

Visie document Data Quality Management

Grip op data is cruciaal...

Versie 0.1 • April 2021

Inhoudsopgave

1. Introductie
2. Strategie
3. PostNL Data Principles
4. Cultuur & Governance
5. Bedrijfsprocessen
6. Besturing & Organisatie
7. Applicatie Architectuur
8. Metrics
9. Roadmap
10. “*Best practices*”



1. Introductie

Waarom een doelarchitectuur

Definitie, scope en positionering

Samenhang met andere data capabilities

Waarom een doelarchitectuur

Richting geven, referentiekader & doorvertaling naar dagelijkse praktijk

Dit document beschrijft de visie van PostNL op Data Quality Management en de praktische doorvertaling daarvan naar de dagelijkse praktijk. Data Quality Management is een van de capabilities binnen Enterprise Data Management en een goede invulling ervan zien we als randvoorwaardelijk voor de digitale transformatie die we binnen PostNL willen doorvoeren.

Het heeft als ambitie om **richting** te **geven** aan alle initiatieven op het gebied van datakwaliteit PostNL breed en verteld hoe wij in de **dagelijkse praktijk** binnen PostNL met Data Quality aan de slag gaan. Zowel qua rollen en verantwoordelijkheden (i.e. Besturing) als qua tooling (i.e. Architectuur).

Ook dient het als **referentiekader** voor toekomstige beslissingen op dit vlak en is het een document waar we altijd op kunnen terugvallen.

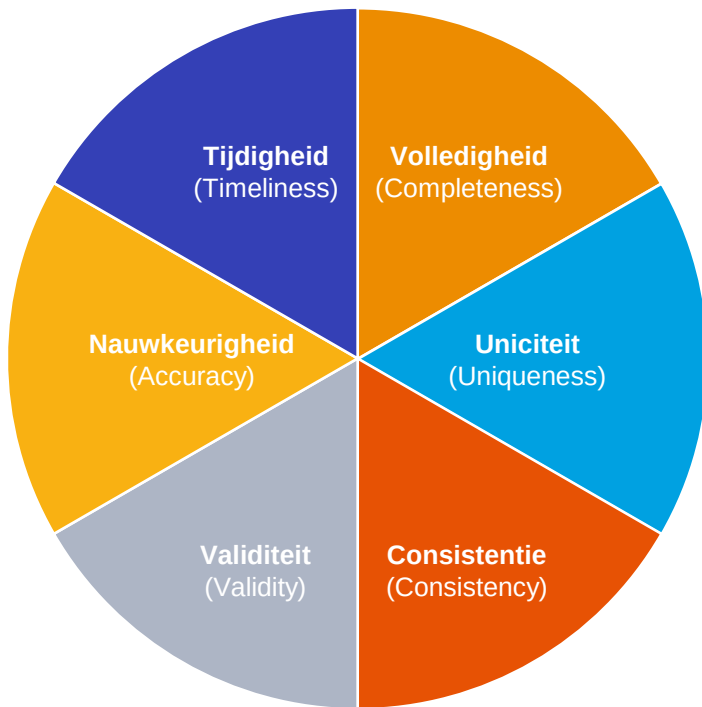
Hoewel getracht is zo volledig mogelijk te zijn, wordt dit document periodiek geijkt om te toetsen of wordt voldaan aan de dan geldende ontwikkelingen, visie en ambities van PostNL.

Geprobeerd is om de hoofdlijnen aan te houden. Praktische hulpmiddelen worden wel benoemd, maar voor verdere detaillering wordt verwezen naar de betreffende documentatie.

Definitie, scope en positionering

Voor héél PostNL én voor álle data

Datakwaliteit is de mate waarin de data(set) geschikt is voor het doel waarvoor het gebruikt wordt. **Data Quality Management** betreft de processen en technologieën om dat voor de gehele PostNL organisatie én voor álle data op een structurele manier in te richten en te borgen. Het is een programma, géén project.

