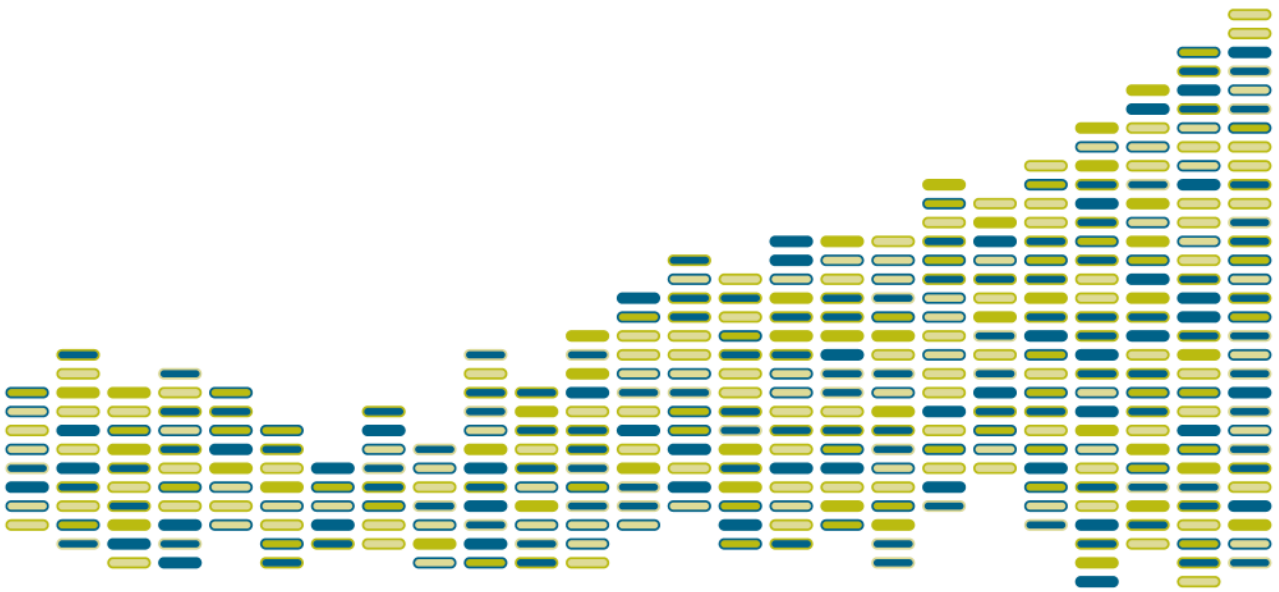


# Deltiq

# Data Management Framework

*Een praktisch hulpmiddel bij het realiseren van Data Management*



Deltiq Group bv  
Breukelen, augustus 2024

[www.deltiq.com](http://www.deltiq.com)

# Over Deltiq

Voor ons is data een bedrijfsmiddel. Onze consultants richten zich op het verbeteren van het managen van data – het vormgeven van Data Management. Op die manier kan data succesvol worden ingezet in bestaande of nieuwe bedrijfsprocessen. Daarmee kunnen overheid en bedrijfsleven echt datagedreven worden en analytics of big-data succesvol inzetten. Deltiq ondersteunt én bij visie- en strategievorming én bij de daadwerkelijke realisatie. Van ontwerp tot bouw en beheer. Van Data Governance tot het inrichten van de data-organisatie. Met integraal aandacht voor alle relevante disciplines om tot succes te komen: mensen, processen, organisatie en techniek. Onze consultants opereren daarbij op elk relevant niveau: van boardroom tot de vloer.

Deltiq beschikt over uitgebreide ervaring in het oplossen van Data Management-vraagstukken. Deze ervaring is opgedaan in complexe en grote organisaties zoals Ahold, de Belastingdienst, de Nationale Politie en de gemeente Amsterdam. Andere klanten zijn het Ministerie van Justitie en Veiligheid, DNB, AFM, ANWB, Heineken, Vomar, COA, bol.com, Albron en de Sint Maartenskliniek.

