

Data Quality Management Proces

Borging van de datakwaliteit door de keten heen...

Enterprise Data Management (EDM) i.s.m. Data Management Organisatie

Versie 1.1 – april 2023

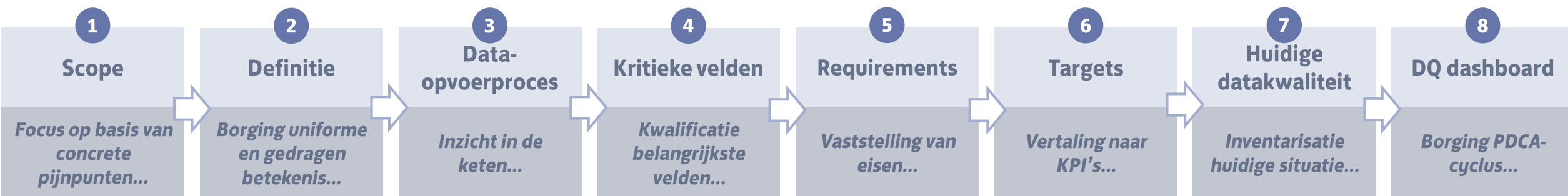
Waarom een Data Quality Management Proces?

Kaders en richtlijnen voor het inrichten van effectief Data Quality Management...

- **Het op een structurele manier monitoren, verbeteren en in stand houden van een zo hoog mogelijke datakwaliteit door de keten heen vraagt om een gedegen en planmatige aanpak.**
- **Dit Data Quality Management Proces is bedoeld als raamwerk waar elke Data Manager voor zijn/haar eigen Business Unit en data domein(en) direct mee aan de slag kan.**
- **Hij/zij ziet er voor zijn/haar eigen data objecten op toe dat de hierna te noemen stappen daadwerkelijk worden uitgevoerd. En dat deze periodiek worden getoetst op actualiteit. Hij/zij is vrij om te kiezen in welke vorm dat gebeurt; als het maar gebeurt.**
- **De Data Manager wordt bij de uitvoering daar waar nodig en gewenst ondersteund door de afdeling Enterprise Data Management (EDM). In het bijzonder door de Data Quality Analisten.**
- **Zij zien er ook op toe dat het proces periodiek wordt getoetst en daar waar nodig wordt geactualiseerd.**

Het Data Quality Management Proces

De stappen...



Het Data Quality Management Proces

Een korte toelichting op elk van de stappen...

	Stap	Toelichting
1	Scope	De eerste stap in het proces is het duidelijk vaststellen van de scope. Dat betekent dat wordt bepaald op welk data object c.q. welke data objecten de focus ligt. Hierbij wordt heel duidelijk gekeken naar wat nu concreet de belangrijkste pijnpunten zijn. Een rondje langs de velden bij een aantal stakeholders zorgt er vaak al voor dat duidelijk wordt welke objecten voor de belangrijkste processen binnen de Business Unit(s) van belang zijn. Dit signaal kan overigens uit verschillende hoeken van de organisatie komen. Belangrijkste in dit stadium is om met die objecten te starten waar of de grootste pijn zit of waar vrij eenvoudig een aantal quick wins kunnen worden gerealiseerd.
2	Definitie	Voor het object dat in stap 1 is gekozen, wordt in dit stadium vastgesteld wat de exacte definitie is. in de praktijk komt het namelijk nog weleens voor dat we denken over hetzelfde te praten, maar dat er in werkelijkheid verschillen zijn. Een verschil dat nu misschien niet van belang lijkt te zijn, maar verderop in het proces wel degelijk tot fundamentele andere beslissingen zou kunnen leiden. Het is dus evident dat er een uniforme én gedragen definitie is.
3	Data opvoerproces	In dit stadium wordt het proces voor het opvoeren, wijzigen en verwijderen van een data object in kaart gebracht. Zo wordt gekeken naar de verschillende applicaties die hierin een rol spelen; met een focus op de applicatie waar het data object wordt opgevoerd c.q. gegenereerd ('System of Entry') en de applicatie waarin de waarheid wordt vastgelegd. ('System of Record') Ook de personen die hierin een rol spelen worden in kaart gebracht en vastgelegd.
4	Kritieke velden	Een data object bestaat doorgaans uit meerdere velden en/of attributen. Niet al deze velden en attributen zijn even belangrijk. Op basis van een door de afdeling Enterprise Data Management (EDM) opgestelde kwalificatie wordt in dit stadium voor elk van de onderliggende velden en attributen bepaald hoe belangrijk elk individueel veld/attribuut is.
5	Requirements	Voor de velden en/of attributen die het meest belangrijk zijn, wordt in dit stadium bepaald welke eisen hieraan worden gesteld. Hierbij wordt dan gekeken naar de door EDM vastgestelde dimensies voor datakwaliteit: volledigheid, uniciteit, tijdigheid, validiteit, nauwkeurigheid en consistentie. Dit zijn feitelijk de beslisregels waar een Data Quality Analyst vervolgens mee aan de slag gaat.
6	Targets	Nadat de requirements zijn opgesteld, worden de targets vastgesteld. Voor elk van de velden en/of attributen wordt dit vertaald naar KPI's. Beginnend op het laagste niveau en vervolgens doorvertaald naar een geaggregeerd niveau zodat deze kunnen worden opgenomen in het dashboard.
7	Huidige datakwaliteit	In dit stadium wordt gekeken naar de huidige kwaliteit van het data object. De opgestelde requirements worden doorvertaald naar de meer technische beslisregels in de applicatie waar de kwaliteit daadwerkelijk wordt gemeten.
8	DQ dashboard	De kwaliteit van het data object dient door de tijd te worden gemeten en gemonitord. Hiervoor wordt gebruik gemaakt van het Data Quality Dashboard. Dit is een dashboard dat door de DMO van de Business Unit wordt gebruikt om periodiek te toetsen wat de status is. En kan worden gebruikt om daar waar nodig bij te sturen. Dit Dashboard is op verschillende aggregatieniveaus beschikbaar zodat deze zowel door de Data Steward (operatie), de Data Manager (DMO Overleg) als de Strategisch Data Eigenaar (Data Domein Overleg en DGB) kan worden gebruikt.