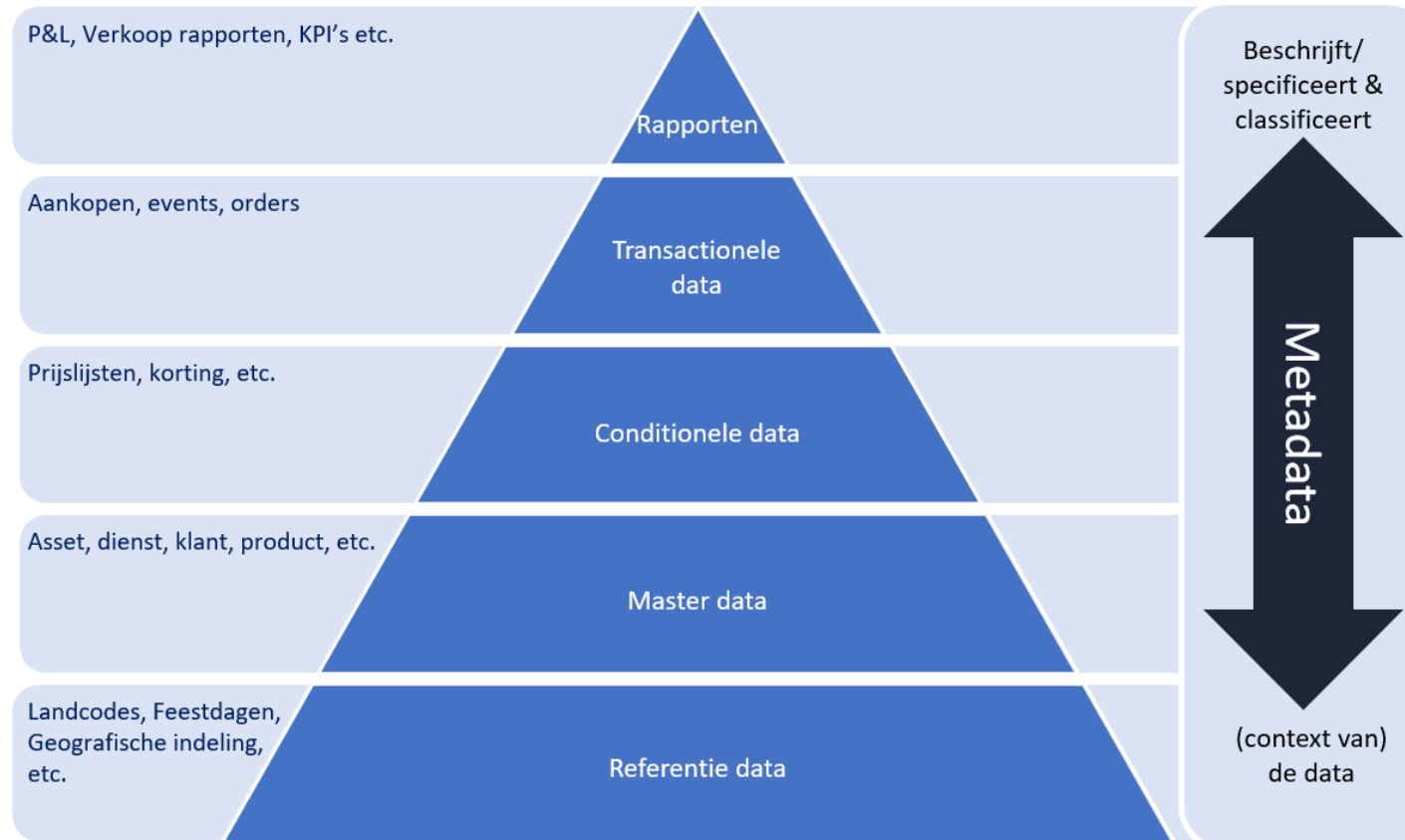


- **Volledigheid**: Mate waarin verwachte data aanwezig is
- **Uniciteit**: Mate waarin de data uniek is, afhankelijk van het gedefinieerde niveau
- **Tijdigheid**: Mate waarin data beschikbaar is op het moment dat het nodig is
- **Validiteit**: Mate waarin de data valt binnen het gedefinieerde formaat, type en reeks
- **Nauwkeurigheid**: Mate waarin de data overeenkomt met de werkelijkheid
- **Consistentie**: Mate waarin de data tussen verschillende datasets overeenkomt

Definitie, scope en positionering

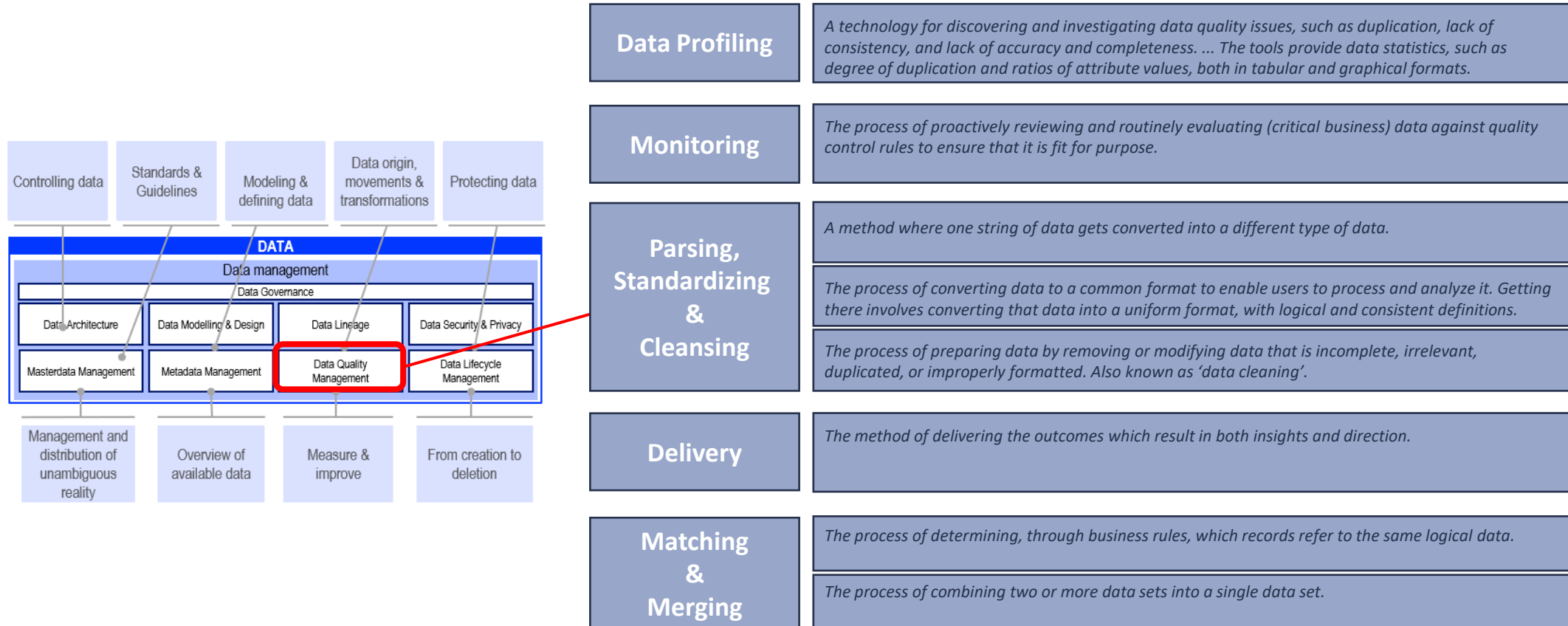
Voor héél PostNL én voor álle data

De doelarchitectuur Data Quality Management is voor héél PostNL én voor álle data.



Definitie, scope en positionering

“Capabilities” binnen Data Quality Management



Samenhang met andere Data Capabilities

Data Quality Management vs. ...

Data Capability	Relatie met Data Quality Management
Data Governance	• Heldere afspraken, kaders, richtlijnen, rollen en verantwoordelijkheden zijn nodig. En data eigenaren geven sturing aan datakwaliteit.
Data Architecture & Modelling	• Een solide data fundament maakt het overbodig vastleggen van objecten inzichtelijk en legt de onderlinge relaties bloot.
Data Lifecycle & - Lineage	• Inzicht in Data Lineage is noodzakelijk om oorzaken van Data Quality issues vast te kunnen stellen en deze op een effectieve en structurele manier te kunnen verhelpen.
Metadata Management	• Heldere en uniforme definities zorgen voor duidelijkheid en juist gebruik. Daarnaast is metadata noodzakelijk om te weten waar data aan moet voldoen om kwalitatief te zijn.
Data Warehousing & Business intelligence	• Sturen op betrouwbare informatie, het nemen van besluiten, het vaststellen van algoritmes en het uitvoeren van analyses zijn alleen mogelijk wanneer de kwaliteit van de data op orde is.



2. Strategie

Missie en visie van PostNL

Het belang van datakwaliteit

Visie op datakwaliteit

Stakeholders binnen PostNL

Missie en visie van PostNL

Wie willen we zijn en wat willen we bereiken?

Wie willen we zijn en wat willen we bereiken?

Hét logistieke e-commerce bedrijf van de Benelux

Jouw favoriete bezorger

Hoe

We houden het simpel. We maken het slim.
En we doen het samen

Oranje kompas



Strategische doelstellingen



Klanten helpen hun business te laten groeien



Leveren van toegankelijke en betrouwbare diensten



Vergroten van duurzame inzetbaarheid



Reduceren van de impact op het milieu



Zorgen voor een winstgevende groei

Digital versnellen en verankeren in ons DNA

Het belang van datakwaliteit

Grip op data is cruciaal...