Deltiq Group bv  
Oud Aa 4b  
3621 LB Breukelen  
Tel 0346 21 67 09  
info@deltiq.com

[www.deltiq.com](http://www.deltiq.com)

© Deltiq Group bv, Breukelen

Niets uit deze uitgave mag zonder voorafgaande schriftelijke toestemming van de uitgever worden openbaar gemaakt of veeleenvoudigd, waaronder begrepen het reproduceren door middel van druk, offset, fotokopie of microfilm of in enige digitale, elektronische, optische of andere vorm of (en dit geldt zondig in aanvulling op het auteursrecht) het reproduceren (i) ten behoeve van een onderneming, organisatie of instelling of (ii) voor eigen oefening, studie of gebruik welk(e) niet strikt privé van aard is.



# Inhoudsopgave

Over Deltiq	2
Inleiding	5
Opzet	7
Doel	7
Doelgroep	7
Referentie	7
Rationale	7
Toepassingsmogelijkheden	8
Praktisch gebruik	9
Uitgangspunten	12
Bedrijfsfunctie of programma	12
Data Governance vs. Data Management	12
HR- en Finance-analogie	12
Follow the data	13
Ontbrekende disciplines	14
Data lifecycle management	14
Indeling	15
Disciplines	15
Aspecten	15
Dataomgeving of Bedrijfsonderdeel	15
Samenhang	16
Toelichting per discipline	17
Data Governance	17
Data Architecture	19
Metadata	21
Data Quality	22
Data Security, Ethics, Privacy en Compliance	23
Data Technology	27



Data Modeling	27
Data Development	29
Data Operations	31



# Inleiding

## *Waarom een framework voor Data Management?*

Van oudsher zijn we gewend te denken in systemen. In sommige organisaties gaat dat zelfs zo ver dat er bijna sprake is van “elk probleem haar eigen systeem”. Inmiddels groeit het besef dat het verstandiger is data en functionaliteit te scheiden. Immers, data heeft over het algemeen een veel duurzamer karakter en is voor meerdere doeleinden inzetbaar, terwijl functionaliteit onder druk van veranderende inzichten frequenter aan wijzigingen onderhevig is en hoogstens resulteert in meer, nieuwe of aanvullende data.

Daarnaast zijn we gaan beseffen dat data niet zomaar een bijproduct is van het door automatisering efficiënter maken van processen maar op zichzelf waarde kan toevoegen. Welke organisatie heeft tegenwoordig datagedreven worden of informatiegestuurd werken niet in haar strategie staan?

Toch vallen de resultaten vaak tegen. De investeringen leveren niet op wat ervan verwacht werd. Initiële resultaten met slimme algoritmes blijken lastig te implementeren en op te schalen in de reguliere operationele processen. Het periodiek beschikbaar krijgen van data met de juiste kwaliteit blijft vaak een uitdaging gegeven de beperkingen van de onderliggende legacy-systemen. De data in de nieuwe cloud-oplossing is uiteindelijk toch heel wat lastiger te ontsluiten voor nieuwe toepassingen dan initieel gedacht. Data scientists komen nauwelijks aan hun eigenlijk werk toe – het toepassen van artificial intelligence (AI) - omdat ze het overgrote deel van hun tijd kwijt zijn aan het beschikbaar krijgen van data....