1. Scope

Focus op basis van concrete pijnpunten...

Doelen	Resultaten	
1.1. Alle objecten binnen het Data Domein zijn vastgelegd. (bijv. m.b.v. de data classificatie piramide)	[...]	
1.2. De pijnpunten op het gebied van data zijn in kaart gebracht én vastgelegd in een 'product backlog'. [de vorm waarin dit wordt bijgehouden is vrij]	[...]	
1.3. De pijnpunten zijn gekoppeld aan een of meerdere data objecten.	[...]	
1.4. De data objecten zijn geprioriteerd en dus is het duidelijk met welk(e) data object(en) wordt gestart.	[...]	

2. Definitie

Borging uniforme en gedragen betekenis...

Doelen	Resultaten	
2.1. De personen die een belang hebben bij het vaststellen van een uniforme en eenduidige formulering zijn in kaart gebracht.	[...]	
2.2. De definitie van het betreffende data object is vastgesteld én wordt gedragen door de belangrijkste stakeholders.	[...]	
2.3. De Strategisch Data Eigenaar is akkoord met de definitie van het data object.	[...]	
2.4. De definitie van het data object is op een centrale en toegankelijke plek vastgelegd.	[...]	

3. Data-opvoerproces (1/3)

Inzicht in de keten...

Doelen	Resultaten	
3.1. Vastgesteld is waar, bij wie het data object ontstaat. (Point of Origin)	[...]	
3.2. Vastgesteld is wie de input aanlevert voor opname in het System of Entry.	[...]	
3.3. Vastgesteld is wie bepaalt of het data object ook daadwerkelijk mag worden opgevoerd.	[...]	
3.4. Vastgesteld is wie (naam, rol en/of functie) het object mag opvoeren.	[...]	
3.5. Vastgesteld is wie toetst of de opvoer van het data object correct heeft plaatsgevonden (i.e. Data Steward).	[...]	
3.6. De applicatie waarin het data object wordt opgevoerd ('System of Entry') is vastgesteld.	[...]	
3.7. Vastgesteld is welke preventieve maatregelen zijn ingericht m.b.t. de opvoer / generatie van data in het System of Entry (denk aan referentietabellen, drop-down menu's, etc.)	[...]	

3. Data-opvoerproces (2/3)

Inzicht in de keten...

Doelen	Resultaten	
3.8. De applicatie waarin de waarheid van het data object wordt vastgelegd ('System of Record') is vastgesteld.	[...]	
3.9. Vastgesteld is op welke manier het data object wordt aangeleverd aan het System of Record.	[...]	
3.10 Het 4-ogen principe is vastgesteld.	[...]	
3.11. Vastgelegd is wie (naam, rol en/of functie) verantwoordelijk is voor de distributie van het data object naar de ontvangende systemen. ('System of Use') [optioneel]	[...]	
3.12. Vastgesteld is wie (naam, rol en/of functie) borgt dat de afnemende systemen (op een juiste manier) gebruik maken van het data object. [optioneel]	[...]	
3.13. De applicaties die gebruik maken van het data object ('System of Use') zijn in kaart gebracht. [optioneel]	[...]	
3.14. Vastgesteld is wie (naam, rol/of functie) mutaties in het data object mag doorvoeren en/of hiervoor akkoord moet geven. [optioneel]	[...]	