Alleen zo...

- ...kunnen we goed aansluiten en anticiperen op de veranderende wensen van onze klanten
- ...kunnen we de marges op onze dienstverlening behouden en daar waar mogelijk zelfs vergroten
- ...sturen we op betrouwbare informatie en maken we de juiste besluiten
- ...wordt de digitale transformatie een succes



Financial performance



Operational performance



Customer experience

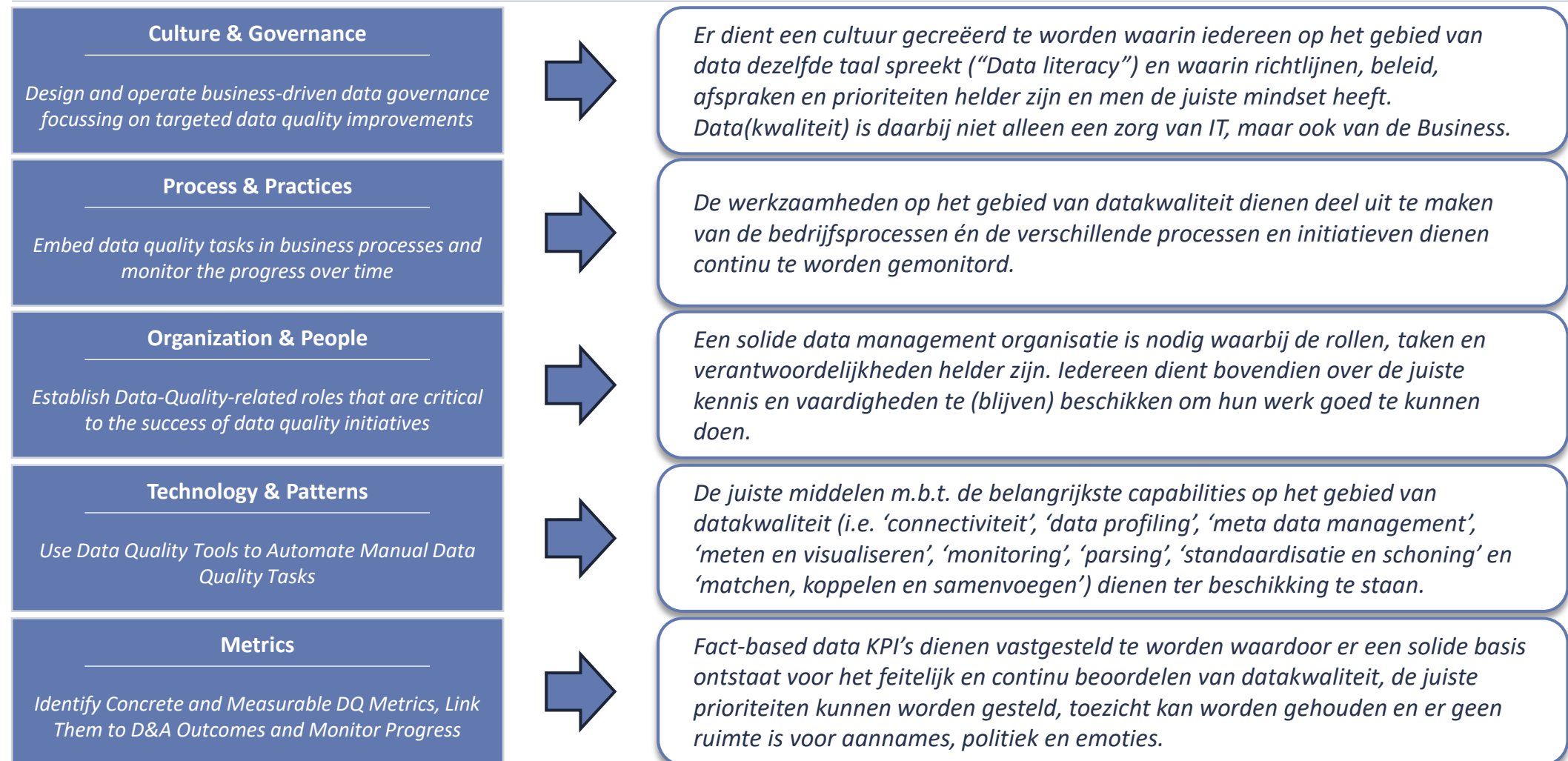


Legal &
regulatory compliance

Business drivers

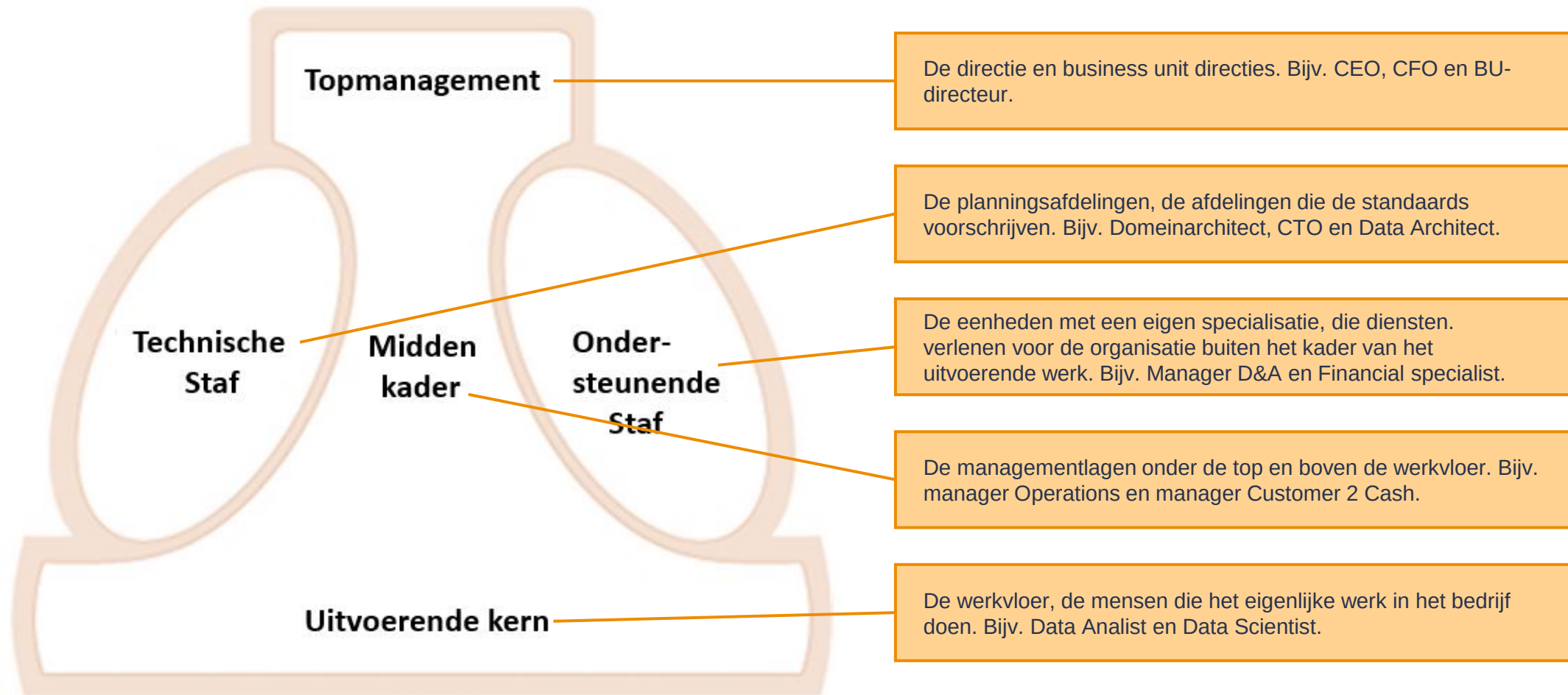
Visie op datakwaliteit

Hoe gaat dat dan in de praktijk? (*)



Stakeholders binnen PostNL

Wie spelen er allemaal een rol op het gebied van data(kwaliteit)?



3. PostNL Data Principles

Data Principles

Data Principles

PostNL breed (1/2)

Data is een asset

De gegevens die wij binnen PostNL verzamelen zijn erg waardevol en leveren ons belangrijke inzichten op. Daarom is data dus cruciaal voor het succes van onze organisatie. We gaan er dan ook zorgvuldig mee om zodat we het gebruik ervan kunnen optimaliseren en waarde kunnen toevoegen in termen van klant ervaringen (favoriete bezorger) en een efficiënte bedrijfsvoering.

Data essentials - 1 van 7



Dat betekent...

- Tooling is per data domein beschikbaar t.b.v. data profiling, data cleansing, het vastleggen van DQ beslisregels, DQ issuebeheer en het meten, monitoren en visualiseren van datakwaliteit
- Resource-allocatie is beschikbaar voor het trainen en opvoeden van collega's die data opvoeren
- Bij de praktische uitvoering van Data Quality Management wordt rekening gehouden met de richtlijnen zoals die in het Data Quality Framework staan
- Focus ligt op data die het meest kritisch is voor het realiseren van de strategische doelen

Data heeft een eigenaar

Elk gegeven dat binnen onze organisatie wordt opgevoerd, kent PostNL breed één specifieke eigenaar. Deze eigenaar is verantwoordelijk voor de kwaliteit van de gegevens en van de processen die nodig zijn om het gegeven op te voeren, te wijzigen en/of te verwijderen.

Data essentials - 2 van 7



Dat betekent...

- Elk business domein heeft een eigenaar die verantwoordelijk is voor de kwaliteit van de opgevoerde data
- Deze persoon heeft in zijn/haar planningsafspraken duidelijke KPI's op het gebied van datakwaliteit
- Voor elk domein zijn alle rollen, taken en verantwoordelijkheden, die PostNL binnen de Data Management Organisatie kent, vastgelegd en bekend binnen het eigen domein

Data wordt één keer vastgelegd

Voor elk gegeven dat wordt opgevoerd, is bepaald in welke centrale bron dé waarheid wordt vastgelegd. Systemen en applicaties hebben de verplichting dit gegeven uit deze bron af te nemen, zodat dit gegeven binnen de gehele PostNL organisatie hetzelfde is en iedereen er een en hetzelfde beeld van heeft.