Wat opvalt is dat de beschikbaarheid en betrouwbaarheid van data als vanzelfsprekend wordt beschouwd, terwijl het dat bijna nooit is, zelfs als de belangen heel hoog zijn zoals in de COVID-19 crisis<sup>1</sup>. Wat deze situaties gemeen hebben is het ontbreken van een expliciet stuur op de data. De data wordt niet *gemanaged* – effectief Data Management ontbreekt. Niet in termen van technologie maar als expliciet ingerichte functie binnen de organisatie. Als we vinden dat data een bedrijfsmiddel is, waarom organiseren we dan data niet net zoals we dat doen met de overige bedrijfsmiddelen? Een functie waarbinnen alle relevante aspecten van Data Management aan de orde komen – niet alleen techniek maar ook de noodzakelijk organisatie, de processen en besturing? Alle noodzakelijke randvoorwaarden om data structureel op een effectieve manier in te kunnen zetten?

Voorlopers in het gebruik van data als strategisch bedrijfsmiddel hebben zich inmiddels gerealiseerd dat er werk aan de winkel is bij het inrichten van Data Management. Maar welke

<sup>1</sup> [The Covid-19 pandemic has catapulted one mysterious data website to prominence, sowing confusion in international rankings, CNN](#)

hulpmiddelen zijn hiervoor beschikbaar, hoe wordt voorkomen dat iedereen het wiel gaat uitvinden? Hiertoe grijpen organisaties vaak naar bekende referentiekaders als de DAMA Data Management Body of Knowledge (DAMA DMBOK) of het Data Management Maturity Model (DMM). Hoewel waardevol als referentiekader is onze ervaring dat een pragmatische insteek in deze raamwerken vaak ontbreekt. De omvang alleen al van DAMA DMBOK (600 pagina's!) schrikt velen af.

Vanuit de in de praktijk ervaren behoefte aan een meer werkbare variant is dit Deltiq Data Management Framework ontstaan. Een gestructureerd en integraal overzicht van alle relevante aspecten die tezamen Data Management vormgeven: grotendeels bekend vanuit bestaande raamwerken, aangevuld met onze ervaring en op een overzichtelijke manier gepresenteerd.

# Opzet

## Doel

Het Deltiq Data Management Framework is bedoeld om op een toegankelijke en laagdrempelige manier overzicht te krijgen van de inrichting van Data Management in een organisatie. Het kan gebruikt worden om zowel de bestaande situatie in kaart te brengen als de gewenste situatie te schetsen. Het belangrijkste doel is dat het dient als praatplaat: het helpt om met verschillende betrokkenen in gesprek te raken. Zie het dus vooral niet als een keurslijf waarbij alle vakjes moeten worden afgevinkt!

## Doelgroep

Het Framework is bedoeld voor iedereen die zich bezighoudt met vraagstukken rondom Data Management en Data Governance. Steeds vaker nemen organisaties het initiatief om Data Management te gaan organiseren. Er wordt dan een Chief Data Officer (CDO), Data Governance manager, Datamanager of vergelijkbare functionaris benoemd om hier richting en inhoud aan te geven. Voor hen biedt het Framework een bruikbaar vertrekpunt. Maar ook voor iedere andere medewerker die wil weten hoe het zit, biedt het Framework houvast.

## Referentie

Als referentie is zowel DAMA DMBOK<sup>2</sup> als het Data Management Maturity Model<sup>3</sup> (DMM) gebruikt. Daarnaast komt de nodige inbreng van onze eigen praktijkervaring. Het Framework wijkt op bepaalde punten af van DAMA of DMM. Dit is bewust gedaan. Op deze manier is het Framework meer praktisch toepasbaar dan wanneer de disciplines uit de verschillende referentiekaders letterlijk zouden zijn overgenomen. Zo komen binnen verschillende DAMA-disciplines soms dezelfde activiteiten voor. In de praktijk leidt dat nogal eens tot verwarring - om niet te spreken van onmogelijke situaties wanneer een organisatie probeert DAMA 1:1 te implementeren.

