

# Plan van aanpak – Datakwaliteit integrale Klantreizen Versie 0.9

## Doel

In het kader van het interne auditrapport datakwaliteit is door de CDO in samenspraak met RvB de behoefte uitgesproken om te starten met het inrichten van gegevens kwaliteit beheerproces middels een "verbetertraject datakwaliteit".

Beheersmaatregelen zullen genomen worden, dewelke borgen dat de kwaliteit van data binnen de dienstverlening aan burgers en werkgevers van UWV aansluit op de beoogde gebruiksdoelen ('fit for use').

## Resultaat

Het uit te voeren plan zal voor de integrale klantreis "Ik ben ziek/arbeidsbeperkt" voor kritische gegevens ("data elementen") een verbetering opleveren van de data kwaliteit op elk drie data kwaliteit criteria.

Een permanente verbetering wordt geborgd door het inrichten en opleveren van een kwaliteit beheerproces ("PDCA-cyclus") per geselecteerd dataelement in de samenwerkende divisies.

De wijze waarop de verbetercyclus en verantwoording rondom de kwaliteit van data is vormgegeven zal helder en eenduidig vastgelegd zijn.

De gerealiseerde resultaten zijn een blauwdruk voor uitbreiding naar overige Integrale Klantreizen en kritische gegevens.

## Plan van Aanpak – Stappenplan

Datakwaliteit verbetering in een organisatie vergt een gedegen aanpak om tot een zinvol resultaat te komen.

Belangrijk is om zorgvuldig en stapsgewijs met implementatie te starten.

Het plan wordt gepresenteerd als een helder stappenplan, zodat er binnen alle divisies op een éénzijdige wijze geacteerd kan worden.

## Uitgangspunten voor succes:

We gebruiken de context van de integrale Klantreizen.

Gekozen is om de betrokken divisies, zijnde SMZ, Uitkeren, Werkbedrijf, Handhaven, Facilitair bedrijf, Klant & Service, Gegevensdiensten, BC&K intensief in de gegevens keten te laten samenwerken.

In de geselecteerde integrale klantreis "ik ben ziek" kiezen dataeigenaren één of meerdere kritische gegevens in het proces, waar fouten met impact voor de cliënt niet mogen voorkomen ("kritische dataelementen"). Hierbij naar eigen inzicht gebruik makend van de "methode Handhaven".

Voor het verbeteren van de datakwaliteit selecteren we drie (3) data kwaliteitscriteria: *Nauwkeurigheid, Volledigheid, Tijdigheid*

Een heldere en afgestemde governance (met RACI) zal aanwezig zijn na vaststelling van het plan door RvB en vóór begin van de start van het traject, waarbij de verantwoordelijkheid uiteraard blijft bij de deelnemende divisie directies.

Coördinatie en voortgang dienen te worden gegarandeerd door (bij voorkeur) één coördinerend product owner, één (gegevens)architect, twee (business)procesdeskundigen, één (klantreis) materiedeskundige aangevuld met experts (ad hoc) met business óf met IT-kennis ter verwerking en het leveren van ontbrekende expertise waar nodig.

Deze ad hoc expertise zal aangeleverd worden met hoge prioriteit.

Rapportage over de voortgang door de divisie directies voor het oplossen van deze RvB auditpunten zal zijn aan de RvB via de UWV CDO. Overige stakeholders zullen geïnformeerd blijven.

Dit initiatief zal ondersteunend zijn aan en aansluiten bij het implementeren van de vier speerpunten en eindbeelden als eerder vastgesteld door de Coalitie Gegevensmanagement.

Reeds lopende initiatieven zullen uitgenodigd worden om gerealiseerde verbeteringen direct in te zetten voor dit traject "Datakwaliteit integrale klantreis".

Stappen:

1. Selecteer een integrale klantreis voor deze datakwaliteit use cases.
  - De geselecteerde klantreis is 'Ik ben ziek/(deels) arbeidsongeschikt'.
2. Selecteer als dataeigenaar in de eigen divisie per integrale klantreis één of meerdere kritische gegevens (dataelementen). Deze moeten essentieel zijn voor de klantreis = als een kritisch gegeven niet van de juiste kwaliteit is, is de klantreis gedoemd te mislukken en heeft dit negatieve impact voor de cliënt en de klantwaarde. Bepalend hierbij is de impact van deze risico's op de dienstverlening aan burgers. Hierbij naar eigen inzicht gebruik makend van de "methode Handhaven".

