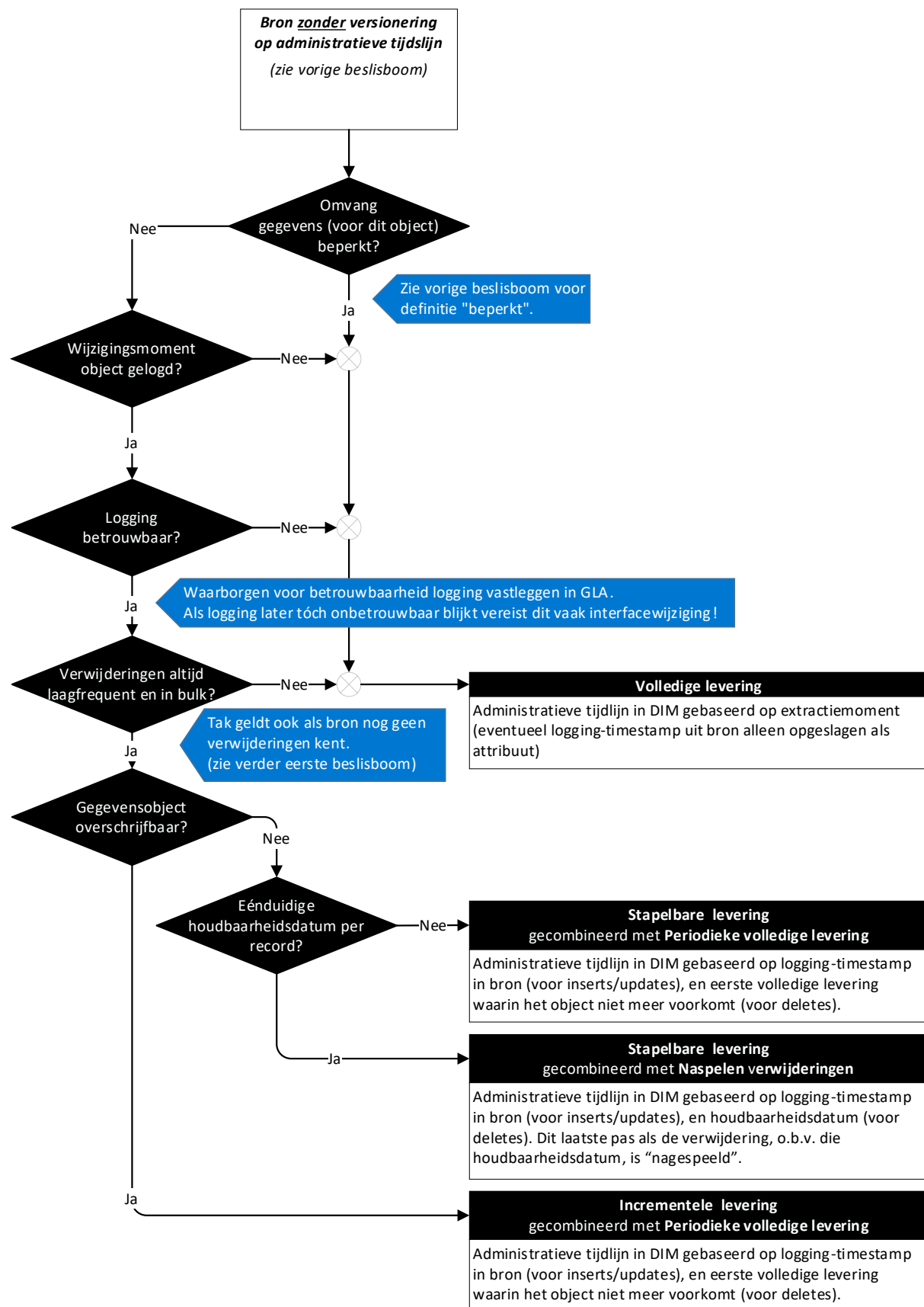
	Event	Volledig	Increment	Stapelbaar
Historie afleidbaar uit levering?	Volledig	Deels	Deels	n.v.t.
Verwijderingen afleidbaar uit levering?	Ja	Ja	Nee	Beperkt
Leverfrequentie	Onbeperkt	Max 1/dag	> 1/dag	> 1/dag
Efficiëntie delta-verwerking	Hoog	Laag	Gemiddeld	Hoog
Zelf reparerend?	Nee	Ja	Nee	Nee
Eisen aan wijzigings-tracking	Hoog	Laag	Hoog	Gemiddeld

Deze combinatie van voor-en nadelen betekent dat er niet één manier van leveren tot standaard kan worden verheven, maar dat dit afhankelijk is van het historisch gedrag van de bron. Onderstaande beslismomen geven de keuzeregels hierbij weer:

¹⁰ Het afleiden van verwijderingen is bij een incrementele levering niet mogelijk. Deze moeten dus op een andere wijze aan het DIM worden doorgegeven (bv. door een periodieke volledige levering).







In de bovenstaande beslisbomen wordt er gesproken over een onbetrouwbare versionering. Onbetrouwbare versionering ontstaat als er in de bron sprake is van versionering (er is een administratieve tijdslijn) maar er wordt iets gewijzigd aan de primaire sleutel van het record.

Hierdoor is het niet meer mogelijk het originele (eerder ontvangen) record te koppelen aan het nieuw ontvangen gewijzigde record. Dit leidt tot o.a. dubbeltellingen.

De betrouwbaarheid van een administratieve tijdlijn zal per tabel in overleg tussen het DIM en de bron worden vastgesteld om te bepalen wat de consequenties zijn.

5.3.1 Event-gedreven levering

Bij een event-gedreven levering resulteert elke wijziging in een gegeven (elke "gegevensgebeurtenis"; het aanmaken, het wijzigen, maar ook het verwijderen van een record) in het leveren van dat record aan het DIM. Daardoor worden alle wijzigingen geleverd aan het DIM (ook als er meerdere wijzigingen per dag zijn, en ook als de bron die wijzigingen niet opslaat) en is het moment van wijzigen altijd herleidbaar in het DIM, zelfs als de bron deze informatie niet opslaat.

Doordat de bron alleen de gegevensgebeurtenissen hoeft te leveren en dat per gebeurtenis doet kan de leverfrequentie hoog zijn; elke levering bevat slechts de data behorende bij één gebeurtenis en is daarmee snel en eenvoudig samen te stellen.

Elke gegevenswijziging kan daarmee (bijna) onmiddellijk beschikbaar zijn voor het DIM. Zelfs als het DIM die wijzigingen niet meteen inleest (maar bijvoorbeeld één keer per dag alle wijzigingen van de voorgaande dag verwerkt) resulteert dit toch in een kleinere vertraging tussen bron en DIM; die dagelijkse verwerking bevat dan namelijk alle wijzigingen tot op het moment van start verwerking binnen het DIM, en niet tot en met start extractie in de bron (meestal enige uren eerder).

De efficiëntie van de delta-verwerking is ook hoog: het DIM ontvangt alleen gegevens die gewijzigd zijn.

Een risico bij event-gedreven levering is dat een gemiste gebeurtenis (aan de bronzijde of in de DIM-verwerking) een blijvende discrepantie tussen bron en DIM oplevert.

Om deze kans zo klein mogelijk te maken is een robuuste tracking (en logging) van wijzigingen in de bron vereist, en daarnaast de ondersteuning van "gegarandeerde levering"¹¹ van bron naar DIM.

Event-gedreven levering heeft daarmee de voorkeur voor bronnen die deze leveringsvorm robuust ondersteunen. In het algemeen zijn dat bronnen die gegevens kunnen leveren via "publish & subscribe" modellen.

5.3.2 Stapelbare levering

Bij een stapelbare levering levert de bron periodiek alle nieuwe gegevens sinds de laatste levering. Al deze gegevens zijn, indien mogelijk, voorzien van een "houdbaarheidsdatum".

Dit type levering is alleen relevant voor gegevens die in de bron per definitie "niet overschrijfbaar" zijn¹². De enige mogelijke gegevenswijzigingen zijn de "aanmaken" en "verwijderen", waarbij "verwijderen" bovendien alleen gebeurt als de bewaartermijn van het gegeven verstreken is.

Voorbeelden hiervan zijn:

- Boeking die alleen door een "tegenboeking" kunnen worden gewijzigd
- Door een geautomatiseerd versiebeheerssysteem (in de bron) gecreëerde versies zonder achteraf toegevoegde einddatum
- Tijdens de levering uit mutaties afgeleide standen met een expliciet "peilmoment"

¹¹ Engels: "guaranteed delivery"

¹² Dit geldt dus ook voor versies zonder achteraf toegevoegde einddatum

Het moment van invoeren moet, binnen de bron, beschikbaar zijn om te kunnen bepalen in welke levering een gegeven dient te worden opgenomen. Om verwijderingen te kunnen herkennen en naspelen zijn er een aantal mogelijkheden.

- De "houdbaarheidsdatum" is binnen de bron beschikbaar (of afleidbaar), of kan binnen het DIM eenvoudig worden afgeleid; verwijderingen worden namelijk, bij deze leveringswijze, niet expliciet doorgegeven aan het DIM, maar moeten wél door dat DIM kunnen worden "nagespeeld".
- Periodieke volledige leveringen t.b.v. het herkennen van verwijderingen in de bron.
- Het gebruik van een verwijder tabel om expliciet de verwijderingen te kunnen naspelen.

De frequentie van levering kan hoog zijn; hoe hoger de frequentie, hoe minder data er in de levering zit, en dus hoe sneller/eenvoudiger deze (door te bron) is samen te stellen.

Meerdere extracties per dag zijn dus mogelijk.¹³

De efficiëntie van de delta-verwerking is ook hoog: het DIM ontvangt alleen nieuwe gegevens, en verwijdering gebeurt o.b.v. de "houdbaarheidsdatum".

Een risico bij stapelbare levering is dat een gemist nieuw gegeven (aan de bronzijde of in de DIM-verwerking) een blijvende discrepantie tussen bron en DIM oplevert.

Om deze kans zo klein mogelijk te maken dienen stapelbare leveringen alleen te worden ingezet bij brongegevens waarin bij aanmaken van een record automatisch de systeemdatum als begindatum/tijd aan het record wordt toegevoegd. Daarnaast dient de levering voldoende metadata te bevatten om bij laden in het DIM te kunnen controleren dat opeenvolgende leveringen compleet en aaneensluitend zijn.

Een ander risico is dat onverwachte wijzigingen in de bron niet aan het DIM worden doorgegeven. Het gaat dan om:

- a) Invoer in het bronsysteem buiten de tracking/logging om. Het gaat daarbij i.h.a. om reparatie-acties door IT.
- b) Verwijdering anders dan vanwege het verlopen van de "houdbaarheidsdatum", bijvoorbeeld omdat een persoon zich heeft beroepen op het "recht op vergetelheid".

Daarnaast kan een vorm van periodieke controle (bijvoorbeeld via reconciliatie tegen een volledige levering) noodzakelijk zijn om eventuele discrepanties tussen bron en DIM (m.n. door onverwachte of anderszins niet gemelde verwijderingen) te kunnen corrigeren.

Stapelbare levering heeft de voorkeur voor (delen van) bronnen die deze leveringsvorm robuust ondersteunen en waarvoor event-gedreven levering geen optie is. Het gaat daarbij eigenlijk altijd om transacties en versie-objecten met een betrouwbare versionering.

5.3.3 Incrementele levering

Bij een incrementele levering levert de bron periodiek alle gewijzigde gegevens sinds de laatste levering.

Het moment van laatste wijziging moet, binnen de bron, beschikbaar zijn om te kunnen bepalen in welke levering een gegeven dient te worden opgenomen. De frequentie van levering kan hoog zijn; hoe hoger de frequentie, hoe minder data er in de levering zit, en dus hoe sneller/eenvoudiger deze (door te bron) is samen te stellen.

Meerdere extracties per dag zijn dus mogelijk.¹⁴

¹³ Binnen de afzienbare toekomst wordt echter geen leverfrequentie hoger dan dagelijks verwacht.

De efficiëntie van de delta-verwerking is gemiddeld: het DIM ontvangt alleen gewijzigde gegevens, maar moet daarvoor nog wel bepalen wat het type wijziging is.

Een risico bij incrementele levering is dat een gemiste gebeurtenis (aan de bronzijde of in de DIM-verwerking) een blijvende discrepantie tussen bron en DIM oplevert.

Om deze kans zo klein mogelijk te maken is een robuuste tracking (en logging) van wijzigingen in de bron vereist. Ook dient de levering voldoende metadata te bevatten om bij laden in het DIM te kunnen controleren dat opeenvolgende leveringen compleet en aaneensluitend zijn.

Een ander risico is dat wijzigingen in het bronsysteem buiten de tracking/logging om niet aan het DIM worden doorgegeven. Het gaat daarbij i.h.a. om reparatie-acties door IT.

Daarom kan een vorm van periodieke controle (bijvoorbeeld via reconciliatie tegen een volledige levering) noodzakelijk zijn om eventuele discrepanties tussen bron en DIM (door gemiste invoer en/of onverwachte verwijderingen) te kunnen corrigeren. Een dergelijke reconciliatie is vaak sowieso noodzakelijk om binnen het DIM te kunnen afleiden of er gegevens uit de bron verwijderd zijn.

Incrementele levering heeft de voorkeur voor (delen van) bronnen die deze leveringsvorm robuust ondersteunen en waarvan de omvang van de gegevens zodanig is dat levering via volledige afslagen duidelijk minder efficiënt is en/of niet mogelijk is in de vereiste frequentie.

Indien in het bronsysteem gegevens buiten de tracking/logging om kunnen worden gewijzigd, dan heeft incrementele levering alleen de voorkeur als (a) de omvang van de gegevens zodanig is dat levering via volledige afslagen niet mogelijk is in de vereiste frequentie, en (b) de bron alleen gegevens verwijderd middels voorspelbare (en laag-frequente) DLM-processen, en (c) het risico dat gegevenswijzigingen gemist worden acceptabel is voor de gegevensafnemers.