Data essentials - 3 van 7



Dat betekent...

- Gegevens worden op één plek vastgelegd
- De opvoer van gegevens wordt zoveel mogelijk gestandaardiseerd o.a. m.b.v. referentietabellen en drill down menu's. Waar dit niet mogelijk is, worden duidelijke voorwaarden gesteld qua validiteit
- Voor elk gegeven is vastgesteld waar de opvoer plaatsvindt, waar dé waarheid wordt vastgelegd. (i.e. System of Record) en welke systemen er gebruik van maken
- Afnemende systemen maken voor masterdata uitsluitend gebruik van het System of Record
- Eventuele correcties op de data worden uitsluitend in de bron doorgevoerd
- Data opvoerprocessen (CRUD) worden vastgelegd in een procesbeschrijving conform het RACI-model

Data is toegankelijk

Iedereen binnen PostNL heeft toegang tot de gegevens die hij /zij nodig heeft voor het uitvoeren van zijn/haar functie. Hierbij wordt er ook rekening mee gehouden met de gevoeligheid van de gegevens.

Data essentials - 4 van 7



Dat betekent...

- Per data domein worden de resultaten van DQ rapportages gedeeld
- Uitkomsten van DQ rapportages worden opgenomen in een DQ-dashboard dat periodiek wordt voorgelegd aan de DGB / EC
- Per domein is vastgelegd over welke data, in welke vorm, in welke frequentie en aan wie wordt gerapporteerd, incl. de KPI's op basis waarvan wordt gerapporteerd

Data Principles

PostNL breed (2/2)

Data is onder beheer

Van alle gegevens die worden opgevoerd, wordt vastgelegd wat hiervan de definitie is en welke eisen er aan worden gesteld. Dit wordt ook op continue basis gemonitord en daar waar nodig bijgestuurd.

Data essentials - 5 van 7



Dat betekent...

- Voor alle (MD) objecten is de meta data (incl. definities) op een centrale en toegankelijke plek vastgelegd
- Data gerelateerde KPI's zijn per databron vastgesteld en hebben een duidelijke link naar de strategische doelen van PostNL
- Aan de opvoer van data worden hele duidelijke voorwaarden gesteld
- Iedereen die data opvoert, is op de hoogte van de eisen die hieraan worden gesteld
- DQ rapportages zijn geïmplementeerd en monitoren de kwaliteit door de keten heen
- Voor elk domein is een proces ingericht voor het identificeren en het uiteindelijk bij de bron oplossen van DQ issues
- Voor elke bron die wordt ontsloten op het Data Lake wordt een rapport opgesteld waarin de uitkomsten van de datakwaliteitsmeting staan. Deze wordt standaard meegestuurd naar de betreffende gebruikers op het Data Lake
- Focus ligt op preventie i.p.v. genezen

Data wordt beschermd

Alle gegevens worden veilig opgeslagen en beschermd tegen oneigenlijk gebruik, zoals ongeautoriseerde toegang, wijziging en publicatie. Hiermee voldoen we aan de Nederlandse (en Europese) wet- en regelgeving.