## Rationale

Bij verschillende opdrachten waarbij we klanten ondersteunen bij de inrichting van Data Management liepen we steeds tegen dezelfde vraagstukken aan: hoe krijgen we snel overzicht van wat er al is gerealiseerd en wat nog niet? Aan de hand van welke visualisatie kunnen we over Data Management communiceren? Hoe brengen we de samenhang in beeld tussen verschillende implementaties of data-omgevingen? Hoe maken we inzichtelijk wat er nog moet gebeuren of wat we bewust niet doen? Wat we merkten was dat de beschikbare raamwerken onvoldoende houvast boden om deze vraagstukken in de praktijk doelmatig te beantwoorden. Te omvangrijk ook in het streven naar volledigheid. Zo is het idee van het Framework ontstaan.

<sup>2</sup> [DAMA/DMBOK2](#)

<sup>3</sup> [CMMI Institute, Data Management Maturity \(DMM\)<sup>SM</sup>](#)

Beoogd wordt om het vooral in de praktijk werkbaar te laten zijn. Zo beschouwd kan wellicht gesproken worden over 'DAMA Light' of een 'Data Management Quick Reference'.

## Toepassingsmogelijkheden

Het Framework kan op meerdere manieren worden gebruikt. In ieder geval zijn de volgende toepassingen te onderscheiden:

1. Als **checklist**: per discipline en activiteit kan worden aangegeven of dit is ingericht of niet. Dit kan worden verfijnd door dit per dataomgeving of bedrijfsonderdeel in te vullen.
2. Voor het beschrijven van de **inrichting**: op iedere cel (kruispunt van een aspect en dataomgeving) kun je aangeven hoe dit voor jouw organisatie is ingericht. Hoe is het georganiseerd, wie is verantwoordelijk voor dit onderdeel? Bestaat er documentatie van? Zijn er tools of oplossingen voor ontwikkeld? Een eenvoudige verwijzing, titel of steekwoord kan volstaan om dit aan te geven. Dit is met name relevant naarmate de ontwikkeling van Data Management langer duurt. Nieuwe teamleden kunnen op deze manier snel inzicht krijgen in de stand van zaken en in eerder gemaakte keuzes. Zo wordt voorkomen dat eerder resultaat uit beeld raakt. Dit is een realiteit waar in de praktijk vaak sprake van is. Het inrichten of herinrichten van Data Management is een traject dat zich doorgaans uitstrekt over meerdere jaren waarbij stapsgewijs resultaten worden opgeleverd. Door deze resultaten in het Framework bij te houden, blijft overzicht behouden.
3. Om een **basale volwassenheidsmeting** uit te voeren en vast te leggen. Zeker wanneer je stap (2) hebt uitgevoerd kun je per cel met een score aangeven wat de kwaliteit van de inrichting is. Er zijn natuurlijk uitgebreide methodieken om een Data Management maturity scan te doen (zoals DMM) maar die vergen veel tijd en inspanning. Het Framework is meer laagdrempelig om bijvoorbeeld met een kleine groep belanghebbenden tot een eerste snelle beoordeling te komen.
4. Als hulpmiddel bij de **prioritering** van werkzaamheden. Het is zinloos te denken dat bij de realisatie van Data Management alles tegelijk zou kunnen, zelfs al waren de middelen onbeperkt. Het Framework ondersteunt de discussie over prioriteitsvolgorde. Het helpt om bewust een keuze te maken over wat nu en wat later opgepakt kan gaan worden. Daarbij kan het gebruikt worden bij periodieke budgetaanvragen.

Wanneer het Framework initieel is ingevuld kunnen allerlei interessante inzichten ontstaan, zoals:

1. **Overlap** in inrichting: bijvoorbeeld wanneer er sprake is van meerdere omgevingen en iedere omgeving zijn eigen technologie, werkwijze of standaarden heeft.
2. **Fit / Gap**: wanneer het Framework initieel is ingevuld ontstaat snel een beeld waar er gaten zijn. Welke disciplines en/of activiteiten zijn gewenst maar kennen bijvoorbeeld nog geen inrichting.