5.3.4 Volledige levering

Bij een volledige levering levert de bron periodiek alle opgeslagen gegevens. Het DIM vergelijkt deze met de laatst bekende situatie (zoals opgeslagen in de bron-zone), en leidt daaruit af welke gegevens nieuw, gewijzigd of verwijderd zijn.

Het DIM vergelijkt deze met de laatst bekende situatie (zoals opgeslagen in de bron-zone), en leidt daaruit af welke gegevens nieuw, gewijzigd of verwijderd zijn. Binnen het DIM kan zo een historie van wijzigingen worden opgebouwd, zelfs als de bron deze informatie niet opslaat. Wel is deze historie nooit nauwkeuriger dan de leverfrequentie. Bij een dagelijkse levering, bijvoorbeeld, zullen meerdere wijzigingen (in de bron) op één dag binnen het DIM gezien worden als één wijziging.

De maximale frequentie van levering is beperkt, aangezien ook bij hogere frequenties de volledige set gegevens (gewijzigd of niet) moet worden geleverd. Typisch is de maximale frequentie dagelijks, maar bij erg grote bronnen kan deze nog lager liggen.

De efficiëntie van de delta-verwerking is laag: het DIM moet alle gegevens (gewijzigd of niet) evalueren om de wijzigingen daaruit af te leiden.

Een voordeel van volledige levering is dat een gemiste gebeurtenis of andere vorm van verkeerde verwerking (aan de bronzijde of in het DIM) géén blijvende discrepantie tussen bron en DIM

¹⁴ Binnen de afzienbare toekomst wordt echter geen leverfrequentie hoger dan dagelijks verwacht.

oplevert. Dit omdat een dergelijke fout "zichzelf repareert" bij de verwerking van de volgende volledige levering.¹⁵

Wel dient de levering voldoende metadata te bevatten om bij laden in het DIM te kunnen controleren dat leveringen in de juiste volgorde geleverd worden.

Volledige levering heeft de voorkeur voor bronnen die geen ondersteuning bieden voor andere vormen van levering. Daarnaast heeft volledige levering de voorkeur voor, qua gegevensomvang, kleine (delen van) bronnen, aangezien andere vormen van levering in dat geval onnodig duur en complex zijn.

Aanvullend op bovenstaande:

Bij implementatie van het DIM zal vaak een "initial load" noodzakelijk zijn. Deze initial load is meestal een volledige levering, al is het soms ook mogelijk om deze te implementeren als een stapelbare of incrementele levering met een hele vroege begindatum.

Volledige levering als correctie middel voor o.a. verwijderingen

Naast bijvoorbeeld incrementele leveringen kunnen er periodieke volledige leveringen worden gebruikt. Dit om eventuele discrepanties (meestal alleen de verwijderingen) tussen bron en DIM te kunnen corrigeren.

De frequentie en of het moment van deze periodieke volledige levering is afhankelijk van verwijderproces bij bron:

- Als dat alleen Archiefwet-DLM is (en dus zeer laag frequent): plannen (ad hoc, geen vast schedule) ná uitvoeren DLM-actie/scripts in de bron (proces om dit te besturen opnemen in GIA)
- Als verwijderproces frequenter en/of onvoorspelbaarder: periodieke volledige levering ook frequenter, en vast gescheduled.

¹⁵ Deze reparatie is overigens niet volledig; een beperkte vervuiling van de administratieve historie blijft bestaan. Deze heeft potentieel impact op rapportages/afleidingen die tussen fout en "reparatie" zijn gegenereerd.

6 STANDAARDEN VOOR DE PAKBON

Metadata die per levering verschilt (of kan verschillen) moet in de "pakbon" van de levering worden meegeleverd. Metadata die kan verschillen per record in de levering moet als veld binnen dat data-record worden geleverd. Metadata die (relatief) stabiel is (en meestal de basis vormt voor gebruik en verwerking van de data) moet in de interface-specificatie worden opgenomen¹⁶.

Niet alle metadata is relevant voor alle leveringswijzen.

De technische invulling van de pakbon verschilt per leveringstechniek Zie hiervoor hoofdstuk 7 en verder.

6.1 Tracking-gegevens

	Event	Stapelbaar	Incrementeel	Dakpan	Volledig
Bron/bronmodule/bron-instantie	X	X	X	X	X
Release/versienummer applicatie ¹⁷	X	X	X	X	X
Release/versienummer extractie-logica	X	X	X	X	X
Volgnummer van de levering		X	X	X	X
Run-ID extractie ¹⁸		X	X	X	X
Extractiemoment	X	X	X	X	X
Begindatum en -tijd van de periode waarvoor wijzigingen geleverd worden		X	X	X	
Einddatum en -tijd van de periode waarvoor wijzigingen geleverd worden ¹⁹		X	X	X	
Eventuele andere voor de levering gebruikte parameters		X	X	X	X
Indicatie of de levering een gecorrigeerde herlevering betreft en, zo ja, van welke eerdere levering ²⁰		X	X	X	X

6.2 Controle-gegevens

	Event	Stapelbaar	Incrementeel	Dakpan	Volledig
Inhoudsopgave ²¹	X	X	X	X	X
Aantal gegevensrecords in de levering		X	X	X	X
Kruistelling op één of meer "kritische gegevenselementen" ²²		(X)	(X)	(X)	(X)

¹⁶ Zie de hoofdstukken 14 (Standaarden voor te leveren gegevensdefinities) en 15 (Standaarden voor specificaties)

¹⁷ Een wijziging in applicatie-versie **kan** impact hebben op de definitie van de geleverde gegevens.

¹⁸ Run-ID kan bijvoorbeeld het technische proces-id zijn waarmee de extractie is uitgevoerd

¹⁹ Als exclusieve grenswaarde; januari loopt dus tot aan 1 februari 00:00, en niet tot en met 31 januari

²⁰ Het Volgnummer van de levering is dan gelijk aan die van de eerdere levering

²¹ De technische invulling hiervan zal verschillen. Voorbeelden zijn XML-tags of JSON-elementnamen voor berichten, header-records voor bestanden, en bestandslijsten voor bestands-verzamelingen.

²² Optioneel; alleen relevant als aan het gegevenselement extreem hoge integriteitseisen gesteld worden.

7 STANDAARDEN M.B.T. LEVERINGSTECHNIEK

7.1 Vergelijking leveringstechnieken

Gegevens kunnen m.b.v. vijf verschillende technieken door de bron aan het DIM geleverd worden:

- via berichten
- via bestanden
- via een database-link
- via een export/import-combinatie (bv. Oracle Data Pump)
- via replicatie m.b.v. "change data capture" (CDC) (bv. Oracle Golden Gate)

Elk van deze technieken heeft zijn voor- en nadelen:

	Berichten	Bestanden	Database-link	Export / Import	CDC
Afhankelijkheid tussen technische implementaties bron en DIM	Volledig ontkoppeld	Volledig ontkoppeld	Deels ontkoppeld	Niet ontkoppeld	Beperkt ontkoppeld
Eisen aan beschikbaarheid bron ("uptime")	Laag (asynchrone koppeling)	Laag (asynchrone koppeling)	Hoog (synchrone koppeling)	Laag (asynchrone koppeling)	Laag (asynchrone koppeling)
Ontwikkeltijd interface (bronzijde)	Hoog	Gemiddeld	Laag	Laag	Gemiddeld
Ontwikkeltijd interface (DIM-zijde)	Hoog	Gemiddeld	Gemiddeld	Gemiddeld	Gemiddeld
Archivering geleverde gegevens	Niet	Eenvoudig	Expliciete DIM-actie vereist	Eenvoudig	Niet
Ondersteuning "logische push"	Volledig	Volledig	Vereist "vrijgave" door bron	Volledig	Volledig
Impact op de bron performance	Laag	Gemiddeld	Gemiddeld tot Hoog	Gemiddeld	Laag

7.1.1 Afhankelijkheid tussen technische implementaties bron en DIM

Deze afhankelijkheid valt uiteen in drie aspecten:

1. De gebruikte opslagtechniek (b.v. relationele database, XML-store, netwerk-database)
2. De leverancier en versie van die opslagtechniek
3. De binnen de opslagtechniek gebruikte datamodellen

	Berichten	Bestanden	Database-link	Export / Import	CDC
Gebruikte opslagtechniek	Volledig ontkoppeld	Volledig ontkoppeld	Bron moet RDBMS zijn	Bron moet RDBMS zijn	Bron moet RDBMS zijn
Leverancier en versie van die opslagtechniek	Volledig ontkoppeld	Volledig ontkoppeld	Oracle en SQL Server ondersteund	Bron en DIM zelfde RDBMS	Oracle en SQL Server ondersteund
Gebruikte datamodellen binnen die opslagtechniek	Volledig ontkoppeld	Volledig ontkoppeld	Deels ontkoppeld	Hard gekoppeld met levering	Hard gekoppeld met levering

Berichten en bestanden zorgen, omdat ze een “tussenformaat” bieden, voor volledige ontkoppeling tussen bron en DIM.

Database-links en CDC eisen wél dat de bron-database een RDBMS is, maar stellen geen eisen aan leverancier/versie van dat RDBMS. Bij CDC is datamodel-ontkoppeling (bijna) niet mogelijk, bij database-links (door inzet van een view-laag en/of een bronkopie) wel.

Export/import is alleen mogelijk als bron en DIM beiden op hetzelfde RDBMS (Oracle) draaien. Oracle-versies kunnen verschillen, maar potentieel vereist dit wijzigingen bij de bron (in de export parameters). Bij export/import is er daarnaast een harde koppeling tussen de fysieke datamodellen van de bron en die van (de landing zone van) het DIM.

7.1.2 Eisen aan beschikbaarheid bron (“uptime”)

Alleen database-links vereisen dat bron en DIM tegelijkertijd “up” zijn op het interface-moment.

Als er historie moet worden afgeleid uit het moment van extractie, dan geldt bij interfacing via database-links bovendien dat dit interfacemoment altijd hetzelfde moet zijn; het is daarnaast ook niet mogelijk om, zonder verlies van historisch detail, extracties uit te stellen (bijvoorbeeld bij gepland onderhoud bij het DIM).

De overige technieken (berichten, bestanden, export/import, CDC) zijn “asynchroon”; het inlezen in het DIM kan op een willekeurig moment na de extractie uit de bron.

7.1.3 Ontwikkeltijd interface (bronzijde)

Berichten/services zijn, zeker als de bron nog niet “service enabled”²³ is, complex en duur om te ontwikkelen. Ook wijzigingen zijn complex, in het bijzonder als er meer dan één afnemer op de berichten is “geabonneerd”.

²³ In deze standaarden wordt een bron als “service enabled” beschouwd als deze:

- standaard interacteert met de buitenwereld via services, én
- het publiceren van gegevensgebeurtenissen via een “publish and subscribe” model ondersteunt.

Het ontwikkelen/wijzigen van bestandsleveringen is, daarmee vergeleken, eenvoudiger en goedkoper, maar vereist nog steeds significante inspanning bij de bron.

Voor database-links, export/import en CDC ligt (bijna) alle ontwikkel-inspanning bij het DIM.

7.1.4 Ontwikkeltijd interface (DIM-zijde)

Aan de DIM-zijde is de verwerking van berichten complex en duur om te ontwikkelen/wijzigen. Het ontwikkelen/wijzigen van de verwerking van de overige leveringstypen is eenvoudiger.

Aandachtsgebied voor export/import (en deels ook CDC) is dat afleidingen niet "on the fly" (tijdens de export) door de bron kunnen worden uitgevoerd. Als de bron deze afleidingen ook niet opslaat kan replicatie van afleidingslogica van bron naar DIM noodzakelijk zijn. In dat geval wordt de ontwikkeltijd "hoog".

7.1.5 Archivering geleverde gegevens (bij de bron)

Archivering van geleverde gegevens is eenvoudig bij bestanden en export/import; in beide gevallen is er sprake van gegevensoverdracht via een bestand.

Bij levering via berichten is archivering alleen mogelijk als UWV hiervoor standaardfaciliteiten biedt. Dit is op dit moment niet het geval. Op zich is dit een beperkt probleem; aangezien bij levering via berichten sowieso een vorm van "gegarandeerde levering" noodzakelijk is.

Bij levering via database-links is archivering bij de bron geen optie, maar ook geen probleem, aangezien het DIM de gegevens zelf ophaalt, en dus ook weet of de levering ontvangen is.

Bij levering via CDC zijn de archiveringsmogelijkheden beperkt, en afhankelijk van de exacte technische inrichting.

N.B. Bij het DIM worden de ontvangen leveringen na verwerking ook, voor een beperkte periode, gearchiveerd. Dit om ze, bij calamiteiten, eventueel opnieuw te kunnen verwerken.

7.1.6 Ondersteuning "logische push"

Alleen bij database-links is sprake van een "technische pull".

7.1.7 Impact op de bron performance

In het algemeen wordt de impact op de performance van de bron door de keuze van het tijdsvenster van de levering zo ingericht dat de impact acceptabel en controleerbaar is.