Per deelnemende divisie zijn de volgende keuzes gemaakt:

- SMZ selecteert Eerste Arbeidsongeschiktheid Dag en tevens laatste Arbeidsongeschiktheid Dag
  - Mogelijke impact cliënt:  
Eerste AOD: Uitkering start niet op tijd of te laat, tevens nadelige gevolgen op de pensioenpremie van de cliënt
  - Intern UWV-risico:  
Laatste AOD: Scheve resultaten in stuur- en verantwoordingsinformatie  
Eerste AOD: Verkeerde conclusies/ uitspraken rond de werkvoorraad
- Uitkeren selecteert "Dagloon".
  - Mogelijke impact cliënt: Uitkering is te hoog/ te laag.
- Facilitair bedrijf selecteert "Pandcode"
  - Mogelijke impact cliënt: Niet tijdige opvolging van afspraken/(her)keuringen door verwijzing naar verkeerde UWV-pand.
- Handhaven selecteert "definitie gezondheidsfraude" (reeds lopend traject).
  - Mogelijke impact cliënt: Discussie rondom verdenking van gezondheidsfraude.
  - Intern UWV-risico: Gebrek aan een exacte definitie kan inefficiënte aanpak van potentiële fraudegevallen veroorzaken
- K&S selecteert "email adres cliënt" en "telefoonnummer cliënt".
  - Mogelijke impact cliënt: UWV kan de cliënt niet tijdig benaderen.
  - Intern UWV-risico: Afbreukrisico vertrouwen cliënt.
- Gegevensdiensten selecteert "werkhervattingskas (WHK) premiepercentage".
  - Mogelijke impact cliënt: De te betalen premie voor de publieke WIA/WGA verzekering te hoog of te laag. Een onnodige bezwaarzaken (bij GD) tot gevolg.
  - Intern UWV-risico: Veroorzaakt onnodig bezwaar-werklast en mogelijk hercalculatie van uitkeringen.
- Werkbedrijf heeft nog geen selectie gemaakt.
- BC&K zal kennis en kunde inzetten en meedenken in het opzetten van de governance rondom datakwaliteit. Dit betreft het meedenken in juiste formulering in

jaarplannen en managementcontracten van directies, het meedenken over rapportageschema's en het opzetten van een kwaliteit beheer- en borgingproces ("PDCA Cyclus") rond datakwaliteit.

3. Bepaal en leg vast hoe de dataelementen zich, de klantreis volgend, door de organisatie (procesmatig én technisch) bewegen. De gehele horizontale data lineage van record of origin (incl. bron van buiten UWV) tot en met aanbidding aan de cliënt van UWV. Vastleggen zal in ieder geval ook visueel in een data flow diagram met relevante metadata geschieden. Dit onder verantwoordelijkheid en akkoord van de dataeigenaar. Hiermee is er inzicht in het datagebruik in klantprocessen, de afnemers van deze data en kwaliteitseisen gesteld aan data.
4. Stel vast waar de kritische dataelementen elders binnen UWV worden gebruikt en geef dit eveneens aan in een data flow diagram (eventueel separaat van de flow in stap 3) – Vaak gaat het om kopieën, andere rapporten, meerdere databases, laptops ten behoeve van gebruiker gedefinieerde applicaties (bijvoorbeeld in excel) etc. ("User Defined Applications"). Dit onder verantwoordelijkheid en akkoord van de dataeigenaar.
5. Bepaal en noteer voor elk data-element de dimensies voor de drie data kwaliteitscriteria in nauwe samenwerking met business experts (zoals domein consultant, domeinhouder, uitvoering) en laat deze accorderen door de verantwoordelijke dataeigenaar.
  - Voorbeeld: De nauwkeurigheid van kritisch gegeven X moet bij de bron liggen tussen de waarden 100 en 14500.
6. Bepaal in samenwerking met de dataeigenaar methoden om de datakwaliteit per onderhanden gegeven structureel ("by design") te garanderen. Dit kan zowel een technische oplossing zijn als een procedurele oplossing.
  - Voorbeeld technische oplossing: Als een waarde niet aan de gestelde nauwkeurigheid voldoet, moet deze worden afgekeurd en kan geweigerd worden of kan verder gaan met een vervangende waarde of kan .....
  - Voorbeeld procedurele oplossing: Stel een kpi in op het percentage minimale correcte aanleveringen voor de gestelde kwaliteitscriteria voor de invoerende partij.
7. De dataeigenaar zorgt zo nodig voor het invoeren van een verandesignaal in de betrokken divisie. Hoge prioriteit zal aangegeven worden door of namens de dataeigenaar en/of domeinhouder voor elke kwaliteit beheer maatregel als onder 6 beschreven. Divisies en directies worden opgedragen het gegevenseigenaarschap door te vertalen door het nader toekennen van taken en verantwoordelijkheden binnen het eigen organisatieonderdeel.
8. Richt een nog vast te stellen UWV Datakwaliteit beheer en -borging proces voor de kritische data-elementen in scope als geselecteerd.
9. Rapporteer periodiek de bevindingen en verbeteringen van de datakwaliteit en de impact op integrale Klantreizen. De Integrale Klantreisleider(s) worden geïnformeerd over de vorderingen en resultaten. In de implementatiefase zal tevens de RACI vastgesteld zijn.
10. Leg de bevindingen uit dit eerste traject vast in een procedure voor volgende verbeter trajecten.
11. Start de volgende fase: Selecteer vervolg-Klantreizen en kritische dataelementen voor een volgende verbeter slag en invoering(en) in PDCA-cyclus in overleg met betrokken divisies en consulteer hiervoor integrale Klantreisleiders om optimaal en doeltreffend maximaal resultaat te bewerkstelligen.
12. Voer de volgende verbeter cyclus uit.