Data essentials - 6 van 7



Dat betekent...

- Voor elke applicatie/tool in de keten zijn de rollen en functionaliteiten vastgelegd
- Voor alle rollen zijn rechten en autorisatieniveaus vastgelegd m.b.t. het opvoeren, bekijken, gebruiken en verwijderen van data

Data is te analyseren

Alle gegevens binnen PostNL – zowel historisch als actueel – zijn beschikbaar voor analyse doeleinden. Hiermee kunnen o.a. dashboards, rapportages en algoritmes worden gemaakt, zodat we een betrouwbaar beeld krijgen van de werkelijkheid en in staat zijn voorspellingen te doen.

Data essentials - 7 van 7



Dat betekent...

- Data afkomstig uit het System of Record voldoet aan de vastgestelde eisen qua kwaliteit, kwantiteit en vorm. (toetsing vooraf)
- Voor elke applicatie zijn requirements vastgelegd voor het ontsluiten van data uit ontvangende systemen en wordt vooraf vastgesteld wat de datakwaliteit is
- Requirements zijn vastgelegd voor het al dan niet toelaten van een nieuwe databron
- De kwaliteit van een data domein wordt getoetst op volledigheid, uniciteit, tijdigheid, validiteit, nauwkeurigheid en consistentie
- Met behulp van rapportages wordt objectief en transparant getoetst of aan de gestelde voorwaarden wordt voldaan
- De betekenis van de (MD) objecten is voor iedereen duidelijk en wordt éénduidig geïnterpreteerd



4. Cultuur & Governance

“Culture & Governance”

The 7 Must Have Foundations

The Seven Must-Have Foundations

Bron : Gartner (*)

The Seven Must-Have Foundations for Modern Data and Analytics Governance



Source: Gartner
ID: 433904

Value and Outcomes

Governance focuses on business value and organizational outcomes, such that governance activities directly connected with achieving specific business priorities.

Accountability and Decision Rights

Accountability and decision rights are unequivocally assigned by those in the organization with the right authority, and are agreed, documented and routinely reviewed.

Trust

Governance is trust-based and acknowledges a distributed ecosystem of data and analytics where an understanding of lineage and curation is important.

Transparency and Ethics

The basis on which governance decisions are made is fully transparent, based on stated enterprise values and organizational ethics.

Risk and Security

Decisions are considered based on risk-adjusted opportunities in context of a dynamic security and risk-control environment.

Education and Training

An ongoing program of education, training and monitoring is in place that develops the right governance skills at the right time.

Collaboration and Culture

Better business results are derived when governance is based on a culture of enablement and collaboration rather than control.



5. Bedrijfsprocessen

“Process & Practices”

Procesonderdelen

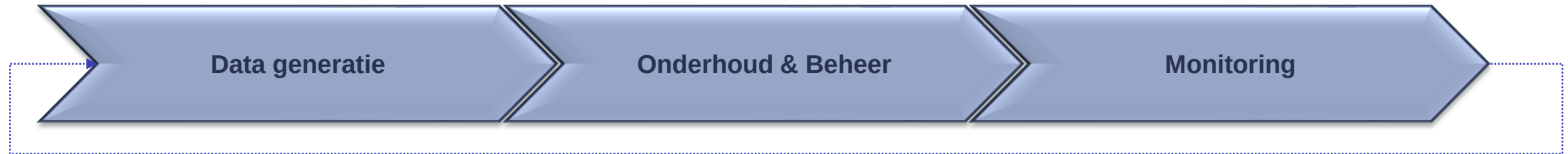
Doelsituatie

Huidige situatie

Procesonderdelen

Data generatie – Onderhoud & Beheer – Monitoring

De werkzaamheden op het gebied van datakwaliteit dienen deel uit te maken van de bedrijfsprocessen én de verschillende processen en initiatieven dienen continu te worden gemonitord.



Het daadwerkelijk (al dan niet geautomatiseerd) opvoeren en/of wijzigen van gegevens in een systeem of applicatie.

Bijv.

- Opvoeren van een nieuwe proceslocatie in BLS
- Wijzigen van een factuuradres in SAP
- Genereren van waarnemingscodes gedurende het verloop van het logistieke proces (events)
- Afsluiten van een opportunity in Salesforce

Alles dat nodig is om ervoor te zorgen dat het werk daadwerkelijk gedaan kan worden. Nu en in de toekomst.

Bijv.

- Trainen en opleiden van nieuwe medewerkers
- Interne communiceren over veranderingen in procedures en/of wijzigingen in systemen
- Zorgen dat alle rollen binnen de eigen BU ingevuld blijven

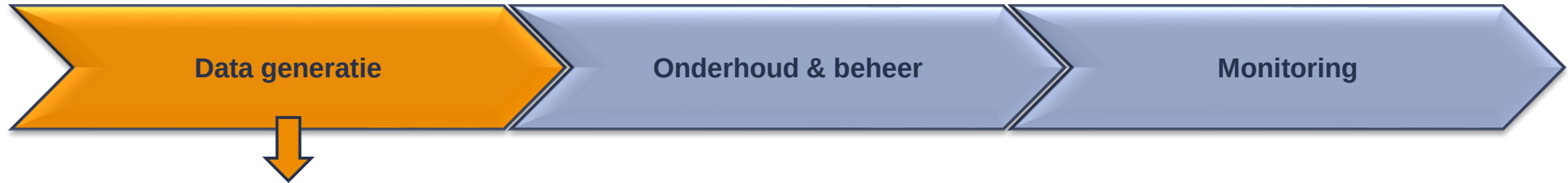
Het periodiek toetsen of de data nog steeds geschikt is voor het beoogde doel en of de opvoer op een juiste manier plaatsvindt.

Bijv.

- Project 'Case Classificatie' in Salesforce
- DQ rapportage op Data Lake of BLS
- DQ Dashboard DGB

Doelsituatie

Data generatie



- Het proces voor het opvoeren, wijzigen en verwijderen van masterdata, transactionele en conditionele data, incl. data eigenaarschap is vastgelegd .
- Vastgelegd is welke data objecten worden opgevoerd. Hierbij wordt kritisch gekeken naar de nut en noodzaak van het vastleggen van elk van de gegevens.
- De opvoer van gegevens is zoveel mogelijk gestandaardiseerd o.a. m.b.v. referentietabellen, drill down menu's en gebruik data vanuit System of Record. Waar dit niet mogelijk is, zijn duidelijke voorwaarden gesteld qua validiteit.
- Medewerkers zijn getraind in het opvoeren van gegevens.
- Voor alle gegevens die worden opgevoerd, is op een centrale en toegankelijke plek vastgelegd wat hiervan de definities zijn.
- Medewerkers die verantwoordelijk zijn voor de opvoer van gegevens hebben in hun planningsafspraken duidelijke KPI's op het gebied van datakwaliteit. Dit geldt ook voor de leidinggevendenden van deze medewerkers.