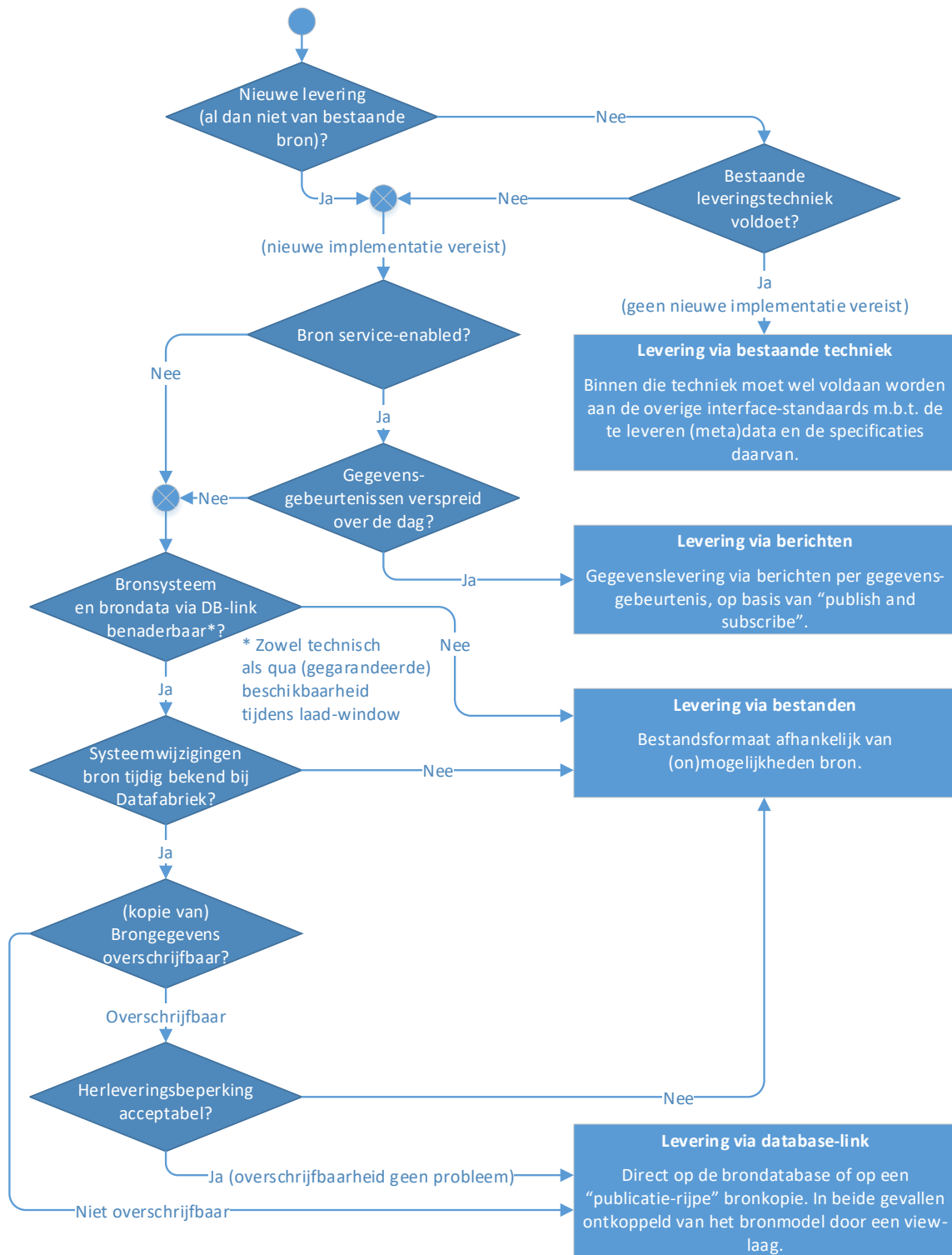
In het geval van berichten en CDC is de impact in het algemeen laag omdat er alleen een actie plaats vindt indien er een actie op record niveau in een tabel in de database plaats vindt. Dit vindt plaats verspreid over de dag en levert dus niet echt een piek belasting op.

Bij bestanden en import export is de impact gemiddeld omdat dit batch verwerkingen. Door het slim kiezen van het tijdsvenster is de impact op de bron acceptabel en controleerbaar.

Bij database-link kan er sprake zijn van een hogere impact om er niet alleen sprake kan zijn van een pull van de gegevens maar ook nog een omgekeerde vergelijking om verwijderingen te kunnen constateren. Het kan dus zo zijn dat tabellen meermaals benaderd worden door verschillende queries. Het tijdstip en het tijdsvenster van toegang is dus een belangrijk gegeven in de GIA vastgelegde afspraken.

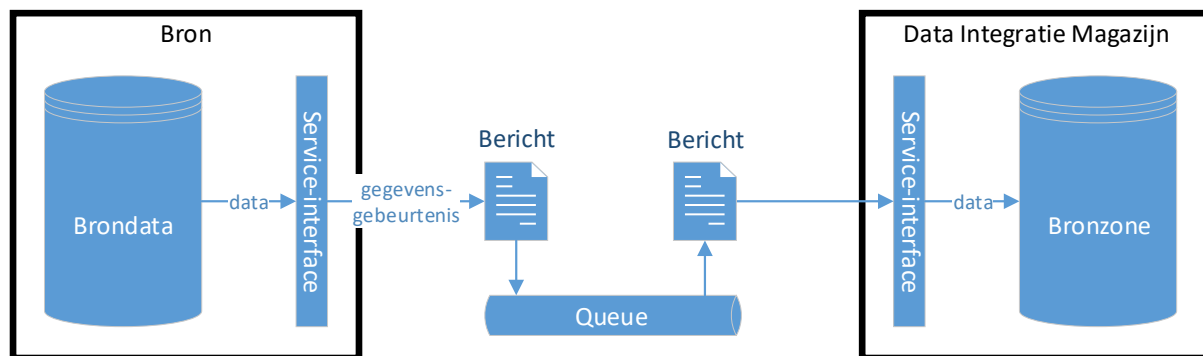
7.2 Inzet leveringstechnieken voor het DIM

Deze combinatie van voor-en nadelen betekent dat er niet één levertechniek tot standaard kan worden verheven, maar dat de voorkeurstechiek afhankelijk is van de (on)mogelijkheden van de bron. Onderstaande beslisboom geeft de keuzeregels hierbij weer:



In de volgende paragrafen worden de diverse opties in meer detail beschreven.

7.2.1 Levering via berichten



Bij levering via berichten levert de bron geen gegevens maar "gegevensgebeurtenissen"; het aanmaken, wijzigen of verwijderen van een gegevensobject.

Deze levering vindt direct na het optreden van de gebeurtenis plaats, in de vorm van een in een queue geplaatst bericht.

De structuur van de gegevens in een bericht is niet hard gekoppeld aan de structuur van die gegevens in de database van de bron, of aan de technische implementatie van die database. Dit resulteert in een "loose coupling" tussen bron en DIM, en daarmee in wijzigingsflexibiliteit. Ook zorgt ervoor dat de bron zowel functioneel als technisch verantwoordelijk is voor de correcte werking van de levering na wijzigingen in het bronsysteem.

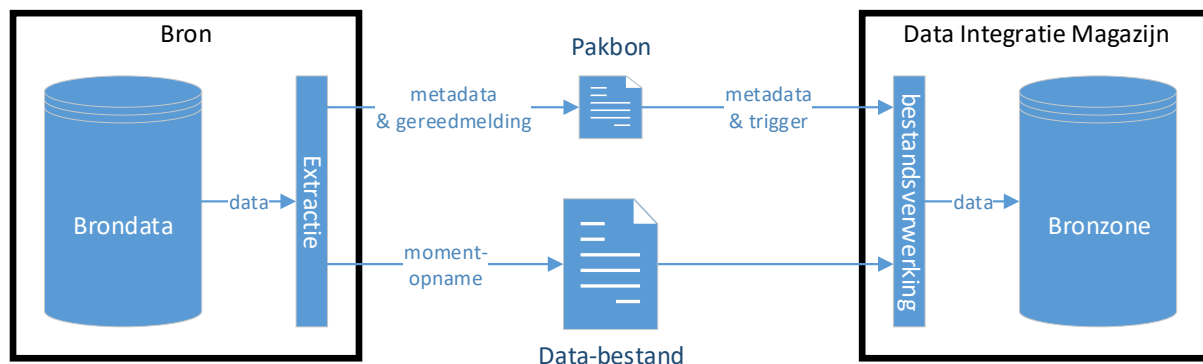
Levering via berichten is echter complex en duur om te bouwen. Vandaar dat deze leveringsvorm alleen de voorkeur heeft als een bron al "service enabled" is, m.a.w. als deze vorm van levering al bij de bron beschikbaar is, en de bron primair via services gekoppeld is met de overige applicaties binnen UWV. Daarnaast moeten deze services qua gegevens-inhoud voldoende "rijk" zijn om te voldoen aan hoofdeis G1 (De gegevens zijn volledig).

Levering via berichten moet voldoen aan de algemene UWV-standaarden op dit gebied²⁴; de Datafabriek stelt geen specifieke technische eisen.

N.B. Soms treden gegevensgebeurtenissen in de bron "in bulk" op, bijvoorbeeld omdat ze het gevolg zijn van een bulk-interface de bron in, of van een bulk-proces binnen die bron. In een dergelijk geval levert levering via berichten extreme piekbelastingen bij de bron op, en verdienen andere leveringstechnieken de voorkeur, zelfs als de bron "service enabled" is

²⁴ Ten tijde van het schrijven van dit document waren deze standaarden nog in ontwikkeling, én waren er nog geen UWV-bronsystemen "service enabled". Deze versie van het document bevat daarom nog geen detaillering van de interface-standaarden bij levering via berichten.

7.2.2 Levering via bestanden



Bij levering via bestanden extraheert de bron op gezette tijden (een deel van) de brongegevens en schrijft deze naar een data-bestand. Dit bestand wordt vervolgens naar het DIM verstuurd (of door het DIM opgehaald). Naast het data-bestand maakt de bron ook een "pakbon" bestand aan. Dit bestand bevat de metadata behorend bij de verstuurde gegevens. De pakbon dient daarnaast als "gereedmelding" (door de bron) van het aanmaken van het databestand, en is daarmee de trigger voor verwerking van dat data-bestand door het DIM.

De structuur van de gegevens in een data-bestand is niet hard gekoppeld aan de structuur van die gegevens in de database van de bron, of aan de technische implementatie van die database. Dit resulteert in een "loose coupling" tussen bron en DIM, en daarmee in wijzigingsflexibiliteit. Ook zorgt ervoor dat de bron zowel functioneel als technisch verantwoordelijk is voor de correcte werking van de levering na wijzigingen in het bronsysteem.

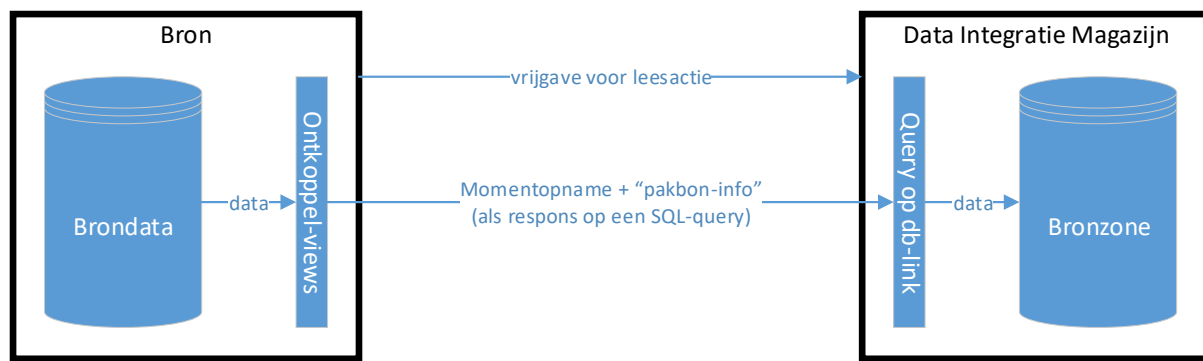
Levering van nieuwe bronnen bij voorkeur via bestanden, mits deze een voorkeursformaat hebben en door de harnessen verwerkt kunnen worden.

Bij levering via een bestand moet de structuur van het bestand voldoen aan de in hoofdstuk 9 beschreven standaarden. De wijze van transport moet voldoen de algemene UWV-standaarden op dit gebied; de Datafabriek stelt hier geen specifieke eisen.

Het geleverde bestand moet bij de bron tenminste beschikbaar blijven totdat het DIM de ontvangst ervan heeft bevestigd.²⁵

²⁵ N.B. Deze verantwoordelijkheid kan ook bij SIP-FT gelegd worden.

7.2.3 Levering via database-link



Bij levering via een database-link extraheert **niet de bron, maar het DIM** op gezette tijden (een deel van) de brongegevens. Dit extraheren mag echter pas plaatsvinden als de bron heeft aangegeven dat zulks "veilig" kan.

Dit is alleen mogelijk als de bron toegang via een dergelijke link ondersteunt²⁶, en de data in die bron via een database-link kan worden getransporteerd²⁷.

Een aandachtspunt is de mate waarin de bron gegevens "versioneert":

- als, in de brondatabase, gegevens overschreven kunnen worden zónder dat daarvan historie wordt bijgehouden, dan is bij **uitvraging** (door het DIM, via de database-link) van een gegevensset bestaande uit meerdere tabellen die set niet per definitie consistent;
- als, in de brondatabase, gegevens overschreven kunnen worden zónder dat daarvan historie wordt bijgehouden, dan is bij **herlevering** (eigenlijk: herbevraging) de respons op de query niet per definitie gelijk aan die van de oorspronkelijke query.