3. **Inconsistenties:** de disciplines en aspecten hangen met elkaar samen. Door de concrete inrichting in het Framework te beschrijven kunnen inconsistenties worden gesignaleerd. Bijvoorbeeld de situatie dat er weliswaar datamodellen worden gemaakt maar deze verder niet worden gebruikt bij het ontwikkelen van gegevensvoortbrengingen.

## Praktisch gebruik

Om het Framework op een praktische manier te kunnen gebruiken wordt het op twee manieren beschikbaar gesteld:

1. Als Excel-sheet: deze versie kan naar eigen inzicht worden aangepast. Vanzelfsprekend kunnen vanuit deze versie verwijzingen worden opnemen naar documentatie, Intranet, applicaties en dergelijke.
2. Als Poster: de poster is geoptimaliseerd om op A0 uit te printen. Zo kan het Framework bijvoorbeeld gebruikt worden in een workshop om het met betrokkenen in te vullen (lekker geeltjes plakken!). Een andere toepassing is om de poster in de project- of werkrimte op te hangen. Iedereen kan er dan op schrijven, aanvullen of tekenen. Zo ontstaat een levendig gebruik van het Framework.



# Deltiq

## Data Management Framework

Het Deltiq Data Management Framework brengt in een beknopte vorm alle relevante onderdelen van de bedrijfsfunctie Data Management samen. Het dient als hulpmiddel om zowel de bestaande als de gewenste situatie in een organisatie in beeld te brengen en verder vorm te geven. Een toelichting is beschikbaar op [deltiq.com/DDF](http://deltiq.com/DDF).

		Dataomgeving / Divisie / Bedrijfsonderdeel			
		1	2	3	
Discipline	Aspect	selecteer een typing			
		<input type="radio"/> Kwadrant I / II	<input type="radio"/> Managed	<input type="radio"/> Kwadrant I / II	
		<input type="radio"/> Kwadrant III / IV	<input type="radio"/> Unmanaged	<input type="radio"/> Kwadrant III / IV	
		<input type="radio"/> Managed	<input type="radio"/> Unmanaged	<input type="radio"/> Managed	
		<input type="radio"/> Unmanaged	<input type="radio"/> Managed	<input type="radio"/> Unmanaged	
I. OVERSIGHT	<b>1</b> Data Governance	<ul style="list-style-type: none"> <li>Strategie en beleid</li> <li>Organisatie, functies en verantwoordelijkheden</li> <li>Roadmap en planning</li> <li>Opleidingen en communicatie</li> <li>Managementinformatie, finance en control</li> <li>Risicomanagement en business case</li> <li>Data-administratie</li> <li>Data-issuemanagement</li> <li>...</li> </ul>			
	<p>Onder deze discipline vallen alle werkzaamheden die tot doel hebben concrete gegevensverwerkingen te realiseren. De aard en invulling van deze werkzaamheden kunnen verschillen per dataomgeving.            *...do not confuse the management of data with ensuring data is managed*</p>				
II. FOUNDATIONAL	<b>2</b> Data Architecture	<ul style="list-style-type: none"> <li>Beleid en spelregels</li> <li>Data(model) architectuur</li> <li>Gegevensbeheersysteem architectuur</li> <li>Datalogistieke architectuur</li> <li>Enterprise Alignment</li> <li>Master- en referentiedata aanpak</li> <li>...</li> </ul>			
	<p>Data Architecture is de discipline die inhoudelijk sturing geeft aan de data zelf en de verwerking ervan.            Daar waar de governance-discipline de overige disciplines inricht en aanstuurt, doet Data Architecture dat voor de data.</p>				
II. FOUNDATIONAL	<b>3</b> Metadata	<ul style="list-style-type: none"> <li>Beleid en spelregels</li> <li>Begrippenwoordenboek</li> <li>Datadefinities</li> <li>Horizontale en verticale lineage</li> <li>Publicatie en collaboratie</li> <li>Technologie</li> <li>...</li> </ul>			
	<p>Metadata(management) is een breed onderwerp en er zijn veel soorten metadata. De invulling is hier beperkt tot wat 'business metadata' wordt genoemd.            Hieronder vallen verschillende vormen van definities die beschrijven wat de data betekent.</p>				
II. FOUNDATIONAL	<b>4</b> Data Quality	<ul style="list-style-type: none"> <li>Beleid en spelregels</li> <li>Datakwaliteitsproces</li> <li>Data Profiling</li> <li>Meting en terugkoppeling</li> <li>Oorzaakanalyse en verbetering</li> <li>Technologie</li> <li>...</li> </ul>			
	<p>Datakwaliteit moet door de hele lifecycle gemanaged worden. Kwaliteitsmanagement is daarom ook een integraal onderdeel van andere disciplines en aspecten.            De uitwerking hier is beperkt tot de aspecten die buiten het reguliere lifecycle management van data vallen.</p>				