### **Planning, budget en tijdslijn**

Nader te bepalen in overleg met betrokken divisies, RvB, CDO en programma informatiehuishouding.

In de eerste fase na akkoord van de RvB zal dit budget verder in detail uit te werken. De eerste 4 maanden na start van de aanpak is reeds bekend qua budget en geborgd. Programma Informatiehuishouding is eveneens in principe akkoord met leveren van aanvullend programma budget na de eerste 4 maanden van het traject.

In de opstartfase zal een verder uitgewerkt plan op hoofdlijnen, inclusief een definitieve inschatting van kosten en financiering voorstel ter goedkeuring aan coalitieoverleg en programma Informatiehuishouding voorgelegd worden.

### **Besluitvorming / Governance lijn**

In de opstartfase ná initiële akkoord voor het plan van aanpak door RvB. Tevens zal een verder uitgewerkt plan op hoofdlijnen, inclusief besluitvorming en governance voorstel ter goedkeuring aan coalitie voorgelegd worden, waarna zo nodig voorlegging aan RvB zal plaatsvinden.

## ***Bijlage 1***

### Conclusie en Aanbevelingen uit Rapport "Audit Datakwaliteit Accountantsdienst UWV Feb2023"

#### Conclusie

Vanaf 2016 hebben CDO en Data-Office, in afstemming met divisies en directies, essentiële bouwstenen opgeleverd, die een bijdrage kunnen leveren aan de ambities op het gebied van datakwaliteit. Wel stellen wij vast dat op een aantal belangrijke aspecten verbetering nodig is om te borgen dat de kwaliteit van data binnen de dienstverlening aan burgers aansluit op de beoogde gebruikersdoelen ('fit for use'). Wij hebben tijdens de uitvoering van ons onderzoek binnen UWV een grote behoefte en bereidheid aangetroffen om het onderwerp 'datakwaliteit' verder op te pakken.

Wij constateren dat:

- een concrete planning van de UWV-doelstellingen rondom datakwaliteit ontbreekt;
- de UWV-strategie datakwaliteit nog niet compleet is;
- het UWV-brede beleid datakwaliteit decentraal nog niet voldoende geoperationaliseerd is;
- het datakwaliteit beheerproces decentraal nog onvoldoende is ingericht;
- er beperkt inzicht is in het datagebruik in klantprocessen, de afnemers van deze data en kwaliteitseisen die zij stellen aan data;
- de invulling van een aantal belangrijke operationele randvoorwaarden ontoereikend is;
- de verantwoording over en het toezicht op datakwaliteit binnen UWV nog niet optimaal is.

Door bovenstaande tekortkomingen bestaat het risico dat de kwaliteit van data onvoldoende is, waardoor de beoogde drempelloze klantreizen niet kunnen worden gerealiseerd. Wij adviseren de Coalitie Gegevensmanagement een UWV-verbeterplan 'datakwaliteit', inclusief planning en verdeling van verantwoordelijkheden, op te stellen en dit verbeterplan door de Raad van Bestuur vast te laten stellen. De bevindingen, risico's en aanbevelingen, zoals in hoofdstuk 7 van dit rapport zijn uitgewerkt, vormen de basis voor de invulling van dit verbeterplan.

Data-Office is, samen met divisies en directies, bezig om de UWV best cases op het gebied van datakwaliteit te inventariseren. Wij zijn van mening dat de wijze waarop de kwaliteit van kritische data binnen de Loonaangifteketen wordt gemonitord een potentiële 'best case' is.