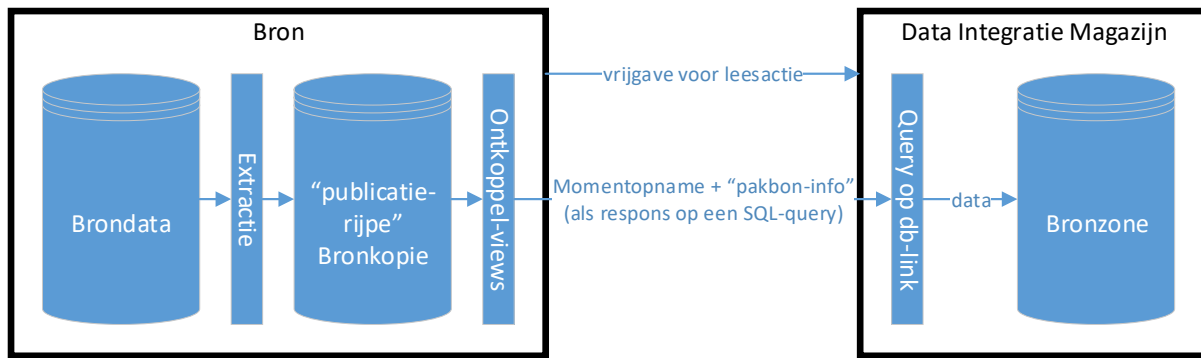
Door aan de bron-zijde van de database-link een view-laag te leggen wordt bereikt dat de structuur van de gegevens in een data-bestand niet hard gekoppeld is aan de structuur van die gegevens in de database van de bron, of aan de technische implementatie van die database. Dit resulteert in een "loose coupling" tussen bron en DIM, en daarmee in wijzigingsflexibiliteit. Ook bij database-links is de bron functioneel verantwoordelijk voor de correcte werking van de levering na wijzigingen in het bronsysteem. De technische verantwoordelijkheid ligt echter deels bij de bron en deels bij het DIM.

Levering via database-links stelt eisen aan de bron (zowel qua opslagvorm als qua beschikbaarheid als qua historische gegevensrijkdom als qua systeemwijzigings-procedures). Wordt daar niet aan voldaan, dan verdient levering via bestanden de voorkeur. Er kan sprake zijn van een serieuze performance impact op de bron. In dat geval heeft de Export/Import de voorkeur als methode van levering.

²⁶ Oracle RDBMS'en zijn eenvoudig vanuit het (Oracle) DIM te benaderen, andere RDBMS'en moeilijker, niet relationele databases niet.

²⁷ Het transport van grote tekstvelden (CLOB's) en grote binaire velden (BLOB's) via een database-link wordt, bijvoorbeeld, maar beperkt ondersteund.

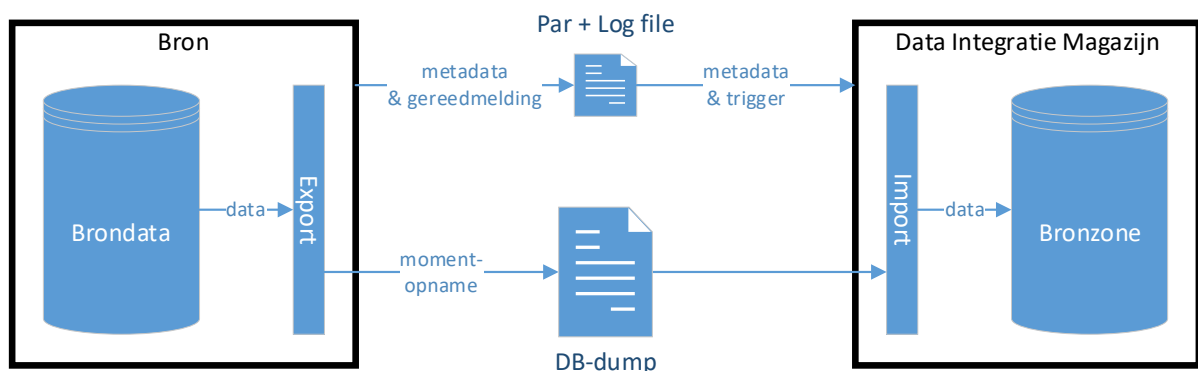
7.2.4 Levering via database-link vanuit bron-kopie



Als een bronsysteem zowel een “primaire” brondatabase als een voor publicatie/interfacing bedoelde kopie-database bevat, dan kan loopt de database-link bij voorkeur van het DIM naar die kopie-database. Vaak is in deze kopie namelijk al een aantal van de data-technische beperkingen van de primaire data “gecompenseerd”. Eis is dan overigens wel dat er, ook binnen het bronsysteem, geen uitval mag zijn tussen primaire database en kopie.

N.B. De hierboven beschreven kopie-database is integraal onderdeel van de bron, en valt dus niet onder beheer van de Datafabriek. Wel is het zinnig als de bron bij het specificeren en implementeren van de kopie de functionele en niet-functionele eisen van DIM en Datafabriek meeneemt. Een voorbeeld hiervan is het opnemen metadata m.b.t. de laatste versie van de kopie-database.

7.2.5 Levering via export/import



Bij levering via export/import exporteert de bron op gezette tijden (een deel van) de bron-database naar een database-dump. Deze dump wordt vervolgens naar het DIM verstuurd (of door het DIM opgehaald) en daar geïmporteerd. De door de exportprogramma's (EXPDP) aangemaakte par- en log-bestanden vervullen hierbij de rol van pakbon.

De structuur van de gegevens in een dump is hard gekoppeld aan de structuur van die gegevens in de database van de bron. Dit resulteert in een “tight coupling” tussen bron en DIM, en daarmee in beperkte wijzigingsflexibiliteit. Daarnaast dienen bron- en doel-RDBMS gelijk te zijn, en moet in sommige gevallen de versie van de doel-database (die van het DIM) bekend zijn bij de bron. Tenslotte is export/import niet bruikbaar als de gegevens in het DIM worden opgeslagen in de (niet-relatieve) datalake-zone.

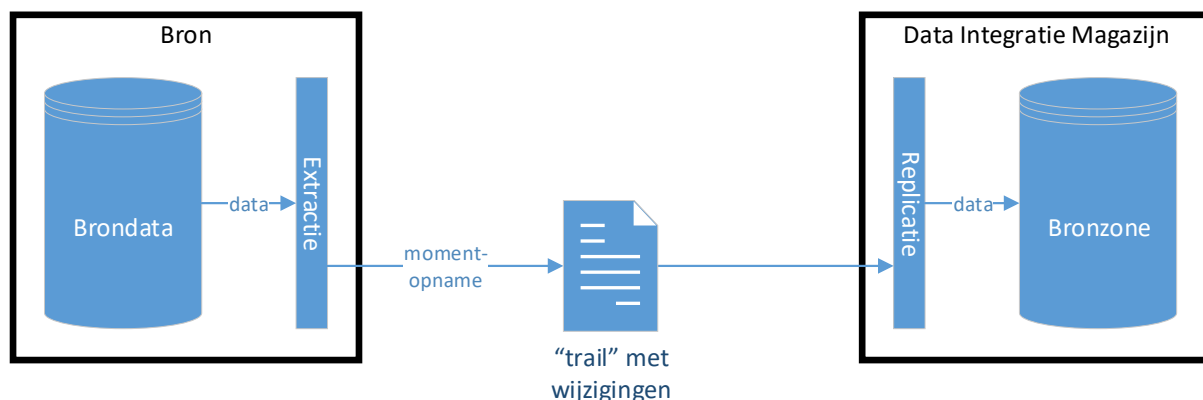
Ook bij export/import is de bron functioneel verantwoordelijk voor de correcte werking van de levering na wijzigingen in het bronsysteem. De technische verantwoordelijkheid ligt echter grotendeels bij het DIM.

Export/Import op basis van een bron-kopie is mogelijk zolang er voldaan wordt aan de in paragraaf 7.2.4 genoemde eisen voor bron-kopie database.

N.B. Ook bij blijvend gebruik van export/import-technologie blijft dekking van de functionele en documentaire eisen (geen afnemer-specifieke filters, pakbon, gedefinieerde data, etc.) vereist.

Daarnaast geldt, net als bij bestanden, dat de geleverde export bij de bron tenminste beschikbaar blijven totdat het DIM de ontvangst ervan heeft bevestigd.²⁸

7.2.6 Levering via CDC



Bij levering via een CDC worden wijzigingen in een brontabel "gerepliceerd" in een DIM-tabel. Dit gebeurt door de wijzigingen in de brondata af te leiden uit het database-log, deze in een "trail" naar het DIM te sturen, en de wijzigingen in het DIM vervolgens weer "na te spelen".

Het aanmaken, versturen en verwerken van een "trail" kan zowel per wijziging als in bulk.

De structuur van de gegevens in een "trail" is hard gekoppeld aan de structuur van die gegevens in de database van de bron. Dit resulteert in een "tight coupling" tussen bron en DIM, en daarmee in beperkte wijzigingsflexibiliteit. Daarnaast is CDC niet bruikbaar als de gegevens in het DIM worden opgeslagen in de (niet-relatieve) datalake-zone.

Ook bij CDC is de bron functioneel verantwoordelijk voor de correcte werking van de levering na wijzigingen in het bronsysteem. De technische verantwoordelijkheid ligt echter vooral bij het DIM.

Levering via CDC biedt in het algemeen te weinig voordelen om inzet ervan te rechtvaardigen, en voor event-gedreven leveringen verdient levering via berichten, zoals beschreven in de diverse doelarchitecturen op het gebied van integratie, de voorkeur.

N.B. Soms wordt CDC ingezet als "tussenstap" bij het aanmaken van berichten. Deze inzet valt binnen de grenzen (en verantwoordelijkheid) van het bronsysteem, en een dergelijk gebruik van CDC is, vanuit DIM-perspectief, dus "gewoon" een levering via berichten.

7.3 Terugkoppeling vanuit het DIM

Het DIM zal tijdens het verwerken van dataleveringen (naar behoefte) terugkoppeling geven naar de bron met betrekking tot:

- Het niet op tijd ontvangen van een levering

²⁸ N.B. Deze verantwoordelijkheid kan ook bij SIP-FT gelegd worden.

- Het ontvangen van een levering
- Problemen met een levering
- Succesvol verwerken van een levering
- Verwerking van brondata naar informatiegebieden
- Verwijdering van gegevens op basis van DLM

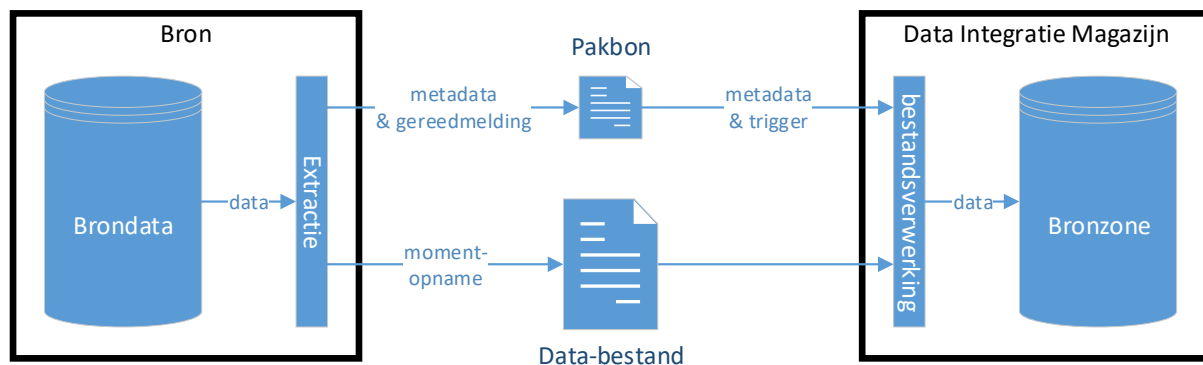
Deze paragraaf zal in de volgende versie van de interface standaarden verder worden uitgewerkt.

8 STANDAARDEN VOOR BERICHT- GEBASEERDE INTERFACES

*De standaarden voor levering van gegevens via berichten zijn grotendeels generiek (UWV-breed).
Dit hoofdstuk zal worden toegevoegd zodra deze UWV-brede standaarden zijn vastgesteld.*

9 STANDAARDEN VOOR BESTANDSLEVERINGE N

9.1 Basis-opzet



Bij levering via bestanden dienen gegevens en metadata gescheiden te zijn; gegevens worden geleverd in één of meer data-bestanden, en de metadata in een "pakbon".

9.2 Data-bestand

9.2.1 Technisch formaat

Er is geen vaste standaard voor het technisch formaat van het gegevensbestand; elk formaat waarvoor de ETL-tooling een "van de plank" inlees-optie biedt is acceptabel. Het voorkeursformaat voor gegevensbestanden is csv.

9.2.2 Eén recordtype per bestand

Elk gegevensbestand bevat slechts één type gegevensregels. Kop- en voet-regels zijn optioneel (aangezien de informatie erin ook is opgenomen in de gegevens inwinnings afspraken (GIA)).

9.2.3 Grote objecten

Niet elk gegevenselement kan eenvoudig in een record-veld worden opgenomen. Denk, bijvoorbeeld, aan tekstvelden met "speciale karakters", figuren, en aangehangen bestanden.

Voor dergelijke gegevens zal per interface een "beste benadering" worden bepaald.

Opties zijn:

- Element negeren
- Element converteren naar "lange string"
- Element in aparte bestanden (één per voorkomen) versturen, en in het data-bestand verwijzingen naar deze bestanden opnemen

9.2.4 Overige punten:

De naamgeving van objecten en attributen zijn conform de Oracle standaarden. Dus in hoofdletters, geenspaties of andere bijzondere tekens en beginnen met een letter.

Een levering kan meerdere data bestanden bevatten met precies dezelfde regelopmaak (RLO). Dit bijvoorbeeld als elk kantoor een individuele database heeft. Het onderscheid wordt op records niveau dan gemaakt doormiddel van een kantoorcode.