Versie 1.0/2020 © Deltiq Group bv, Breukelen www.deltiq.com

II. FOUNDATIONAL

**5**  
Data Security, Ethics, Privacy en Compliance

- Beleid en spelregels
- Issuemanagement
- Meting, monitoring en control
- Data-autorisatie
- Dataclassificatie
- Dataretentie
- Data Masking / encryptie / pseudo- en anonimisering
- Identificatie van persoonsgegevens
- Verwerkingsbeoordeling / Privacy Impact Assessment (PIA)
- Verwerkingsregister
- Verwerkersovereenkomst
- ...

Analoog aan Data Quality moeten security, ethics, privacy en compliance – door de hele lifecycle heen gemanaged worden. Deze discipline brengt dus ook elken met zich mee voor andere disciplines. Daarnaast zijn er eigenschappen die expliciete aandacht behoeven, ofwel omdat ze buiten de lifecycle van data om gemanaged moeten worden, ofwel omdat ze zo belangrijk worden gevonden dat aparte vermelding op zijn plaats is.

**6**  
Data Technology

- Beleid en spelregels
- Technologie
- Opslag
- Data-Integratie
- Analytics en data science
- Dashboarding, reportage en visualisatie
- ...
- Voortbrenging en versiebeheer
- Infrastructuur
- ...

Dit bevat alle (basis)technologieën die nodig zijn om gegevens te verwerken en te gebruiken. Bepaalde specifieke technologieën komen terug bij andere disciplines. In het kader van deze discipline is het nodig keuzes te maken over welke technologieën voor welk doel in welke dataomgeving gebruikt gaan worden.

**7**  
Data Modeling

- Modelleerwijze en -standaarden
- Modelleertools en schematechnieken
- Voortbrenging en versiebeheer
- Conceptuele modellering
- Logische modellering
- Fysieke modellering
- ...

Voor gedegen data management is data modeling onontbeerlijk. Modellen zijn de placeholder voor aanvullende eisen met betrekking tot kwaliteit, beveiliging, privacy en compliance. Elementaire keuzes ten aanzien van modellering worden hier verder ingevuld.

III. LIFECYCLE

**8**  
Data Development

- Ontwikkelstandaarden en -patronen
- Data-requirements-analyse
- GLO-management
- Voortbrenging, testen en versiebeheer
- Data Services
- Datalogistiek
- Data in OTAP
- Datamigratie
- ...

Onder deze discipline vallen alle werkzaamheden die tot doel hebben concrete gegevensverwerkingen te realiseren. De aard en invulling van deze werkzaamheden kunnen verschillen per dataomgeving.

**9**  
Data Operations

- Gebruikersondersteuning
- Incident & Problemmanagement
- Service Delivery Management
- Data Processing Management
- Data-archivering
- Beheer datakwaliteit
- Beheer begrippen en datadefinities
- Beheer master- en referentiedata
- ...