#### Bevindingen, risico's en aanbevelingen

##### Strategie datakwaliteit

###### *Risico*

Doordat bij de doelstellingen op het gebied van datakwaliteit een concrete planning ontbreekt, bestaat het risico dat niet duidelijk is wanneer de ambitie visie Datagedreven werken wordt gerealiseerd.

###### *Aanbeveling*

Wij adviseren de Coalitie Gegevensmanagement concrete en haalbare (meerjaren) doelstellingen op datakwaliteit te laten formuleren, die het realiseren van de ambitie Datagedreven werken ondersteunen. Wij adviseren tevens deze doelstellingen, inclusief bijbehorende planning, door de Raad van Bestuur te laten bekrachtigen

## Strategie datakwaliteit nog niet compleet

### *Risico*

Door het issuegerichte karakter van de strategie bestaat het risico dat data-kwaliteitsdoelstellingen in relatie tot het verbeteren van dienstverlening onvoldoende vorm krijgen.

### *Aanbeveling*

Wij adviseren de Coalitie Gegevensmanagement om de pull strategie aan te vullen met een meer proactieve benadering en de klantgerichte elementen in de strategie te verankeren, zoals het integreren van datakwaliteit in klantreizen en voortbrengingsprocessen van applicaties. Wij adviseren tevens de aangevulde strategie te laten bekrachtigen door de Raad van Bestuur.

## Best cases datakwaliteit

### *Aanbeveling*

Wij adviseren de CDO om divisies en directies kennis te laten nemen van de Kwaliteitsmonitor LAK en met divisies en directies de toepasbaarheid van deze Kwaliteitsmonitor voor de klantprocessen over de divisies heen te onderzoeken.

## Beleid nog niet voldoende geoperationaliseerd

### *Risico*

Doordat divisies en directies het beleid niet hebben geoperationaliseerd kan geen sturing worden gegeven aan de verbetering van datakwaliteit. Hierdoor bestaat het risico dat de kwaliteit van data niet aansluit op gebruikersdoeleinden van afnemers ('fit for use') en komt de kwaliteit van data niet op het beoogde niveau.

### *Aanbeveling*

Wij adviseren de Coalitie Gegevensmanagement, na te (laten) gaan wat nodig is om het beleid op het gebied van datakwaliteit door divisies en directies te operationaliseren.

## Data-organisatie

### Verdere inrichting data-organisatie UWV breed

#### *Risico*

Doordat binnen divisies taken en verantwoordelijkheden op het gebied van datakwaliteit niet zijn belegd, ontbreekt bij hen het fundament om in de toekomst daadwerkelijk het geveenseigenaarschap in te vullen.

#### *Aanbeveling*

Wij adviseren de Coalitie Gegevensmanagement om divisies en directies het geveenseigenaarschap door te laten vertalen door het nader toekennen van taken en verantwoordelijkheden binnen het eigen organisatieonderdeel.

### Verbeteren UWV brede samenwerking op datakwaliteit

#### *Risico:*

Doordat geen overstemming is over de taken, verantwoordelijkheden en bevoegdheden op centraal en decentraal niveau wordt de UWV brede samenwerking op datakwaliteit bemoeilijkt.

#### *Aanbeveling:*

Wij adviseren de Coalitie Gegevensmanagement om, mede in het kader van de herinrichting van de data-organisatie, te (laten) onderzoeken in hoeverre de verdeling van taken,

verantwoordelijkheden en bevoegdheden tussen centraal en decentraal aangepast moet worden. Een eventuele verruiming van de bevoegdheden van de CDO kan hierin betrokken worden.

## Datakwaliteit beheerproces

### Datakwaliteit beheerproces nog onvoldoende ingericht

#### *Risico*

Door het niet gestructureerd inrichten van het datakwaliteit beheerproces bestaat het risico dat de dienstverlening gebaseerd is op foutieve data en op inconsistenties in externe stuur- en verantwoordingsinformatie.

#### *Aanbeveling*

Wij adviseren de Coalitie Gegevensmanagement te (laten)onderzoeken wat benodigd is om:

- conform beleid datakwaliteit, beheerprocessen binnen divisies en directies in te richten;
- divisie/directie overstijgende datakwaliteit beheerprocessen in te richten waarlangs correcties van data in bronsystemen plaats kunnen vinden.