9.3 Pakbon

Bij bestandsleveringen dient de pakbon in json-formaat te worden opgeleverd. Pakbonnamen mogen **geen** hoofdletters bevatten, omdat we op Linux werken waar bestandsnamen hoofdlettergevoelig zijn. De pakbon bevat de volgende elementen:

Element	Toelichting
interface_naam	De in de GIA vastgelegde unieke code voor de interface. Interfacenaam in kleine letters, en deze mag geen underscore (" _ ") en/of spatie bevatten.
bron	De bron (of bronmodule, of bron-instantie) van de levering, zoals gedefinieerd in de GIA. De bronnaam in kleine letter, en deze mag geen underscore (" _ ") en/of spatie bevatten.
versienr_applicatie	Versienummer van de applicatie, en van de extractielogica, zoals gebruikt bij beheer van de bronapplicatie.
versienr_extractie_logica	
volgnr_levering	Het volgnummer van de levering. Wordt één hoger bij elke volgende levering. Bij gecorrigeerde herleveringen is het volgnummer gelijk aan dat van de levering waarvoor de herlevering plaats vindt, en bevat element "herlevering" de waarde "J".
herlevering	"J" als het hier een gecorrigeerde herlevering betreft, "N" als dat niet zo is.
run_id_extractie	Unieke identificatie van deze bestandslevering. Kan bijvoorbeeld het technische proces-id zijn waarmee de extractie is uitgevoerd.
extractiemoment	Het moment waarop de extractie is gestart. Formaat: jjjj-mm-dd uu:mm:ss met de 24-uurs notatie
periode_start	<i>(optioneel, alleen verplicht voor incrementele, dakpan en stapelbare leveringen)</i> Begindatum en -tijd en einddatum en -tijd van de periode waarvoor wijzigingen geleverd worden ²⁹ Formaat: jjjj-mm-dd uu:mm:ss met de 24-uurs notatie
periode_einde	
peildatum	<i>(optioneel alleen verplicht voor volledige leveringen)</i> De datum en tijd tot waar de geleverde data is bijgewerkt. Formaat: jjjj-mm-dd uu:mm:ss met de 24-uurs notatie
parameters	<i>Optioneel</i> , lijst-indicator, met daarbinnen voor elke bij de bronextractie gebruikte parameter een identifier en een waarde.
parameternaam	Naam van de parameter, uniek binnen de pakbon
parameterwaarde	Waarde van de parameter
bestanden	Lijst-indicator, met daarbinnen voor elk geleverd data-bestand ³⁰ de volgende elementen:

²⁹ Einddatums zijn exclusief; een maand loopt dus van de eerste van de maand tot aan de eerste van de volgende maand.

³⁰ Het pakbon bestand komt dus niet in deze lijst voor.

Element		Toelichting
	bestandsnaam	Naam van het geleverde bestand. De bestandsnaam in kleine letters.
	records	Lijst-indicator, met daarbinnen voor elk object binnen het geleverd data-bestand de volgende elementen:
	object	Het gegevensobject, in hoofdletters, dat in het bestand wordt geleverd. ³¹ Deze dient gelijk te zijn aan de technische entiteitsnaam in de RLO.
	object_ind	Optioneel, de waarde in de kolom die de verschillende record types identificeert. <i>Dit element weglaten tenzij er na overleg meerdere record types in een data-bestanden worden aangeleverd!</i>
	attributenlijst	Opsomming van alle geleverde attributen voor het object, in hoofdletters, gescheiden door een staand streepje (" "). De attribuutnamen dienen gelijk te zijn aan de technische attribuutnamen in de RLO. Bij afwijkingen wordt een levering afgekeurd en niet verwerkt.
	aantal records	Het aantal voor dit object geleverde gegevensrecords (dus exclusief eventuele header- en footer-records).
	sha256_controle_getal	Optioneel, een checksum om te controleren of de gegevens in een databestand in de oorspronkelijke staat, zoals door de bron klaargezet, ontvangen zijn. Het DIM genereert op basis van het databestand ook een sha256 checksum, en controleert of deze overeenkomt met de sha256_controle_getal in de pakbon. Hiermee wordt de integriteit van het databestand gewaarborgd.

Aandachtspunten voor de pakbon:

- Alle waarden in de pakbon worden, ongeacht het datatype in de bron, als tekst behandeld, en staan dus tussen dubbele quotes.
- JSON elementen mogen geen spaties in de benaming hebben (ETL-tool randvoorwaarde).
- Optionele elementen niet opnemen in de pakbon als die voor een levering niet van toepassing zijn (ETL-tool randvoorwaarde).
- JSON detail informatie kan gevonden worden op <https://www.json.org/json-en.html>
- De standaard voor JSON encoding is UTF-8

9.3.1 Afwijkend pakbon-formaat

Voor bronnen die alleen met grote inspanning aan de pakbon-eisen kunnen voldoen (m.n. COBOL-bronnen) kan, in overleg met de Datafabriek en na analyse van de totale kosten van de aansluiting (bron én DIM), van deze eis worden afgeweken, en gekozen worden voor maatwerk-aansluiting op het DIM.

9.3.2 Voorbeeld pakbon-formaat

Onderstaand is een fictief voorbeeld van de pakbon in het geval er na overleg (met Datafabriek) een uitzondering is dat er meerdere objecten in een data-bestand worden geleverd.

```
{
  "interface_naam": "persoon-mnd.test",
  "bron": "persoons.master-test",
  "versienr_applicatie": "<vn_r_app>",
```

³¹ Afhankelijk van de technische inrichting van bron-applicatie en levering kan dit refereren naar een entiteit, een tabel, of een record-type.

```

"versienr_extractie_logica": "<vnr_app_log>",
"volgnr_levering": "121",
"herlevering": "N",
"run_id_extractie": "8791",
"extractiemoment": "2019-09-01 03:35:02",
"periode_start": "2019-08-01 10:20:30",
"periode_einde": "2019-09-01 11:22:33",
"peildatum": "",
"parameters": [
  {
    "par_naam": "maand", "par_waarde": "201908"},
  {
    "par_naam": "omgeving", "par_waarde": "acceptatie"}
],
"bestanden": [
  {
    "bestandsnaam": "PERSOON-MND_PERSOON_i_20190801_20190101_20190901.csv",
    "records": [
      {
        "object": "PERSOON",
        "object_ind": "B2",
        "attributenlijst":
"PERSOON_ID|ACHTERNAAM|VOORNAAM|GEBORTE DATUM"},
      {
        "object": "FIETS",
        "object_ind": "fiets",
        "attributenlijst": "FIETS_ID|FIETSNAAM|FIETSVORM|FIETSDATUM"}
    ],
    "aantal_records": "319547"},
  {
    "bestandsnaam": "persoon-mnd_04_beroep_i_20190801_20190101_20190901.csv",
    "records": [
      {
        "object": "BEROEP",
        "object_ind": "",
        "attributenlijst": "PERSOON_ID|PRIMAIRE_BEROEP|SECUNDAIRE_BEROEP"}
    ],
    "aantal_records": "2984"},
  {
    "bestandsnaam": "persoon-mnd_05_beroep_i_20190801_20190101_20190901.csv",
    "records": [
      {
        "object": "BEROEP",
        "object_ind": "",
        "attributenlijst": "PERSOON_ID|PRIMAIRE_BEROEP|SECUNDAIRE_BEROEP"}
    ],
    "aantal_records": "1129"}
]
}

```

9.4 Volgorde levering bestanden

Bij het doorsturen van de interface-bestanden dienen eerst de databestanden verstuurd te worden (via SIP-FT), en pas als de rest is verstuurd de pakbon. Dit maakt het mogelijk om ontvangst van de pakbon binnen het DIM als "trigger" te gebruiken voor de start van de verwerking van de gehele gegevenslevering.

9.5 Bestandsnamen

De naamgeving voor bestanden is afhankelijk van de leveringswijze. Bestandsnamen mogen **geen** hoofdletters bevatten, omdat we op Linux werken waar bestandsnamen hoofdlettergevoelig zijn.

Leveringswijze	Pakbon-bestand
Stapelbaar	<interface-naam>_s_<startdatum>_<einddatum>_<extractiedatum>.json
Incrementeel	<interface-naam>_i_<startdatum>_<einddatum>_<extractiedatum>.json
Volledig	<interface-naam>_v_<peildatum>_<extractiedatum>.json

Leveringswijze	Data-bestand
Stapelbaar	<interface-naam>_s_<startdatum>_<einddatum>_<extractiedatum>_<object>.<ext>
Incrementeel	<interface-naam>_i_<startdatum>_<einddatum>_<extractiedatum>_<object>.<ext>
Volledig	<interface-naam>_v_<peildatum>_<extractiedatum>_<object>.<ext>

waarbij:

<interface-naam>	de in de GIA vastgelegde unieke code is voor de interface (in kleine letters, en zonder underscore "_")
<startdatum>	de startdatum is van de periode waarvoor de interface wordt geleverd
<einddatum>	de einddatum is van de periode waarvoor de interface wordt geleverd ³²
<peildatum>	de datum is tot waar de geleverde data is bijgewerkt ³³
<extractiedatum>	de datum is waarop de extractie is uitgevoerd ³⁴
<object>	het gegevensobject dat in het bestand wordt geleverd. Afhankelijk van de technische inrichting van bron-applicatie en -levering kan dit refereren naar een entiteit, een tabel, of een record-type
<ext>	de extensie van het bestand (bijvoorbeeld csv of dat), zoals vastgelegd in de GIA. Indien een bestand opgedeeld moet worden in delen dan is het eerste bestand de gewone extensie en daarna 001, 002 etc.

(alle datums zijn in JJJJMMDD formaat)

Een bestandsnaam moet toepasbaar zijn op Unix platform (gebruik kleine letters, underscores, streepjes, puntjes; gebruik geen tekens als "/", ">", "<", "|", ":", "&", "%", "\$" etc.)

Bestandsnamen in de pakbon van data bestanden moeten exact overeenkomen met de feitelijke naam.

Indien er in de bron voor een object geen gewijzigde gegevens zijn dit als een leeg bestand aanleveren. Hierdoor is het duidelijk dat er geen gewijzigde gegevens zijn en dat het niet zo is dat er iets fout is met de levering.

Er zijn bronnen welke tabellen zullen leveren met verschillende leveringswijzen. De levering zal dan incrementeel of stapelbaar volgen met betrekking tot bestandsnaamgevingsconventies. Indien een levering een combinatie van volledig, stapelbaar en incrementeel bevat, of een combinatie van volledig en incrementeel, of incrementeel en stapelbaar, dan zal de indicator "i" van incrementeel in de naamgeving opgenomen worden. Indien een levering een combinatie van volledig en

³² Einddatums zijn exclusief; een maand loopt dus van de eerste van de maand tot aan de eerste van de volgende maand.

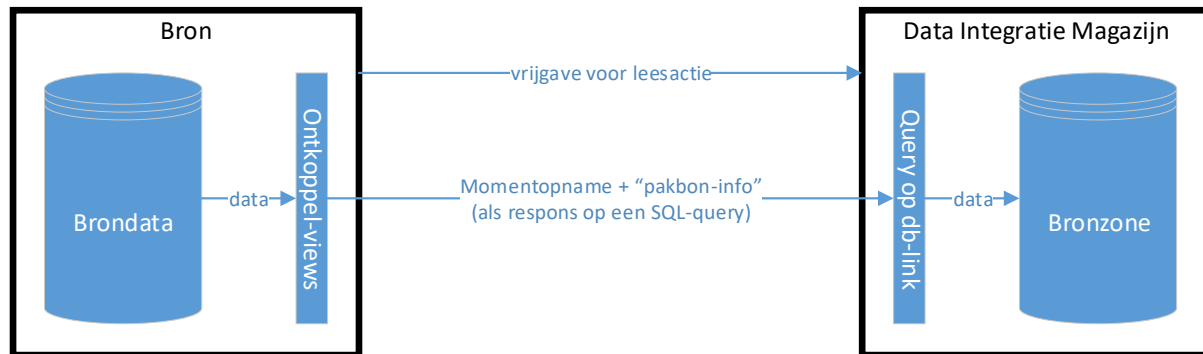
³³ Ook de peildatum is exclusief; een extractie t/m maand ultimo krijgt dus als peildatum de eerste van de volgende maand.

³⁴ Het gaat hier om de kalenderdatum; een extractie om 02:00 's nachts heeft dus **niet** de voorgaande kantoordag als extractiedatum

stapelbaar bevat, dan zal de indicator "s" van stapelbaar in de naamgeving opgenomen worden. In het RLO zal duidelijk worden opgenomen welke tabellen incrementeel/stapelbaar of volledig geleverd worden.

10 STANDAARDEN VOOR DATABASE-LINKS

10.1 Basis-opzet



Bij levering via een database-link extraheert **niet de bron, maar het DIM** de bron op gezette tijden (een deel van) de brongegevens. Dit extraheren mag echter pas plaatsvinden als de bron heeft aangegeven dat zulks "veilig" kan.

Extractie gebeurt in de regel met een vaste frequentie (meestal dagelijks of maandelijks), en de respons op de extractie-query kan dan gezien worden als een momentopname van de gegevens in de bron.

Een aandachtspunt is de ondersteuning van herlevering; als, in de brondatabase, gegevens overschreven kunnen worden zónder dat daarvan historie wordt bijgehouden, dan is bij herlevering (eigenlijk: herbevraging) de respons op de query niet per definitie gelijk aan die van de oorspronkelijke query.

Bij gebruik van een database-link is, aan de bron-zijde, een view-laag met ontkoppel-views verplicht. Dit om te zorgen dat de structuur van de gegevens in een data-bestand niet hard gekoppeld is aan de structuur van die gegevens in de database van de bron, of aan de technische implementatie van die database. Dit resulteert in een "loose coupling" tussen bron en DIM, en daarmee in wijzigingsflexibiliteit.

10.2 Ontkoppel-views

Toegang, vanuit het DIM, naar de bron-database mag alleen via de ontkoppel-views lopen.

Toegang, vanuit het DIM tot andere objecten in de bron-database mag op geen enkele wijze zijn geautoriseerd. De view-laag dient drie doelen:

1. Creëren "gegevensvenster" op bron-database, zodat doelbinding voor de toegang eenvoudig kan worden bewaakt, en eventuele niet relevante data kan worden uitgefilterd.
2. Bieden van een (uitwijk) ontkoppelvlak, zodat, mocht dit noodzakelijk zijn, het DIM niet synchroon met de bron hoeft mee te wijzigen
3. (indien relevant) Opnemen van "virtuele kolommen"³⁵ in het gegevensvenster, zodat deze door het DIM gezien worden als integraal onderdeel van het datamodel.