Deze discipline representeert het operationeel gebruik van data en de ondersteuning die daarbij nodig is. Ook voor gegevenstoepassingen, of dit nu gegevensbeheersystemen zijn of analytische toepassingen, is beheer nodig. Afhankelijk van de context blijft de inrichting van beheer soms onderbelicht.



# Uitgangspunten

Bij het opstellen van het Framework zijn een aantal uitgangspunten gehanteerd. Kennis van deze uitgangspunten helpt om het Framework te begrijpen en toe te passen.

## Bedrijfsfunctie of programma

Data Management is een bedrijfsfunctie. Dat impliceert dat Data Management een permanente positie heeft in de organisatie. Het is dus niet iets wat tijdelijk bestaat in de vorm van een project of programma om daarna weer te verdwijnen. Bestaande literatuur en methodologieën hanteren echter vaak zo'n programmatische invalshoek: ze zijn aanpak gericht om Data Management te 'doen'. Het Framework gaat echter eerder uit van een permanente inrichting: wanneer Data Management is ingericht in mijn organisatie, wat zou er dan allemaal moeten zijn? Op z'n minst helpt het Framework om die vraag te beantwoorden, om daarna misschien via een programmatische aanpak Data Management te realiseren....

## Data Governance vs. Data Management

Over de betekenis van deze termen en het onderscheid worden de meest uiteenlopende interpretaties en beschouwingen gehanteerd. We volstaan hier met aan te geven dat we de benadering van DAMA volgen, die Data Governance als centrale discipline positioneert, sturend op de andere Data Management disciplines. Of, zoals John Ladley<sup>4</sup> het mooi verwoordt:

“...do not confuse the management of data with ensuring data is managed.”

## HR- en Finance-analogie

Wanneer Data Management als bedrijfsfunctie wordt gezien, kan de vergelijking gemaakt worden met andere bedrijfsfuncties, zoals HR en Finance. Dat is een bruikbare analogie die echter wel enige toelichting verdient.

Het onderscheid tussen Data Management en Data Governance, is ook zichtbaar bij de HR- en Finance-functie. Een centrale HR- of Finance-functie zorgt voor de kaders, richtlijnen en werkwijzen voor het 'doen' van HR en Finance. In het geval van HR is dat bijvoorbeeld een functiebouwwerk of salarishuis dat voor een organisatie is ingericht; in het geval van Finance bijvoorbeeld alle afspraken en regels die gelden voor de boekhouding en vastgelegd zijn in een accounting manual. Het beheer van dergelijk beleid en standaarden is – met opzet - vrijwel altijd centraal belegd. Anders gezegd: het komt zelden voor dat afdelingen, teams of

<sup>4</sup> *Data Governance: How to design, deploy and sustain an effective Data Governance program*, J. Ladley, Morgan Kaufman, 2012

divisies zelf functies mogen bepalen; conform een eigen systematiek het functioneren van medewerkers mogen beoordelen of zelf bepalen wat een prettige afschrijvingstermijn is. Ze volgen daarin de kaders die voor de hele organisatie gelden.

De centrale HR- en Finance-functie zorgt niet alleen voor het opstellen en beheren van kaders en richtlijnen, maar verleent ook ondersteuning bij specifieke uitvoerende werkzaamheden, bijvoorbeeld het beheer van de personeelsadministratie. Niettemin is een groot deel van noodzakelijke Finance- en HR-*werkzaamheden* in de lijn belegd. Iedere manager is verantwoordelijk voor personeelsmanagement van zijn eigen medewerkers; beheer van begrotingen en bewaken van uitgaven. Niet HR maar de eigen manager voert de reguliere voortgangsgesprekken met medewerkers. Samenvattend – kaders en richtlijnen worden centraal bepaald en bewaakt, de uitvoering vindt in de lijn plaats.