Doelsituatie

Onderhoud & Beheer



- Bij wijzigingen in de opvoersystemen worden medewerkers die belast zijn met de opvoer van gegevens hierover proactief geïnformeerd en daar waar nodig bijgespijkerd.
- Tooling is beschikbaar t.b.v. data profiling, data cleansing, het vastleggen van DQ beslisregels, DQ issuebeheer en het meten, monitoren en visualiseren van datakwaliteit.
- Er is een proces ingericht voor het identificeren en het uiteindelijk en uitsluitend bij de bron oplossen van DQ issues.
- Alle rollen, taken en verantwoordelijkheden, die PostNL binnen de Data Management Organisatie kent, zijn vastgelegd en bekend binnen het eigen domein.

Doelsituatie

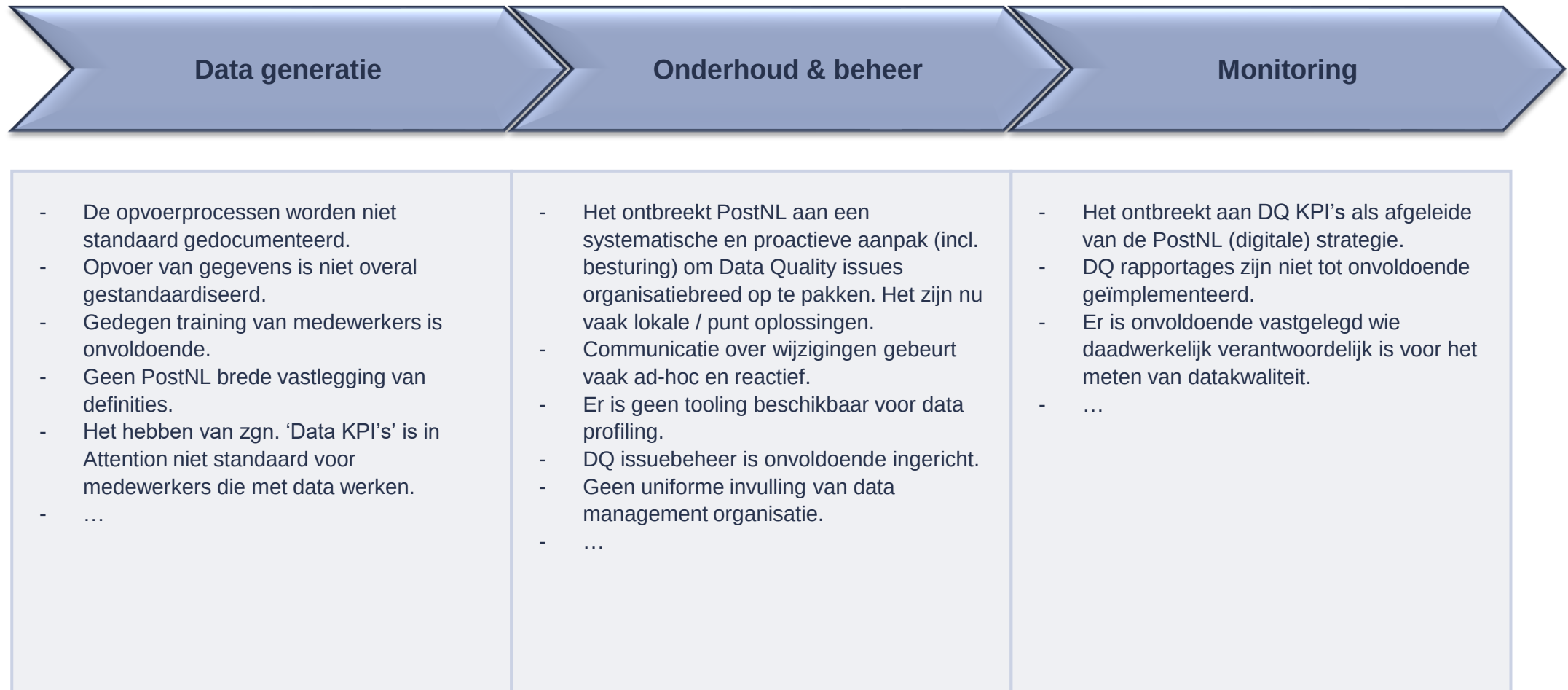
Monitoring



- Vastgelegd is over welke data, in welke vorm, in welke frequentie en aan wie wordt gerapporteerd, incl. de KPI's op basis waarvan wordt gerapporteerd.
- DQ rapportages zijn geïmplementeerd waarmee wordt getoetst of aan de gestelde voorwaarden wordt voldaan.
- Vastgesteld is wie verantwoordelijk is voor het daadwerkelijk meten van de datakwaliteit.
- Uitkomsten van DQ rapportages worden gedeeld.

Huidige situatie

Wat is nog niet op orde?





6. Besturing & Organisatie

“Organization & People”