De ontkoppel-views zijn uitdrukkelijk niet bedoeld voor complexe transformaties en afleidingen. Eenvoudige transformaties (bv. typecasting) zijn alleen toegestaan als deze niet kunnen resulteren in verlies van informatie.

³⁵ Zoals gedefinieerd op bron-database niveau, al dan niet door aanroep van stored procedures of vergelijkbaar

De definitie van een ontkoppel-view is daarmee typisch iets als:

```
SELECT
    <subset van kolommen>,
    <eventuele virtuele kolommen>
FROM
    <bron-object>
WHERE
    <eventuele bron-specifieke filters>;
```

Met:

<subset van kolommen>

Een expliciete lijst van kolommen (dus geen "SELECT *") die potentieel relevant zijn voor DIM-afnemers; de lijst mag dus niet afnemer-specifiek zijn.

<eventuele virtuele kolommen>

Een lijst met alle niet-triviale (en voor DIM-afnemers potentieel relevante) afleidingen waarvoor de bron verantwoordelijk is.

<bron-object>

De brontabel (of bronview, of vergelijkbaar) waar de ontkoppelview toegang toe biedt.

<eventuele bron-specifieke filters>

Eventuele filters om technische en/of niet relevante rijen uit het gegevensvenster te verwijderen.

N.B. Filters op "vluchtige" criteria (zoals het VIP-zijn of niet) zijn niet toegestaan.

10.3 Technische inrichting

De inrichting van de database-links moet "Secure By Design" zijn. Dit is van invloed op zowel de bron- als de DIM-zijde van de link.

10.3.1 Bron-zijde

Dedicated non-expiring OWNER account met sterk wachtwoord.

Dit account heeft toegesneden rechten:

- create session privilege
- create private synonym
- create view
- select rechten op de bron-tabellen

De private synonyms verwijzen naar de bron-tabellen.

10.3.2 DIM-zijde

Dedicated non-expiring OWNER account met sterk wachtwoord.

Account heeft toegesneden rechten:

- create session privilege
- create databaselink
- create private synonym
- select rechten op de ontkoppel-views

De private synonyms verwijzen via het SYNONYM attribuut DBLINK naar de ontkoppel-views op de remote database. Binnen dit account is het object database-link gedefinieerd.

De database-link is een PRIVATE link en koppelt aan het dedicated account op het bronsysteem.

Zie Oracle documentatie voor creatie database links:

https://docs.oracle.com/database/121/SQLRF/statements_5006.htm#SQLRF01205

Voor de verwerking van (distributed) LOB's zie:

<https://docs.oracle.com/en/database/oracle/oracle-database/12.2/adlob/distributed-LOBs.html#GUID-7E450E86-3E4E-4714-A164-FD36B93722F6>

10.4 Pakbon info

Het database-link harnas verzamelt zelf de benodigde informatie om zorg te dragen dat er een controle kan plaats vinden om te verifiëren dat de levering correct is verwerkt.

10.5 Beschikbaarheid en vrijgave voor leesactie

Een database-link levering zal altijd een in de GIA vastgelegde starttijd (en eindtijd/tijdsvenster) hebben om de verwerking van de levering te starten. Er zijn verschillende mogelijkheden met betrekking tot het initiëren van het data leveringsproces:

- Trigger-record in trigger-tabel in de bron
- Trigger-file geleverd via SIP-FT
 - o Bestandsnaam van de trigger heeft het formaat <interface-naam>_<yyyymmdd>.trg
- Vaste starttijd

Indien er een trigger afgesproken is en deze niet binnen het afgesproken tijdsvenster gezet wordt krijgt DIM beheer een melding vanuit het ETL proces dat de levering niet heeft plaats gevonden en zal deze doorzetten naar de bron.

10.6 Ondersteuning herlevering

De mogelijkheden van een herlevering zijn afhankelijk van de mogelijkheden van de verschillende gegevens objecten in de bron. De opties zijn:

- Een volledige verwerking starten (indien er geen administratieve tijdlijn is)
- Een incrementele verwerking (indien er een administratieve tijdlijn is) waarbij het correcte peilmoment meegegeven wordt zodat er geen of minimaal (in geval er records overschreven worden) gegevens verlies op treedt.

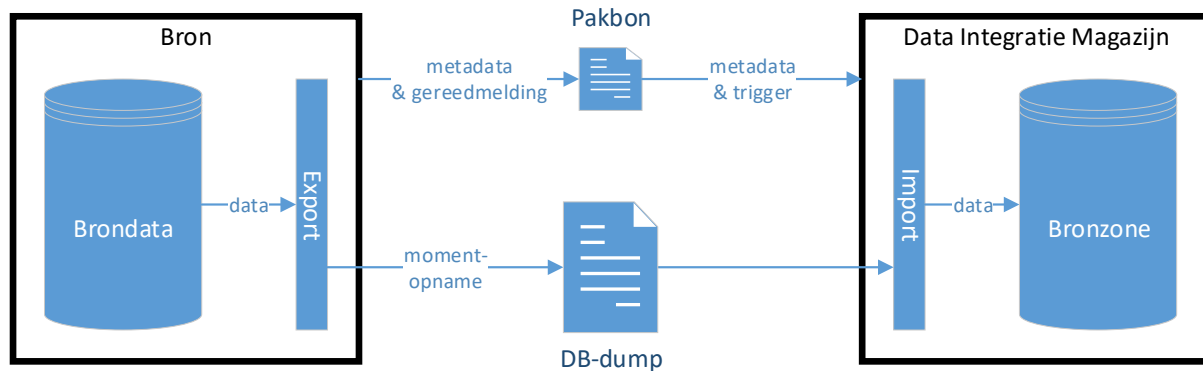
In beide gevallen is het mogelijk dat door het latere verwerkings tijdstip wijzigingen in de bron zijn gemist omdat deze zijn overschreven.

10.7 Inzet bron-kopie

Als een bronsysteem zowel een "primaire" brondatabase als een voor publicatie/interfacing bedoelde kopie-database bevat, is het de verantwoordelijkheid van het bronsysteem om zorg te dragen dat er in de synchronisatie tussen de originele en de kopie database geen filtering, geen aggregatie etc wordt toegepast welke tot gegevens verlies leidt. Of in andere woorden het is de verantwoordelijkheid van de bron om in hun synchronisatie proces te voldoen aan de in dit document genoemde interface standaarden.

11 STANDAARDEN VOOR EXPORT/IMPORT

11.1 Basis-opzet



Bij levering via export/import exporteert de bron op gezette tijden (een deel van) de bron-database naar een database-dump. Deze dump wordt vervolgens naar het DIM verstuurd (of door het DIM opgehaald) en daar geïmporteerd. De door de exportprogramma's (EXPDP) aangemaakte par- en log-bestanden vervullen hierbij de rol van pakbon.

De structuur van de gegevens in een dump is hard gekoppeld aan de structuur van die gegevens in de database van de bron. Dit resulteert in een "tight coupling" tussen bron en DIM, en daarmee in beperkte wijzigingsflexibiliteit. Daarnaast dienen bron- en doel-RDBMS gelijk te zijn, en moet in sommige gevallen de versie van de doel-database (die van het DIM) bekend zijn bij de bron. Tenslotte is export/import niet bruikbaar als de gegevens in het DIM worden opgeslagen in de (niet-relatieve) datalake-zone.

Ook bij export/import is de bron functioneel verantwoordelijk voor de correcte werking van de levering na wijzigingen in het bronsysteem. De technische verantwoordelijkheid ligt echter grotendeels bij het DIM.

Levering via export/import verdient daarmee nooit de voorkeur, en vereist expliciete acceptatie (door de Datafabriek, namens haar afnemers) van de tekortkomingen ervan. In de context van het project DataFabriek zullen op export/import gebaseerde leveringen alleen acceptabel zijn waar deze nu al, en met acceptabele betrouwbaarheid, worden ingezet richting de "legacy" datawarehouses.

N.B. Ook bij blijvend gebruik van export/import-technologie blijft dekking van de functionele en documentaire eisen (geen afnemer-specifieke filters, pakbon, gedefinieerde data, etc.) vereist. Daarnaast geldt, net als bij bestanden, dat de geleverde export bij de bron tenminste beschikbaar blijven totdat het DIM de ontvangst ervan heeft bevestigd.

11.2 Pakbon info

Bij export/import fungeert de combinatie van par- en log-file als pakbon. Beiden dienen dus met de data-bestanden te worden meegeleverd. De namen van de par-en log-file mogen **geen** hoofdletters bevatten, omdat we op Linux werken waar bestandsnamen hoofdlettergevoelig zijn.

11.2.1 Technische inrichting

Alle gebruikte parameters dienen te zijn gedefinieerd in een parameter file; PARFILE is de enige parameter die in de command line mag staan. Een deel van de parameters is optioneel (O) en een deel is verplicht (V). In de onderstaande tabel is dit aangegeven:

Parameter	Levering				Beschrijving
	Stapelbaar	Incrementeel	Dakpa	Volledig	
DIRECTORY	V	V	V	V	De locatie waar de .dmp en de .log bestanden worden geplaatst. De Datafabriek stelt geen eisen aan de gebruikte directory. De parameter dient wel in de parameter file te staan (en niet in de command line).
LOGFILE	V	V	V	V	De naam van het logbestand. Voor naamgevingsspecificaties zie de bijbehorende sub paragraaf.
DUMPFIL	V	V	V	V	De naam van het .dmp bestand. Voor naamgevingsspecificaties zie de bijbehorende sub paragraaf.
FILESIZE	O	O	O	O	Indien er een maximum filesize voor de dumpbestanden is kan deze parameter gebruikt worden om deze in te stellen en meerdere dumpfiles te maken indien de filesize overschreden wordt. In het parameter bestand wordt in de waarde van de DUMPFIL parameter een substitutie variabele _%U opgenomen. De _%U is in dit geval verplicht.
VERSION	V	V	V	V	De parameter VERSION bevat de laagste van de twee Oracle versies waartussen de data levering plaats vindt. ³⁶
METRICS	V	V	V	V	Het aantal objecten en de tijdsduur van de export worden in het .log bestand vastgelegd. Gebruik: METRICS=YES
LOGTIME	V	V	V	V	Tijdstempels op status als log file berichten. Gebruik: LOGTIME=ALL
COMPRESSION	V	V	V	V	COMPRESSION=ALL als de bron dit ondersteund, COMPRESSION=METADATA_ONLY als dat niet zo is
CONTENT	V	V	V	V	Altijd zowel de metadata als de data. Gebruik: CONTENT=ALL
PARALLEL	V	V	V	V	Het aantal parallelle processen dat gebruikt wordt. Default waarde is 1. Geldige waardes 1 tot en met 99. Elk parallel proces maakt een eigen .dmp bestand. In het parameter bestand wordt in de waarde van de DUMPFIL parameter een substitutie variabele _%U opgenomen. (_%U is in de bestandsnamen niet nodig als PARALLEL=1. Let op wel verplicht als de FILESIZE parameter gebruikt wordt!) Dit zorgt ervoor dat de .dmp bestanden een unieke naam krijgen.

³⁶ Het DIM heeft momenteel Oracle versie 12.2.0.1

Parameter	Levering				Beschrijving
	Stapelbaar	Incrementeel	Dakpa	Volledig	
EXCLUDE	V	V	V	V	Grants, triggers en indices dienen niet in het .dmp bestand te zijn opgenomen: EXCLUDE=GRANT,TRIGGER,INDEX
TABLES	V	V	V	V	De lijst met tabellen welke in het .dmp bestand geëxporteerd zijn. Dit is de enige toegestane vorm van scoping op tabellen. Dat betekent geen scoping op het niveau van tablespaces, schemas e.d. Tabellennamen dienen volledig te worden uitgeschreven zonder wildcards. TABLES=('<schema naam>.<tabel 1>', '<schema naam>.<tabel 2>')
QUERY	V	V	V	O	Voor elke tabel kan een eigen query opgegeven worden. Indien filters noodzakelijk zijn dienen in overleg met het DataFabriek project gebruikt te worden.
FLASHBACK_TIME	V	V	V	V	Met behulp van de parameter FLASHBACK_TIME wordt het SCN geselecteerd die het beste overeenkomt met de gespecificeerde tijd. ³⁷ De Data Pump export wordt uitgevoerd met data die consistent is tot deze SCN. Dit zorgt ervoor dat de geëxporteerde tabellen onderling consistent blijven, zelfs als ze tijdens de export wijzigen. N.B. Het wordt nog steeds aangeraden de Data Pump export in een tijdsvenster uit te voeren waarin weinig tot geen wijzigingen plaats vinden. De waarde van FLASHBACK_TIME staat meestal op SYSTIMESTAMP, oftewel het begin van het extractieproces. Zie paragraaf 11.5 (Ondersteuning herlevering) voor meer detail.

Om de parameter file in het DIM automatisch te kunnen verwerken gelden de volgende aanvullende eisen:

- Géén spatie tussen parameternamen en =-teken
- Verplichte parameters in de volgorde van bovenstaande beschrijving. Overige parameters mogen ervoor, er na, of ertussen.
- Lijsten zijn komma gescheiden met waardes direct achter de komma of een einde regel na de komma en de waarde op de volgende regel

Meer informatie over Data Pump kan gevonden worden in:

³⁷ Het System Change Number (SCN) is een interne teller van het Oracle RDBMS – elke keer dat er een commit gedaan wordt wordt de teller opgehoogd. Het SCN markeert daarmee een punt in de tijd waarop de database consistent is.

Meer informatie over SCN op http://docs.oracle.com/cd/B28359_01/server.111/b28286/functions149.htm

<https://docs.oracle.com/en/database/oracle/oracle-database/12.2/sutil/oracle-data-pump-export-utility.html>

11.3 Volgorde levering bestanden

Bij het doorsturen van de interface-bestanden dienen eerst de databestanden (*.dmp) verstuurd te worden (via SFTP), en pas als de rest is verstuurd de logfile en, als laatste, de par-file. Dit maakt het mogelijk om ontvangst van de par-file binnen het DIM als "trigger" te gebruiken voor de start van de verwerking van de gehele gegevenslevering.