3. Data-opvoerproces (3/3)

Inzicht in de keten...

Doelen	Resultaten	
3.15. Vastgesteld is wie (naam, rol/of functie) (de inhoud van) het data object mag verwijderen. [optioneel]	[...]	

4. Kritieke velden

Kwalificatie belangrijkste velden...

Doelen	Resultaten	
4.1. Alle velden en/of attributen van het data object zijn vastgelegd. [Dit kan een export zijn van een bepaalde data set.]	[...]	
4.2. Elk van deze velden en/of attributen is gekwalificeerd (qua impact, mate waarin data kritiek is). [De data manager bepaalt wat 'kritiek' is; dit kan per proces verschillen.]	[...]	

5. Requirements

Vaststelling van eisen...

Doelen	Resultaten	
5.1. Voor elk van de velden en/of attributen zijn de requirements vastgesteld. [Hierbij wordt gekeken op welke van de DQ dimensies gemeten wordt.]	[...]	
5.2. Vastgesteld is in welke systemen en/of applicaties de kwaliteit wordt gemeten. [Idealiter gebeurt dit conform de doelarchitectuur in Colibra DQ; maar de data manager is hier vrij in.]	[...]	
5.3. De requirements zijn afgestemd met de data managers van de BU's die deze velden en/of attributen afnemen.	[...]	

6. Targets

Vertaling naar KPI's...(1)

Doelen	Resultaten	
6.1. Voor elk van de velden en/of attributen zijn de targets vastgesteld. [Dit kan ook voor een combinatie van velden gelden.]	[...]	
6.2. De huidige datakwaliteit is afgezet tegen de eerder opgestelde targets.	[...]	
6.3. De targets zijn door Strategisch Data Eigenaar definitief vastgesteld.	[...]	
...	[...]	

(1) In de praktijk kan het pragmatischer zijn om eerst de doelen onder stap 7 te doen alvorens met stap 6 te beginnen. De data manager heeft de vrijheid om dit te doen. Belangrijkste is dat er kritisch wordt nagedacht over wat de targets zouden moeten zijn en dat er niet vanuit gemak gekozen wordt voor een relatief lage target. Tegelijkertijd is ook het stellen van het doel dat alles 100% moet zijn niet realistisch. Wat kan helpen, is om eerst een 'voorkwalificatie' te maken met bijv. met low, medium en high. Om na het vaststellen van de huidige datakwaliteit de daadwerkelijke targets vast te stellen.

7. Huidige datakwaliteit

Inventarisatie huidige situatie...(1)

Doelen	Resultaten	
7.1. Vastgesteld is wie de kwaliteit van de data objecten meet en in welke frequentie dit gebeurt.	[...]	
7.2. Voor elk veld en/of attribuut zijn de opgestelde requirements concreet vertaald naar de meer 'technische' beslisregels in de DQ-tool.	[...]	
7.3. De tooling en de te gebruiken methodiek is vastgesteld.	[...]	

(1) In de praktijk kan het pragmatischer zijn om eerst deze doelen op te pakken alvorens met stap 6 te beginnen. De data manager heeft de vrijheid om dit te doen. Belangrijkste is dat er kritisch wordt nagedacht over wat de targets zouden moeten zijn en dat er niet vanuit gemak gekozen wordt voor een relatief lage target. Tegelijkertijd is ook het stellen van het doel dat alles 100% moet zijn niet realistisch. Wat kan helpen, is om eerst een 'voorkwalificatie' te maken met bijv. met low, medium en high. Om na het vaststellen van de huidige datakwaliteit de daadwerkelijke targets vast te stellen.

8. DQ Dashboard

Borging PDCA-cyclus...

Doelen	Resultaten	
8.1. Er is een proces ingericht dat de datakwaliteit op een frequente en geautomatiseerde manier weergeeft in het DQ Dashboard.	[...]	
8.2. De manier waarop de uitkomsten worden gedeeld, met wie en in welke vorm is vastgelegd.	[...]	
8.3. Afspraken zijn gemaakt over het eventueel herijken van de bestaande opzet.	[...]	
8.4. De aansluiting op het DQ issue proces is ingeregeld.	[...]	
8.5. Er is een proces ingericht dat de kwaliteit op een frequente en continue basis monitort.	[...]	