Federale model

Rolverdeling m.b.t. Data Quality Management

Federale model

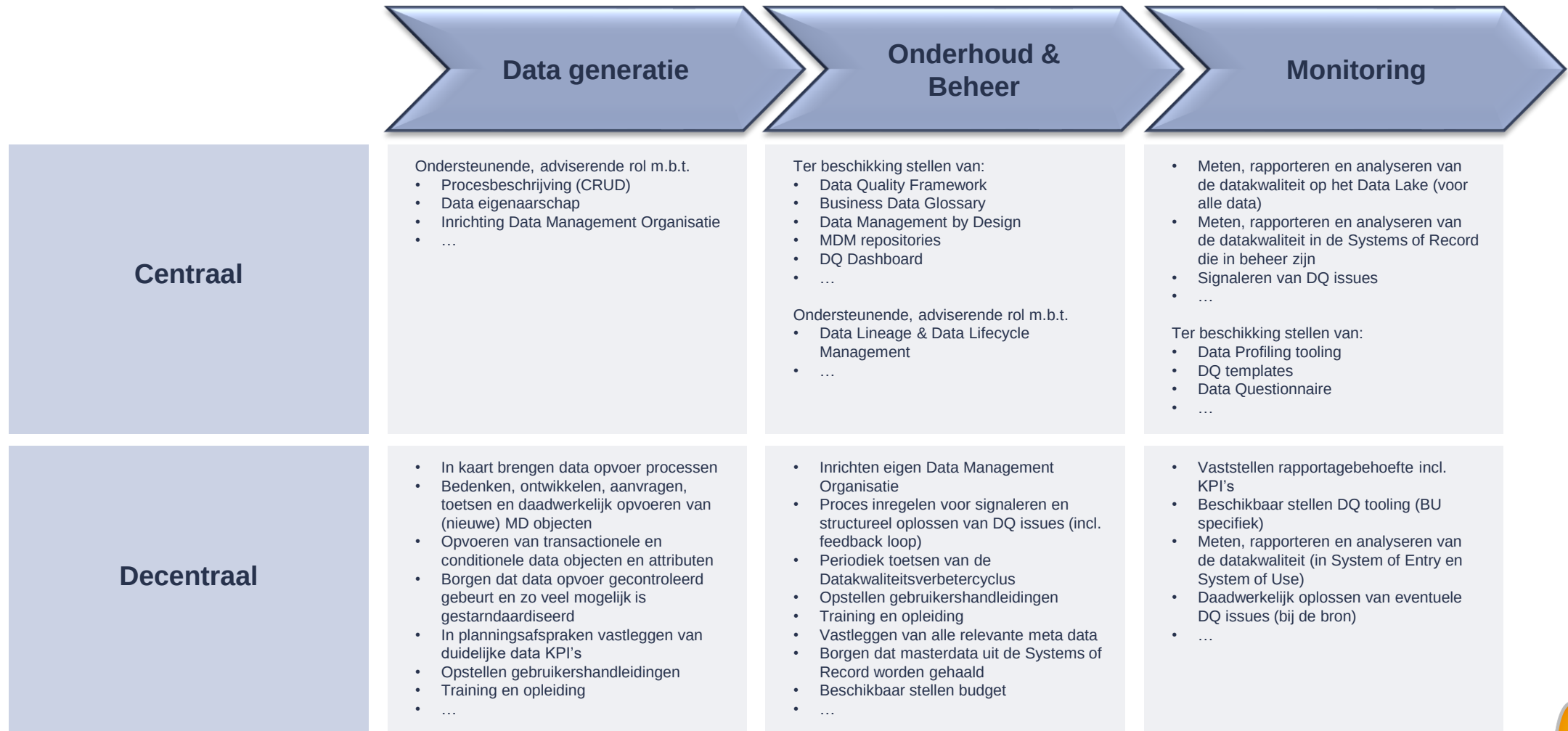
Rollen binnen elk Business domein

Een solide data management organisatie is nodig waarbij de rollen, taken en verantwoordelijkheden helder zijn. Iedereen dient bovendien over de juiste kennis en vaardigheden te (blijven) beschikken om het werk goed te kunnen doen.

Strategisch Data Eigenaar	Tactisch Data Eigenaar	Data Manager	Platform Owner	Data Steward
Is accountable en zorgt ervoor dat de noodzakelijke mensen en middelen ter beschikking worden gesteld om de kwaliteit van de aan hem/haar toegewezen objecten te borgen.	Verantwoordelijk voor de data binnen een specifiek data domein. Hij/zij zorgt ervoor dat de data binnen zijn/haar domein correct wordt beheerd over verschillende systemen en processen (bedrijfsactiviteiten) en dat, door het inrichten van processen en systemen, de kwaliteit van de data optimaal is.	Verantwoordelijk voor een doelmatige en effectieve organisatie van het data management proces binnen zijn/haar eigen domein. Hij/zij ziet toe op de kwaliteit van de beheerde data, implementeert en waarborgt de data gerelateerde KPI's en SLA's en stelt het datamodel op.	Zorgt er voor dat de door de data manager voorgestelde oplossingen daadwerkelijk in de systemen, applicaties en/of platforms worden geïmplementeerd.	Heeft inhoudelijke kennis van de data objecten en de gerelateerde bedrijfsprocessen en is verantwoordelijk voor de controle van de kwaliteit van de data, achteraf.

Rolverdeling

Centraal vs. decentraal





7. Applicatie Architectuur

“Technology & Patterns”

Onderbouwing

Architectuur building blocks

Solution building blocks [volgt later]