## Gegevensverwerking- en uitwisseling

### Beperkt inzicht in datagebruik

#### *Risico*

Door het ontbreken van inzicht in datagebruik bestaat het risico dat divisies en directies niet in staat zijn om kwaliteitseisen van afnemers te inventariseren en conform 'fit for use' te werken aan de verbetering van de kwaliteit van data in klantprocessen.

#### *Aanbeveling*

Wij adviseren de Coalitie Gegevensmanagement te zorgdragen voor een uitwerking van de data-lineage per klantproces. [= *Horizontale lineage oftewel data stromen, die de klantreis tonen*])

## Afnemers en afnemerseisen

### Beperkt inzicht in afnemers en afnemerseisen

#### *Risico*

Doordat onduidelijk is welke kwaliteitseisen afnemers aan data stellen, bestaat het risico dat de data die zij in hun processen gebruiken niet betrouwbaar zijn. Dit kan negatieve gevolgen hebben voor de kwaliteit van de dienstverlening aan burgers en de gegevens- en informatie-uitwisseling met externe organisaties.

#### *Aanbeveling*

Wij adviseren CDO om te monitoren dat divisies en directies afnemerseisen inventariseren en vastleggen.

## Operationele randvoorwaarden

### Invulling operationele randvoorwaarden ontoereikend

#### *Risico*

Het risico bestaat dat algemeen directeuren/directieteams, CDO en RvB door de huidige informatievoorziening niet in voldoende mate in staat worden gesteld om te sturen op en te verantwoorden over (de verbetering van) datakwaliteit.

#### *Aanbeveling*

Wij adviseren CDO, divisies en directies een PDCA-cyclus op het gebied van datamanagement in te richten, aan de hand waarvan:

- inzicht is in de kwaliteit van data in de klantprocessen van UWV;
- sturing op verbetering van de kwaliteit van data in klantprocessen kan plaatsvinden.

De conclusies en aanbevelingen in het auditrapport worden middels het plan van aanpak opgevolgd en aangepakt, als verwoord in de voorlegger voor de RvB vergadering van 23 juni 2023:

- o Wij adviseren de Coalitie Gegevensmanagement concrete en haalbare (meerjaren) doelstellingen op datakwaliteit te laten formuleren, die het realiseren van de ambitie Datagedreven werken ondersteunen. Wij adviseren tevens deze doelstellingen, inclusief bijbehorende planning, door de Raad van Bestuur te laten bekrachtigen.
  - o In het plan van aanpak wordt de basis gelegd voor een integrale en gestructureerde aanpak, die aansluit bij de (meerjaren) doelstellingen op datakwaliteit en de roadmaps datagedreven werken in de divisies. Deze doelstellingen en plannings zijn vastgesteld in het coalitieoverleg.
- o Wij adviseren de Coalitie Gegevensmanagement om de pull strategie aan te vullen met een meer proactieve benadering en de klantgerichte elementen in de strategie te verankeren, zoals het integreren van datakwaliteit in klantreizen en voortbrengingsprocessen van applicaties. Wij adviseren tevens de aangevulde strategie te laten bekrachtigen door de Raad van Bestuur.
  - o Zie plan aanpak en we hanteren nog steeds de pull strategie, die aangevuld wordt met een stevigere regievoering vanuit de Data Office – CDO
  - o Een risicogerichte aanpak is gevolgd langs de lijn van de klantreis "ik ben ziek/ arbeidsongeschikt"
- o Wij adviseren de CDO om divisies en directies kennis te laten nemen van de Kwaliteitsmonitor LAK en met divisies en directies de toepasbaarheid van deze Kwaliteitsmonitor voor de klantprocessen over de divisies heen te onderzoeken.
  - o Gegevensmanagers in divisies zijn op de hoogte gebracht en hebben de opdracht om de toepassing in de klantprocessen te onderzoeken.
- o Wij adviseren de Coalitie Gegevensmanagement, na te (laten) gaan wat nodig is om het beleid op het gebied van datakwaliteit door divisies en directies te operationaliseren.
  - o Zie plan van aanpak
- o Wij adviseren de Coalitie Gegevensmanagement om divisies en directies het geveenseigenaarschap door te laten vertalen door het nader toekennen van taken en verantwoordelijkheden binnen het eigen organisatieonderdeel.
  - o In het plan wordt in elke divisie door de "eigenaar" van het gegeven (in de meeste gevallen de domeinhouder) gegeven(s) geselecteerd, die ook in de kwaliteitsbeheer en -borging cyclus opgenomen zal worden.
- o Wij adviseren de Coalitie Gegevensmanagement te (laten) onderzoeken wat benodigd is om: conform beleid datakwaliteit beheerprocessen binnen divisies en directies in te richten; divisie/directie overstijgende datakwaliteit beheerprocessen in te richten waarlangs correcties van data in bronsystemen plaats kunnen vinden.
  - o Het plan van aanpak bevat het invoeren van Governance, waaronder taken, verantwoordelijkheden en beheerprocessen.
- o Wij adviseren de Coalitie Gegevensmanagement zorg te dragen voor een uitwerking van de data-lineage per klantproces.
  - o Dit is één van de stappen in het plan van aanpak en in ruimere zin geadresseerd in de Roadmap Gegevensmanagement 2023-2025