11.4 Bestandsnamen

De naamgeving voor bestanden is afhankelijk van de leveringswijze. Bestandsnamen mogen **geen** hoofdletters bevatten, omdat we op Linux werken waar bestandsnamen hoofdlettergevoelig zijn.

Leveringswijze	Data-bestand
Stapelbaar	<interface-naam>_s_<startdatum>_<einddatum>_<extractiedatum>_%U.dmp
Incrementeel	<interface-naam>_i_<startdatum>_<einddatum>_<extractiedatum>_%U.dmp
Dakpan	<interface-naam>_p_<startdatum>_<einddatum>_<extractiedatum>_%U.dmp
Volledig	<interface-naam>_v_<peildatum>_<extractiedatum>_%U.dmp

Leveringswijze	Log-bestand
Stapelbaar	<interface-naam>_s_<startdatum>_<einddatum>_<extractiedatum>.log
Incrementeel	<interface-naam>_i_<startdatum>_<einddatum>_<extractiedatum>.log
Dakpan	<interface-naam>_p_<startdatum>_<einddatum>_<extractiedatum>.log
Volledig	<interface-naam>_v_<peildatum>_<extractiedatum>.log

Leveringswijze	Par-bestand
Stapelbaar	<interface-naam>_s_<startdatum>_<einddatum>_<extractiedatum>.par
Incrementeel	<interface-naam>_i_<startdatum>_<einddatum>_<extractiedatum>.par
Dakpan	<interface-naam>_p_<startdatum>_<einddatum>_<extractiedatum>.par
Volledig	<interface-naam>_v_<peildatum>_<extractiedatum>.par

waarbij:

- <interface-naam> de in de GIA vastgelegde unieke code is voor de interface
 - <startdatum> de startdatum is van de periode waarvoor de interface wordt geleverd
 - <einddatum> de einddatum is van de periode waarvoor de interface wordt geleverd³⁸
 - <peildatum> de datum is tot waar de geleverde data is bijgewerkt³⁹
 - <extractiedatum> de datum is waarop de extractie is uitgevoerd⁴⁰
 - _%U het volgnummer van het bestand bij parallelle extractie (PARALLEL parameter > 1).
- Als PARALLEL = 1, dan is dit deel van de bestandsnaam niet relevant.

(alle datums zijn in JJJJMMDD formaat)

³⁸ Einddatums zijn exclusief; een maand loopt dus van de eerste van de maand tot aan de eerste van de volgende maand.

³⁹ Ook de peildatum is exclusief; een extractie t/m maand ultimo krijgt dus als peildatum de eerste van de volgende maand.

⁴⁰ Het gaat hier om de kalenderdatum; een extractie om 02:00 's nachts heeft dus **niet** de voorgaande kantoor dag als extractiedatum

11.5 Ondersteuning herlevering

Voor bronsystemen die **geen** administratieve tijdslijn bijhouden, is het uitgangspunt om in het DIM een administratieve tijdslijn op te bouwen met de SYSTIMESTAMP van de bron, omdat dat het meest exacte moment van het extractieproces is en we hiermee een consistente set van tabellen in de exports krijgen. Deze kan opgehaald worden uit de waarde van het extractiemoment in de log file, welke beschouwd kan worden als het peilmoment. Dat levert in het geval van een herlevering bij systemen zonder administratieve tijdslijn, wel een minimale afwijking (hooguit een aantal seconden) op in de administratieve tijdslijn, maar deze afwijking is bekend, en daarmee goed te documenteren.

De waarde van FLASHBACK_TIME staat meestal ook op SYSTIMESTAMP. In theorie is het mogelijk om bij gecorrigeerde herleveringen de FLASHBACK_TIME op die van de originele levering te zetten (en niet op SYSTIMESTAMP). Aandachtspunt is dan dat de SCN slechts een beperkte bewaartermijn heeft, en dat het werkelijk peilmoment dus later kan liggen dan de opgegeven FLASHBACK_TIME. Vandaar dat ook voor herleveringen de FLASHBACK_TIME op SYSTIMESTAMP moet worden gezet.

11.6 Inzet bron-kopie

Er zijn bron systemen welke om de operationele database te ontzien een kopie database hebben ten behoeve van backup en queries. Het is de verantwoordelijkheid van de bron om zorg te dragen dat er in de synchronisatie tussen de originele en de kopie database geen filtering, geen aggregatie etc wordt toegepast welke tot gegevens verlies leidt. Of in andere woorden het is de verantwoordelijkheid van de bron om in hun synchronisatie proces te voldoen aan de in dit document genoemde interface standaarden.

12 STANDAARDEN VOOR CDC

Directe (voor het DIM zichtbare) inzet van Change Data Capture (CDC) is geen gewenste leveringstechniek. Zie paragraaf 7.2.6 (Levering via CDC).

Als CDC, binnen een bronsysteem, wordt ingezet als "tussenstap" bij het aanmaken van berichten, dan valt deze inzet binnen de grenzen (en verantwoordelijkheid) van dat bronsysteem, en een dergelijk gebruik van CDC is, vanuit DIM-perspectief, dus "gewoon" een levering via berichten.

13 STANDAARDEN M.B.T. BEHEERSPROCESSEN

13.1 Bewaking keten-integriteit

In het algemeen kan voor alle leveringswijzen de bewaking van keten-integratie en reconciliatie door middel van een periodieke volledige levering uitgevoerd worden.

13.1.1 Bewaking keten-integriteit bij event-gedreven levering

Zoals in paragraaf 5.3.1 beschreven is een nadeel van event-gedreven levering dat een gemiste gebeurtenis (aan de bronzijde of in de DIM-verwerking) een blijvende discrepantie tussen bron en DIM oplevert.

Om deze kans zo klein mogelijk te maken beschikken bronnen die “publish & subscribe” ondersteunen daarom over een robuuste tracking-mechanisme van wijzigingen. Daarnaast wordt het transport van de gegevensgebeurtenissen van bron naar afnemers (de “geabonneerden”) ondersteund door functionaliteit t.b.v. “gegarandeerde levering”⁴¹ zoals verzend- en ontvangstqueues.⁴²

Typisch zal bij event-gedreven levering geen reconciliatie tussen bron en DIM vereist zijn; het “publish & subscribe” mechanisme is immers standaard (lees: niet alleen voor de levering aan het DIM) ingericht op betrouwbare tracking en levering.

Mocht de betrouwbaarheid van de levering toch niet hoog genoeg zijn om te voldoen aan de, door de afnemers van het DIM gestelde, eisen t.a.v. gegevensintegriteit, dan verdient inzet van een andere leveringswijze de voorkeur boven periodieke reconciliatie van het DIM met een volledige afslag van de bron.

13.1.2 Bewaking keten-integriteit bij stapelbare levering

Zoals in paragraaf 5.3.2 beschreven is een nadeel van stapelbare levering dat onverwachte wijzigingen in de bron niet aan het DIM worden doorgegeven.

Het gaat dan om:

- a) Invoer in het bronsysteem buiten de tracking/logging om. Het gaat daarbij i.h.a. om reparatie-acties door IT.
- b) Verwijdering anders dan vanwege het verlopen van de “houdbaarheidsdatum”, bijvoorbeeld omdat een persoon zich heeft beroepen op het “recht op vergetelheid”.

Ad a)

Dit risico word geminimaliseerd door hiervoor, in de GIA, een uitwijkscenario vast te leggen. Daarin gedekte aspecten zouden, tenminste, moeten zijn:

- Ook bij reparatie-invoer wordt automatisch de systeemdatum als begindatum/tijd aan het record wordt toegevoegd. Dit wordt na reparatie gecontroleerd.
- Reparatie-wijzigingen worden, waar mogelijk doorgevoerd via tegenboekingen (of vergelijkbaar), en gedragen zich daarmee weer als invoer. Is dat niet mogelijk, dan dienen

⁴¹ Engels: “guaranteed delivery”

⁴² End-to-end “handshakes” zijn meestal niet mogelijk, omdat dit van een “publish & subscribe” model de facto weer een point-to-point oplossing maakt

de betreffende wijzigingen via een eenmalige levering ter beschikking gesteld te worden aan het DIM, zodat ze daar, indien nodig en ook eenmalig, kunnen worden verwerkt.

Ad b)

Deze verwijderingen kunnen het eenvoudigst in het DIM verwerkt worden door periodiek een volledige levering aan het DIM te doen, en deze in het DIM, uitsluitend te gebruiken om verwijderingen uit het bronsysteem in het DIM "na te spelen.

Dit proces is ook bruikbaar als de bron geen "houdbaarheidsdatum" kan leveren.

13.1.3 Bewaking keten-integriteit bij incrementele levering

Zoals in paragraaf 5.3.3 beschreven is een nadeel van incrementele levering dat een gemiste gebeurtenis (aan de bronzijde of in de DIM-verwerking) een blijvende discrepantie tussen bron en DIM oplevert.

Meestal wordt dit veroorzaakt door wijzigingen in het bronsysteem buiten de tracking/logging om. Het gaat daarbij i.h.a. om reparatie-acties door IT.

Dit risico wordt geminimaliseerd door hiervoor, in de GIA, expliciet vast te leggen dat ook reparatie-acties binnen het tracking/logging proces moeten vallen.

Mocht de betrouwbaarheid van incrementele levering niet hoog genoeg zijn om te voldoen aan de, door de afnemers van het DIM gestelde, eisen t.a.v. gegevensintegriteit, dan verdient volledige levering de voorkeur. Is dit, vanwege gegevensomvang en/of bronsysteembeperkingen, niet mogelijk in de door de afnemers van het DIM gewenste frequentie, dan dient de via incrementele leveringen opgebouwde inhoud van het DIM periodiek gereconcilieerd te worden met een volledige afslag van de bron. Deze reconciliatie kan, eventueel, ook ad hoc ingezet worden als er, onverhoopt, toch IT-reparaties in de bron zijn uitgevoerd die niet correct in de incrementele levering zijn verwerkt.

13.2 Missende levering

Er is sprake van een missende levering wanneer een levering **wel** door de bron is verstuurd, maar **niet** door het DIM is ontvangen.⁴³ In dat geval dient de missende levering alsnog geleverd te worden.

Standaard is om te voorkomen dat deze situatie kan optreden door het implementeren van een vorm van "gegarandeerde levering":

- Bij levering via berichten gebeurt dit door inzet van queues, e.e.a. conform de UWV-standaarden op dit gebied; het DIM stelt hieraan geen specifieke eisen.
- Bij levering via bestanden (inclusief export/import) gebeurt dit door inzet van SIP-FT. De bron mag een naar de SIP-FT server verstuurd bestand pas uit de eigen omgeving verwijderen als ontvangst ervan door SIP-FT is gevalideerd. Deze validatie volgt de SIP-standaards op dit gebied; het DIM stelt hieraan geen specifieke eisen.
- Bij levering via DB-links kan dit probleem niet optreden.

⁴³ Het DIM draagt zelf zorg voor de archivering van reeds ontvangen leveringen t.b.v. eventuele herverwerking ervan bij interne foutsituaties

Als een bestandslevering niet via SIP-FT verloopt, dan mag een bron een naar DIM verstuurd bestand pas uit de eigen omgeving verwijderen als ontvangst ervan door DIM is bevestigd.

13.3 Herlevering

Deze paragraaf zal worden uitgebreid in een latere versie van dit document, als meer duidelijk is over de (on)mogelijkheden van de bronnen op het gebied van herlevering, en over de impact die dit heeft op de laadprogrammatuur voor de bronzone (het "harnas").

Een herlevering oftewel een gecorrigeerde levering is noodzakelijk als een eerdere levering:

- a) (formaat)technisch corrupt was
- b) filtertechnisch incorrect was
- c) functioneel incorrect was

In de pakbon van de herlevering zal het element Herlevering de waarde 'J' hebben.

Ad a)

De meest voorkomende vormen van technische corruptie zijn onvolledig gevulde bestanden en bestanden met een afwijkend formaat.

Technisch corrupte leveringen worden in het algemeen bij aanvang van de verwerking in het DIM als zodanig herkend, op basis van verschillen tussen interface-lading en pakbon-informatie en/of interface-lading en interface-specificatie.⁴⁴ De verwerking wordt dan gestopt, en eventuele deelverwerkingen teruggedraaid.

In de GIA worden specifieke afspraken gemaakt over de wijze waarop een corrupte levering aan de bron wordt gemeld en over de termijn waarbinnen alsnog een correcte levering plaats moet vinden.

Ad b)

De meest voorkomende vormen hiervan zijn het leveren van een verkeerd (bijvoorbeeld oud) bestand, of het leveren van een met de verkeerde filterparameters aangemaakt bestand.

Verkeerd gefilterde leveringen worden in het algemeen al voor aanvang van de verwerking in het DIM als zodanig herkend, op basis van verschillen tussen pakbon-informatie en verwachte pakbon-informatie. Verwerking wordt dan niet gestart.

In de GIA worden specifieke afspraken gemaakt over de wijze waarop een verkeerde levering aan de bron wordt gemeld en over de termijn waarbinnen alsnog een correcte levering plaats moet vinden.

Ad c)

Een "functioneel incorrecte" levering is een levering die niet voldoet aan de functionele specificaties. Dit zou bij goed geteste interface-functionaliteit alleen mogen voorkomen als er in de brondata zelf iets onverwachts is gebeurd. Meest voorkomende geval is een incrementele levering waarin niet alle gegevenswijzigingen in de bron zijn meegenomen, omdat die wijzigingen niet op correcte wijze, en in afwijking van afgesproken procedures, in de bron zijn ingevoerd.