Onderbouwing

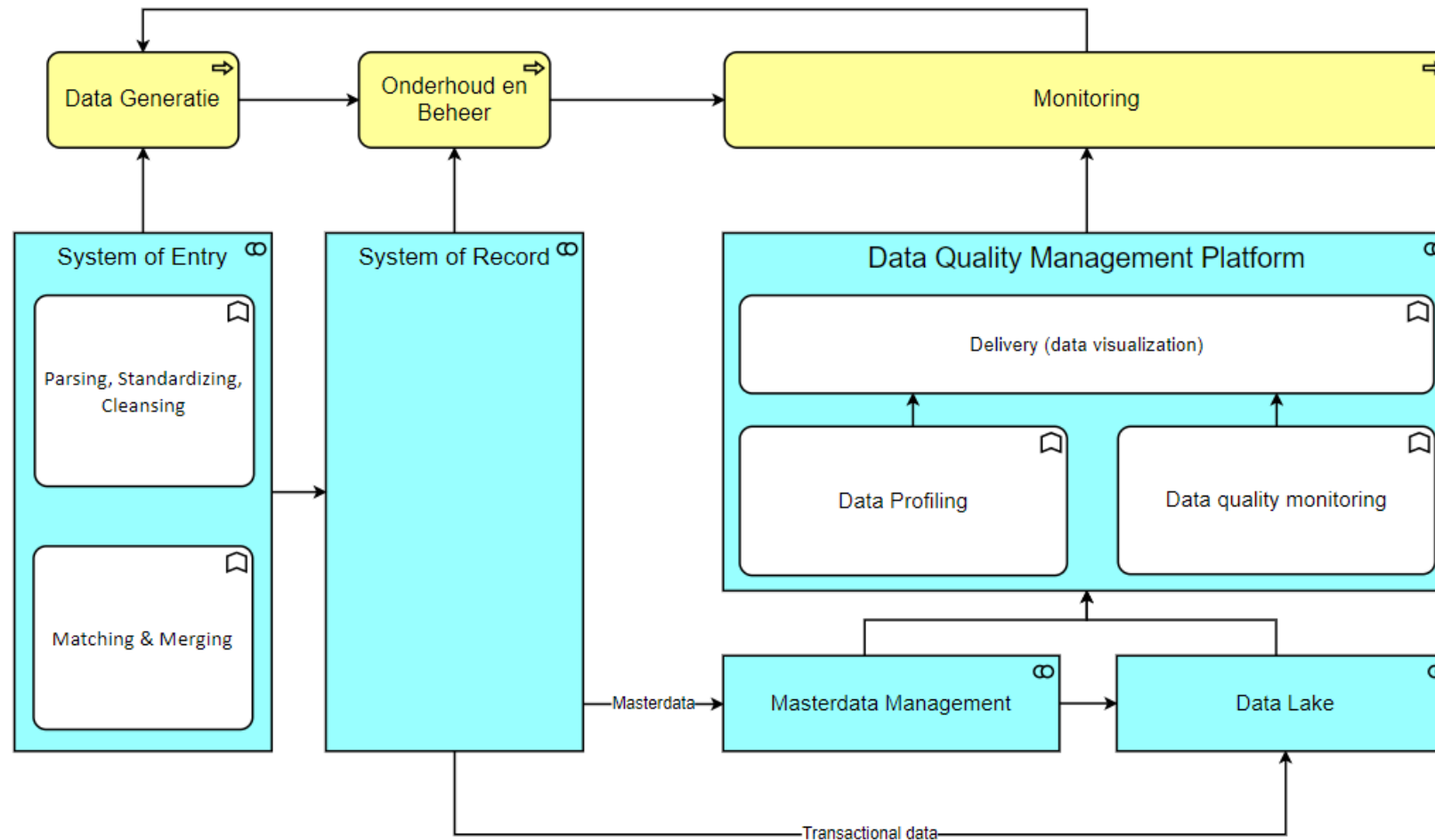
Focuspunten korte termijn

De juiste middelen m.b.t. de belangrijkste capabilities op het gebied van datakwaliteit (i.e. 'data profiling', 'monitoring', 'parsing, standaardisatie en schoning', 'delivery' en 'matching & merging') dienen ter beschikking te staan.

- Vanuit de business zijn 'Business Partner', 'Consument' en 'Locatie' benoemd als de belangrijkste masterdata objecten waar op korte termijn de focus ligt qua datakwaliteit.
Waar mogelijk worden de onderliggende transactionele en conditionele objecten en attributen meegenomen.
- Datakwaliteit wordt idealiter bij de bron (i.e. System of Entry) geadresseerd en aangepakt.
- Echter, het is ondoenlijk om op korte termijn de verschillende DQ capabilities los te laten op het grote aantal Systems of Entry binnen PostNL.
- Daarom focus op data profiling, monitoring en delivery op die platformen waar dat op korte termijn wel mogelijk is; Data Lake en EBX Platform.
- DQ issues worden hier gesignaleerd en geadresseerd. Middels een nog in te richten proces dient een close-loop voor het daadwerkelijk bij de bron oplossen van het probleem te zorgen.

Architectuur Building Blocks

Concept doelarchitectuur



8. Metrics

Value Pyramid [nog uit te werken]

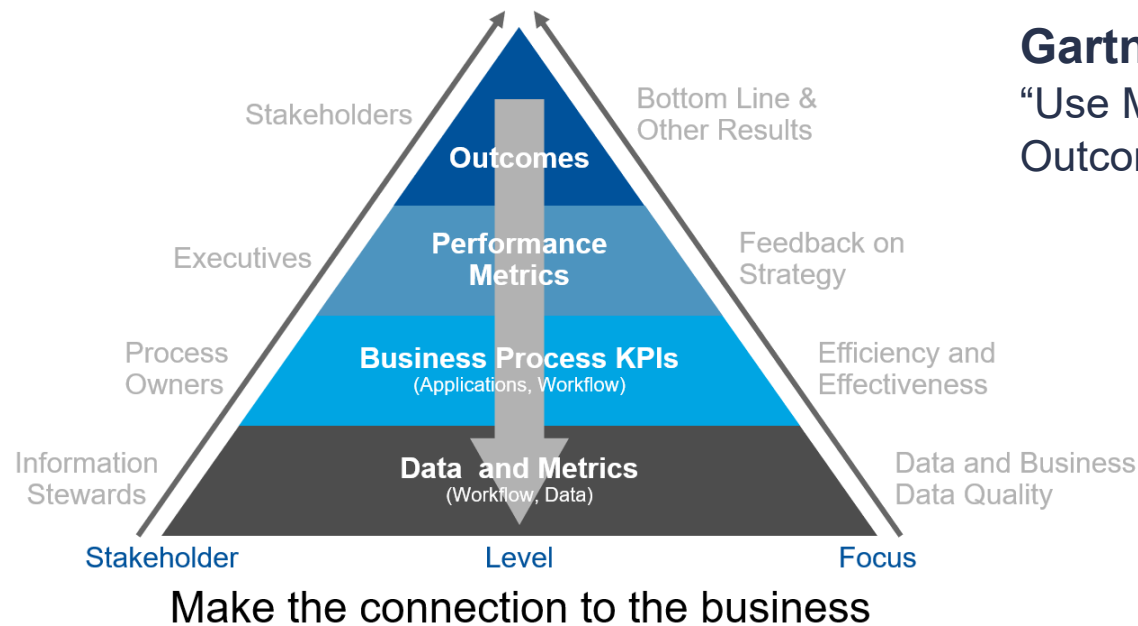
DQ Dashboard [nog uit te werken]



Value Pyramid

Fact-based data KPI's dienen vastgesteld te worden waardoor er een solide basis ontstaat voor het feitelijk en continu beoordelen van datakwaliteit, de juiste prioriteiten kunnen worden gesteld, toezicht kan worden gehouden en er geen ruimte is voor aannames, politiek en emoties.

[Input volgt uit 'DQ Dashboard' traject met Marcel en Bart.]



Gartner's Value Pyramid

"Use Metrics Framework to Explore the Link Between Outcome and Data "





9. Roadmap

Focus 2021

Tijdslijnen

Focus 2021

Belangrijkste aandachtsgebieden

- DQ Dashboard
 - *Vaststellen Data specifieke KPI's & targets*
 - *Actualiseren en formaliseren data eigenaarschap (incl. DMO)*
 - *Inrichten DQ rapportages en DQ issue management*
- DQ Toolselectie
 - *Data Profiling op Data Lake en MDM Platform*
 - *Monitoring en Delivery*
- Data Cataloging
 - *Vastleggen meta data t.b.v. bronnen Data Lake*

10. “Best practices”

...

...



Best practices

“Ondersteunde hulpmiddelen / producten”

Opsomming van de ‘producten’ die beschikbaar zijn om in de praktijk direct met DQ aan de slag te gaan. Bijv.

- Data Quality Framework
- Data Management by Design
- ...
- ...



Concrete vragen aan leden Architectuurboard

Van een visie document naar een doelarchitectuur

Te beantwoorde vragen...

- Is de visie op Data Quality Management helder?
- Zijn er aanvullende adviezen voor dit visie document?
- Zijn de benoemde focuspunten (korte termijn) duidelijk en herkenbaar?
- Wat zijn de next steps om tot een Doelarchitectuur voor Data Quality Management te komen?
- ...