- Wij adviseren CDO om te monitoren dat divisies en directies afnemereisen te inventariseren en vastleggen.
  - Afnemereisen beginnen met het vaststellen van de criteria voor de gegevens in dit plan. Meer afnemereisen vaststellen is een vervolgstap na uitvoering van dit plan van aanpak.
- Wij adviseren de Coalitie Gegevensmanagement om divisies en directies opdracht te geven om maatregelen te treffen op nog niet ingevulde operationele randvoorwaarden en zich hieromtrent te verantwoorden.
  - Divisies hebben in de Coalitie Gegevensmanagement eindbeelden vastgesteld ter invulling van de strategie van UWV. Deze eindbeelden zijn opgesteld om operationele randvoorwaarden op vier speerpunten, te weten Data Kwaliteit, Metadata Management, Masterdata Management en Data Lifecycle Management te realiseren. Divisies hebben daartoe "roadmaps" opgesteld en rapporteren de voortgang in de coalitie gegevensmanagent ter bewaking van de voortgang en onderlinge afstemming.
- Wij adviseren CDO, divisies en directies een PDCA-cyclus op het gebied van datamanagement in te richten, aan de hand waarvan: inzicht is in de kwaliteit van data in de klantprocessen van UWV; sturing op verbetering van de kwaliteit van data in klantprocessen kan plaatsvinden.
  - Het op te leveren Datakwaliteit Beheer- en Borging proces ("PDCA cyclus") is een essentieel onderdeel van dit initiatief en vult daarmee deze aanbeveling in.
- DO is stelselverantwoordelijk voor datakwaliteit en geeft aan op welke wijze de architectuur een rol speelt in datakwaliteit.
  - Data Office heeft de UWV Data Architectuur beschreven en afgerond. Een "Proof of Concept" is in voorbereiding en zal de architectuur op onderdelen toetsen. De Data Architectuur zal eveneens in de AB behandeld worden. Deze architectuur ondersteunt het verbeteren van de datakwaliteit.

## **Bijlage 2 - Meetbaarheid van datakwaliteit**

Datakwaliteit moet meetbaar zijn om de risico's op schade te beperken. De eisen hiervoor zijn afhankelijk van het doel waarvoor de data wordt vastgelegd en gebruikt. Een algemene lijst met criteria bestaat niet; dit is voor iedere klantreis ("customer journey" en iedere afdeling verschillend. Maar kwaliteitscriteria zijn essentieel om goed te kunnen meten.

Door ervaringen in datakwaliteit projecten, zijn kwaliteitscriteria zeer belangrijke criteria om te toetsen hoe het gesteld is met de data van een klantreis:

### **Nauwkeurigheid:**

Is de ingevoerde data correct? In welke mate reflecteert de data de werkelijkheid? Dit vergt kennis van de inhoud van de data ten opzichte van hoe het zich weerlegt in het echt. Aan de hand van kennis van de waarden in de business kun je testen of je data de werkelijkheid goed weergeeft;

*Toelichting:* Het netto jaarinkomen van een bepaalde persoon wordt geschat op € 28.000 terwijl dit werkelijk € 25.589 bedroeg.

### **Volledigheid:**

Is alle data aanwezig? Denk hierbij aan het hebben van missende waarden en het hebben van de juiste variabelen om je onderzoek te kunnen doen of je dienstverlening uit te kunnen voeren;

*Toelichting:* Onderdeel a) betreft het verschil tussen de variabelen die geleverd worden en die gevraagd zijn. Bij b) gaat het feitelijk om de volledigheid van de metadata omschrijving, bevat deze alle voor de gebruiker benodigde definities en classificaties (dit zegt dus *niets* over de duidelijkheid). Onderdeel c) slaat op het benodigde detailniveau van de gegevens. Zijn voldoende gegevens op het detailniveau aanwezig waarover de gebruiker uitspraken wil.