In de GIA worden specifieke afspraken gemaakt over de wijze waarop een verkeerde levering aan

⁴⁴ Een bijzondere vorm van dat laatste treedt op als een interface-wijziging in de bron nog niet in het DIM is verwerkt.

de bron wordt gemeld en over de termijn waarbinnen alsnog een correcte levering plaats moet vinden. Deze kunnen echter minder strikt zijn dan bij de eerdere gevallen, omdat de reparatie-inspanning per geval kan verschillen.

14 STANDAARDEN VOOR TE LEVEREN GEGEVENSDEFINITIES

Onderstaande beschrijft de standaarden voor de (door de bron) met de interface-specificatie mee te leveren gegevensdefinities.

14.1 Definities

14.1.1 Gegevensdefinities eenduidig vastgelegd

Van elk gegevensobject (en elk veld daarbinnen) in een broninterface moet de gegevensdefinitie bekend zijn, en beschreven in eenduidige terminologie, bij voorkeur zoals vastgelegd in FUGEM en/of CGM.

N.B. Deze definitie moet conform het **werkelijk** gebruik van het object zijn, dus inclusief “workarounds” en ander creatief gebruik.

14.1.2 Formaat vastgelegd

Van elk gegevensobject (en elk veld daarbinnen) in een broninterface moet het gegevensformaat, zowel technisch (in de brondatabase) als functioneel (in het bronproces) gedefinieerd zijn.

Denk hierbij aan:

- Technische en functionele data typen
- Verplicht of niet (zowel technisch als functioneel, voor de functionele versie ook beschreven of het veld alleen in bepaalde gevallen verplicht is)
- Toegestane waarden of waardebereiken
- Tekstformaten en maximale lengtes
- Datumformaten
- Precisie
- Meeteenheid of muntsoort

14.1.3 Sleutels gedefinieerd

Van elk gegevensobject in een broninterface moet bekend zijn welk veld (of combinatie van velden) in dat object kan worden gebruikt als de (tijds-invariante) primaire sleutel van dat veld. Hieruit volgt dat sleutels niet hergebruikt mogen worden. Indien sleutels wel hergebruikt worden dient al in/bij het RLO uitgezocht te worden hoe hier in het DIM mee omgegaan moet worden. Dit kan bijvoorbeeld door het toevoegen van een additioneel veld aan de sleutel om het onderscheid te kunnen maken.

14.1.4 Historisch gedrag gedefinieerd

Van elk gegevensobject (en elk veld daarbinnen) in een broninterface moet het historisch gedrag bekend zijn.

Voorbeelden hiervan zijn: kan het object/veld in de bron na aanmaak nog wijzigen of niet (is het "overschrijfbaar"), is het object in de bron al geversioneerd⁴⁵, is het object of veld een momentopname van een afgeleid object of veld.

14.2 Gebruiksgrenzen

14.2.1 Vertrouwelijkheid gedefinieerd

Van elk gegevensobject (en elk veld daarbinnen) in een broninterface moet de vertrouwelijkheidsclassificatie bekend zijn.

N.B. Meestal is de vertrouwelijkheid voor alle voorkomens van het gegevenselement gelijk.

Kunnen ze per "data-record" verschillen (bijvoorbeeld bij VIP-info, eigen personeel of bijzondere gevalsbehandeling), dan dienen ze óf als onderdeel van dat data-record geleverd te worden, óf de binnen het DIM toe te passen afleiding moet zijn gedefinieerd in de interface-specificatie.

14.2.2 Maskerings-eisen gespecificeerd

Van elk veld binnen een gegevensobject in een broninterface moet bekend zijn of het, vanuit AVG-perspectief, als identificerend beschouwd moeten worden en, zo ja, welke maskering erop dient te worden toegepast.⁴⁶

14.2.3 Gebruiksbeperkingen gedefinieerd

Van elk gegevensobject (en elk veld daarbinnen) in een broninterface moet bekend zijn of er, om wat voor een reden dan ook, beperkingen zijn op de verwerking ervan (in de breedste zin des woords) t.b.v. rapportage, analyse en ketenintegratie.

⁴⁵ M.a.w. resulteert elke wijziging van een gegeven in een (opgeslagen, en zelf niet meer wijzigbare) nieuwe versie ervan.

⁴⁶ Het gaat hierbij niet om het te gebruiken maskerings-algoritme, maar om de "maskerings-groep" (bv. BSN's) waarbinnen het veld valt.

15 STANDAARDEN VOOR SPECIFICATIES

Dit hoofdstuk wordt vervolledigd in v2 van dit document. Het zal, waar mogelijk, zijn gebaseerd op bestaande standaarden m.b.t. functionele en technische ontwerpen (FO's en TO's).

15.1 Gegevenslijst ("RLO")

Het belangrijkste onderdeel van de specificaties is de definitie van de geleverde gegevens. Hiervoor is een Excel-template beschikbaar, met bijbehorende "invulhulp" (als Word-document).

In dat template is onder andere opgenomen:

- Welke gegevensobjecten (record-typen, tabellen) worden geleverd, met voor elk object:
 - wat een voorkomen van het object (een record, een rij) functioneel representeert
 - hoe het object zich, in de bron, historisch gedraagt (bv. overschrijfbaarheid)
 - welke attributen (velden, kolommen) voor het object worden geleverd
 - wat de unieke identificatie (primaire sleutel) van dat object is
 - of het object incrementeel, dakpan, stapelbaar of volledig geleverd wordt
- Voor elk attribuut:
 - wat het attribuut functioneel representeert
 - wat het technische formaat is
 - of het veld beschouwd moet worden als "identificerend" moet worden en, zo ja, in welke maskeringsgroep⁴⁷ het valt
 - wat de unieke identificatie (primaire sleutel) van dat object is
 - of het attribuut een administratieve tijdlijn veld is

15.2 FO en TO

Het **functionele** doel van een gegevenslevering aan het DIM is het leveren van gegevens in volledig detail. Meestal betreft het hier levering van de **technische** (fysiek opgeslagen) velden in de bron. Voor gegevensleveringen zullen FO en TO naar verwachting dus nauw verweven zijn.

Voor het **formaat** van FO's en TO's bestaan i.h.a. al standaarden bij de bron. Het DIM stelt hier geen specifieke eisen aan.

Voor wat betreft de **inhoud** zal de beheersorganisatie Datafabriek (en haar voorloper: het projectteam DataFabriek) conform het reguliere IV-voortbrengingsproces controleren dat beide documenten eenduidig en van voldoende detail zijn.

N.B. Uitdrukkelijk wordt ook het TO door de Datafabriek gevalideerd.

⁴⁷ Het opstellen van een eerste lijst van maskeringsgroepen (met hun bijbehorende maskeringseisen) is onderdeel van fase 3 van het project. Deze lijst zal in vervolgfases verder worden uitgebreid.

16 STANDAARDEN VOOR GIA'S

*Voor de GIA's gelden de standaarden/processen van Gegevens Diensten Uitvoeren.
Op basis van de inhoud van dit document zal gecontroleerd worden dat de opgestelde GIA's alle
voor het DIM benodigde afspraken bevat.*

BIJLAGE A: AFKORTINGEN

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

Afkorting	Verklaring, eventueel met toelichting
AB	Architectuur Board
AVG	Algemene Verordening Gegevensbescherming
BGB	Bijzondere gevalsbehandeling (bijvoorbeeld mensen in een blijf van mijn lijf huis) (classificatie met impact op vertrouwelijkheid)
CDC	Change Data Capture
CGM	Canoniek gegevensmodel
DA DWH	Design Authority afdeling DWH
DBA	Database administrator
DF	DataFabriek (het project) of Datafabriek (de organisatie)
DIA	Data Integratie en Analyse
DIM	Data Integratie Magazijn
DLM	Data Lifecycle Management
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.
EP	Eigen Personeel (classificatie met impact op vertrouwelijkheid)
EPK	Kinderen van Eigen Personeel (classificatie met impact op vertrouwelijkheid)
EPP	Partners van Eigen Personeel (classificatie met impact op vertrouwelijkheid)
ESB	Enterprise Service Bus
FO	Functioneel ontwerp
FUGEM	Functioneel gegevensmodel
GIA	Gegevens Inwinnings Afspraak
GLO	Gegevensleveringsovereenkomst (Oude term niet meer gebruiken)
RDBMS	Relationeel DataBase Management Systeem
SNO	Serviceniveauovereenkomst
TO	Technisch ontwerp
UPA	UWV Personen Administratie (master data personen)
VIP	Very Important Person; in de context van UWV een indicatie dat deze persoon zó bekend is dat zijn/haar gegevens met extra vertrouwelijkheid dienen te worden behandeld.

BIJLAGE B: BEKENDE INTERFACE UITZONDERINGEN

De in deze bijlage beschreven interface uitzonderingen zijn het resultaat van tekortkomingen in de bron welke momenteel nog niet kunnen worden opgelost door de bron.

16.1 Dakpan Levering

Een dakpan levering is een bijzondere vorm van een incrementele levering. Bij een dakpan levering levert de bron periodiek alle gewijzigde gegevens waarbij er een overlap met gegevens uit voorgaande leveringen is. Dit leid tot een beperkte mate van zelf reparerendheid met betrekking tot gemiste gebeurtenissen.

Een voorbeeld van een situatie waarin een dakpan levering gedaan wordt is als de bron bestaat uit een verzameling van identieke databases (op bijv. kantoor niveau) welke elk een levering doet naar een centraal verzamelpunt. Op gezette tijden wordt vanuit het verzamelpunt de data geleverd naar het DIM. Het kan voorkomen dat één van deze leveringen naar het centrale verzamelpunt te laat is en deze dus niet is meegenomen in de levering naar het DIM. Door een overlappende periode te hebben kan bij de volgende levering naar het DIM er automatisch een reparatie van de missende data uitgevoerd worden omdat deze nogmaals is geleverd.

Dakpan leveringen worden alleen in overleg geaccepteerd. De verantwoordelijkheid om binnen de periode van de dakpan levering alle wijzigingen te leveren ligt bij de bron.

Technisch wordt de levering op dezelfde wijze verwerkt als een incrementele levering. In het vervolg van dit document geldt voor dakpan leveringen hetzelfde als voor incrementele levering tenzij er expliciet iets over de dakpan levering verteld wordt.

De voor- en nadelen van een dakpan levering zijn:

	Dakpan
Historie afleidbaar uit levering?	Deels
Verwijderingen afleidbaar uit levering?	Nee
Leverfrequentie	> 1/dag
Efficiëntie delta-verwerking	Laag/Gemiddeld
Zelf reparerend?	Deels ⁴⁸
Eisen aan wijzigings-tracking	Hoog ⁴⁹

Het moment van laatste wijziging moet, binnen de bron, beschikbaar zijn om te kunnen bepalen in welke levering een gegeven dient te worden opgenomen.

De frequentie van levering kan hoog zijn; hoe hoger de frequentie, hoe minder data er in de levering zit, en dus hoe sneller/eenvoudiger deze (door te bron) is samen te stellen. Echter de reden voor dakpan levering is dat er bij de levering naar het centrale verzamelpunt

⁴⁸ Het zelfreparerend zijn geldt alleen voor de data die alsnog binnen de periode aangeleverd wordt. Het is dus beperkt en niet een reconciliatie zoals bij een volledige levering mogelijk is.

⁴⁹ Dakpan leveringen stellen dezelfde eisen aan de administratieve tijdlijn als een incrementele leveringen.

vertraging/verlies kan optreden. Een frequentie hoger dan de aanlevering naar het centrale verzamelpunt van de bron heeft weinig toegevoegde waarde.

Meerdere extracties per dag zijn dus mogelijk maar is alleen zinnig als dit aansluit op een hogere levering naar het centrale verzamelpunt van de bron.⁵⁰

De efficiëntie van de delta-verwerking is gemiddeld. Echter het is wel lager dan bij een pure incrementele levering: het DIM ontvangt gewijzigde gegevens plus van een kleine periode al verwerkte gegevens, maar moet daardoor nog wel bepalen wat er gewijzigd is en het type wijziging is.

Een risico bij incrementele levering is dat een gemiste gebeurtenis (aan de bronzijde of in de DIM-verwerking) een blijvende discrepantie tussen bron en DIM oplevert. Dakpan leveringen proberen deze gemiste gebeurtenissen nogmaals te leveren en het risico van de incrementele levering te verkleinen (mits het in het tijdframe van de periode van de dakpan levering valt).

Identiek als bij incrementele leveringen geldt voor dakpan leveringen dat om deze kans zo klein mogelijk te maken is een robuuste tracking (en logging) van wijzigingen in de bron vereist. Ook dient de levering voldoende metadata te bevatten om bij laden in het DIM te kunnen controleren dat opeenvolgende leveringen compleet en aaneensluitend zijn.

Identiek als bij incrementele leveringen geldt voor dakpan leveringen geldt dat een ander risico is dat wijzigingen in het bronsysteem buiten de tracking/logging om niet aan het DIM worden doorgegeven. Het gaat daarbij i.h.a. om reparatie-acties door IT.

Daarom kan een vorm van periodieke controle (bijvoorbeeld via reconciliatie tegen een volledige levering) noodzakelijk zijn om eventuele discrepanties tussen bron en DIM (door gemiste invoer en/of onverwachte verwijderingen) te kunnen corrigeren. Een dergelijke reconciliatie is vaak sowieso noodzakelijk om binnen het DIM te kunnen afleiden of er gegevens uit de bron verwijderd zijn.

Dakpan levering heeft alleen en in overleg de voorkeur indien er vertraging kan zijn in de bron met betrekking tot het verzamelen van alle door te leveren wijzigingen. Door de periode overlap is het mogelijk om gemiste gegevens alsnog in de volgende levering geleverd te krijgen. Hierdoor is er een beperkte mate van zelfreparerendheid mogelijk binnen de periode van de levering.

Naamgevingstandaarden

Dakpan leveringen vallen onder de indicator "i" voor de naamgevingconventies die gehanteerd worden.

⁵⁰ Binnen de afzienbare toekomst wordt echter geen leverfrequentie hoger dan dagelijks verwacht.