**Tijdigheid:** De mate waarin de gewenste datum van beschikbaarheid (het vrijgeven) van gegevens in de tijd afstaat van het einde van de periode waarop de gegevens betrekking hebben (Is niet hetzelfde als Punctualiteit). Is de data beschikbaar op het moment dat je het nodig hebt? Een voorbeeld is data die nodig is voor maand- of kwartaalrapportages. Wanneer de data niet op tijd beschikbaar is, geef je niet altijd een compleet beeld weer voor hetgeen dat je middels data wilt laten zien. *Toelichting:* Gegevens over de maand mei van het jaar 2008 zijn in december 2008 voor afnemers beschikbaar.

### **Samenhang:**

Kan data in verschillende systemen gebruikt worden? Is de data gelijk in alle systemen? Dit kan misgaan wanneer je bijvoorbeeld data vanuit verschillende bronnen trekt en deze verbindt ondanks de verschillen. Het is daarom in ieder geval belangrijk om de metadata (tekst en uitleg over je data) goed bij te houden;

### **Authenticiteit**

De mate waarin bekend is dat een gegeven is gemaakt of verzonden door de persoon of organisatie die beweert het te hebben gemaakt of verzonden.

*Toelichting:* Voor een afnemer moet onomstotelijk vast staan dat een bepaalde gegevensset afkomstig is van een bepaalde leverancier. Dit geldt uiteraard ook voor de gegevens van individuen.

### **Beschikbaarheid (Toegankelijkheid)**

De mate waarin gegevens aanwezig en verkrijgbaar zijn (geredeneerd vanuit de afnemer). Kan ook op de verschillend vormen van de fysieke drager van de gegevens slaan (bijv. papier, tape en elektronisch). Hoe makkelijk kunnen personen beschikken over de data?

*Toelichting:* Bij voorkeur dient de leverancier van de gegevens alle door de afnemer gevraagde gegevens aan te bieden. Soms is dat echter niet mogelijk, bijv. wanneer een gevraagd gegeven niet aanwezig is of omdat hiervoor geen toestemming is verkregen of kan worden gegeven.

### **Beveiliging**

De mate waarin opzettelijk of abusievelijk ongeoorloofde toegang wordt voorkomen (Een IT-interpretatie).

*Toelichting:* Het versleutelen van berichten die van de ene naar de andere overheidsinstelling worden verzonden is een voorbeeld van beveiliging.

### **Continuïteit**

De mate van samenhang van gegevens van dezelfde variabelen in de tijd. Dit wordt ook wel vergelijkbaarheid in de tijd genoemd.

*Toelichting:* De lengte van de tijdsperiode waarin gegevens nog direct vergeleken kunnen worden. Door revisies of wijzigingen in de definitie van een variabele kan deze periode beperkt zijn tot bijvoorbeeld 5 of 10 jaar.

### **Controleerbaarheid**

De mate waarin het eindresultaat (de output) geheel herleid kan worden tot de begingegevens (de input). Zou je ook Reproduceerbaarheid of Traceerbaarheid kunnen noemen.

*Toelichting:* Deze is optimaal wanneer de verwerking van de gegevens volledig kan worden gereconstrueerd.

### **Dekking**

De mate waarin de eenheden in de gegevens set overeenkomen met de eenheden in de beoogde te beschrijven populatie.

*Toelichting:* Het CBS wil een bestand met loongegevens van alle werknemers in Nederland gebruiken voor een publicatie over het gemiddelde inkomen van de bewoners van Nederland. Hiertoe ontbreken in het bestand echter o.a. de gegevens van zelfstandigen. Dit wordt onder dekking genoemd. Ook overdekking komt voor, zo bevat het bestand gegevens van personen die niet in Nederland wonen maar er wel werken; de z.g. grensarbeiders. Deze wonen vaak in Duitsland of België.

### **Duidelijkheid**

De mate waarin:

- 1) er sprake is van toereikende metadata,
- 2) gegevens worden geïllustreerd met behulp van grafieken en kaarten,
- 3) er informatie beschikbaar is over de kwaliteit van de gegevens,
- 4) de beperkingen in het gebruik van de gegevens zijn beschreven en
- 5) er wordt voorzien in aanvullende ondersteuning.

*Toelichting:* Er wordt gezorgd dat voldoende (meta)informatie aanwezig voor de gebruiker om de gegevens correct te kunnen interpreteren. Daarnaast kan ook extra ondersteuning worden geboden door de leverancier.

### **Gedetailleerdheid**

De mate waarin deelpopulaties kunnen worden onderscheiden in een bestand.

*Toelichting:* Geeft de mate van de beschikbaarheid van detailinformatie aan binnen een gegevens set. Hoe gedetailleerder, hoe meer specifieke gegevens over individuele eenheden in een bestand zijn opgenomen. Een bestand met gegevens over gemeenten is minder gedetailleerd dan een bestand met gegevens over de inwoners van die gemeenten.

### **Juistheid**

De mate waarin de gegevens in een bestand overeenkomen met de werkelijke gegevens van de betreffende variabelen voor elke eenheid.

*Toelichting:* Maat voor de afwezigheid van fouten (de foutloosheid) in een gegevensset. Voor de gegevens van de Loonaangifteketen is de loonstrook de administratieve weergave van de feitelijke werkelijkheid. Fouten zullen dan ook met name betrekking hebben op verschillen tussen de gegevens in het bestand en die op de loonstrook

### **Koppelbaarheid**

- 1) De mate waarin een gegevensset aan een andere gegevensset (met dezelfde eenheden) gerelateerd kan worden

2) Gemak waarmee een systeem gegevens kan uitwisselen met andere systemen (Zou je Interoperabiliteit kunnen noemen).

*Toelichting:* Onderdeel 1) heeft betrekking op bestanden. Het betreft de mate waarmee gegevens over individuen in 2 of meer verschillende bestanden succesvol kunnen worden samengevoegd. Hierbij wordt gestreefd naar zoveel mogelijk correcte koppelingen en zo min mogelijk miskoppelingen. Onderdeel 2) heeft betrekking op systemen. Het betreft de mate van succes waarmee systemen onderling gegevens kunnen uitwisselen. Bij niet-koppelbare systemen is uitwisseling alleen indirect mogelijk

### **Privacy**

De relatie tussen het doel waarvoor gegevens gebruikt worden en de mate waarin het noodzakelijk is de individuele eenheden in de gegevens te herkennen. (Zou je Vertrouwelijkheid of Doelbinding kunnen noemen).

*Toelichting:* In een tabel met het gemiddelde inkomen van bepaalde beroepen in de gemeente Amsterdam is de waarde voor het beroep 'burgemeester' verwijderd. Hiervan is er immers maar één.

### **Punctualiteit**

De tijd die zit tussen het werkelijke aanlevertijdstip en het geplande aanlevertijdstip.

*Toelichting:* Gegevens die volgens afspraak uiterlijk om 10:00 uur op de 1e werkdag van de maand moeten worden geleverd blijken in de praktijk (regelmatig) pas de 2e werkdag afgegeven te worden.

### **Relevantie**

De mate waarin gegevens voldoen aan de behoefte van huidige en potentiële gebruikers. Is de data bruikbaar voor het doel waar jij het voor nodig hebt? Het is belangrijk om doelen te stellen op basis van de analyse voordat je data verzamelt. Voordat je begint met (bijvoorbeeld) een analyse, controleer altijd of de data wel relevant is voor het gegeven doel. Dit kan veel tijd besparen.

### **Validiteit**

Is de data in het juiste formaat en is het in lijn met de regels binnen jouw organisatie?

Denk hierbij aan checks zoals dat de datum ook als juiste datum soort is opgeslagen, dat getallen als geheel of decimaal getal moeten worden opgeslagen, enz;

### **Redundantie**

Is dezelfde data op verschillende plekken terug te vinden en/of opgeslagen?

Wanneer dit het geval is wordt er niet efficiënt omgegaan met de data. Als gevolg hiervan kan de database groter zijn dan nodig is, wat leidt tot hogere kosten en fouten door kopiëren of niet bijhouden van data.

*Toelichting:* Het gebruik van gegevens in de openbaar toegankelijke statistische database van het CBS levert informatie op over de behoefte aan gegevens. Van statistieken die niet of nauwelijks worden geraadpleegd kan men zich afvragen of deze (nog) relevant zijn.

### **Samenhang**

De mate waarin overeenkomstige gegevens uit verschillende bronnen gebruikt en gecombineerd kunnen worden. Dit hangt af van de mate waarin gegevens voldoende nauwkeurig zijn en de variabelen, die bij die gegevens horen, dezelfde betekenis hebben.

*Toelichting:* Het CBS produceert statistieken over verwante onderwerpen. Hierbij wordt gebruik gemaakt van zelf verzamelde gegevens en ook van gegevens die van anderen afkomstig zijn. De mate waarin beide sets van gegevens vergelijkbare eindresultaten opleveren (over soortgelijke fenomenen) duidt op samenhang.