Vanuit deze analogie kan ook het onderscheid Data Governance – Data Management worden beschouwd. Data Governance is hierbij de centrale discipline die zorgt voor kaders en richtlijnen, het schrijft de inrichting voor en vervult een aantal centrale taken. Data Management representeert daarnaast de verzameling disciplines en activiteiten die nodig zijn om data daadwerkelijk te managen. De meeste van deze activiteiten zullen – opnieuw naar analogie met HR en Finance – bij bestaande organisatorische eenheden in de lijn belegd worden. De Data Governance functie betreft typisch een verantwoordelijkheid die bij een Chief Data Officer belegd zou kunnen worden, mits hij of zij om effectief te kunnen zijn in voldoende mate gepositioneerd is binnen de organisatie.

### **Follow the data**

Effectief Data Management betekent betrouwbare, betekenisvolle, veilige en toegankelijke data. Het daadwerkelijk gebruik van data is altijd gerelateerd aan de *fysieke verschijningsvorm* ervan. Bijvoorbeeld in concrete systemen, zoals een CRM-systeem of in dataplatformen die weer andere toepassingen ondersteunen – zoals een datawarehouse of data lake. Of in andere dataverwerkende of datahoudende componenten zoals een API, een fileshare waar aangeleverde bestanden worden opgeslagen, maar ook in zelf gemaakte applicaties zoals in Excel, SAS of MS Access.

Het is op dit (fysieke) niveau waar het resultaat van Data Management zichtbaar wordt. Beleidsnotities, modellen, methodologieën, begrippenwoordenboeken en dergelijke zijn allemaal relevant en nodig. Maar als het niet leidt tot verbeteringen in de manier waarop met concrete data wordt omgegaan, zijn al dit soort initiatieven nutteloos. Die relatie tussen – goedbedoelde – Data Management-initiatieven en de verwerking van concrete (fysieke) data ontbreekt nog wel eens.

Vandaar dat we zeggen: Follow the data! Dit principe betekent een constant streven naar het leggen van de relatie tussen Data Management-activiteiten en de (verwerking van) concrete fysieke gegevens.

## Ontbrekende disciplines

Er is voor gekozen om – refererend aan DAMA DMBOK – de disciplines Master- en Reference Data Management, Content en Record-management en BI en Datawarehousing niet apart terug te laten komen. Voor Content- en Record-management is deze afweging gemaakt omdat deze discipline op veel terreinen zijn eigen karakteristieken kent. Ook Master Data Management (MDM) en BI en Datawarehousing zijn weggelaten als aparte discipline. Dat komt omdat de onderliggende activiteiten overlappen met andere disciplines als Data Architecture, Data Modeling en Data Development. Deze laatstgenoemde disciplines moeten in beginsel ieder soort data – dus ook master- en reference data en analytics – afdekken.

## Data lifecycle management

Ook dit onderwerp komt niet terug als aparte discipline (zoals dat wel bijvoorbeeld in DMM gebeurt). De reden is dat alle activiteiten die DMM onder lifecycle management schaaft, ook als discipline of aspect al elders terugkomen in het Framework. Data is aan een levenscyclus onderhevig en dat blijkt uit de verschillende aspecten die dit adresseren: van architectuur, requirements, development, operations tot en met archivering en retentie. Dat rechtvaardigt echter onvoldoende een aparte vermelding in het Framework. De vraag is dan immers: wat doet een data lifecycle management-discipline zelf nog wat niet door andere disciplines of aspecten wordt afgedekt? Het lijkt ook een beetje een onzinnige activiteit om sec de lifecycle van data te managen.

# Indeling

## Disciplines

Het Framework onderscheidt negen disciplines. De term *discipline* is grotendeels vergelijkbaar met wat DAMA een *Knowledge Area* noemt. Om verwarring te voorkomen wordt hier het begrip discipline gehanteerd.

## Aspecten

Van iedere discipline worden verschillende aspecten opgesomd. Een *aspect* is ofwel een resultaat ('artefact') wat er zou moeten zijn, of een proces c.q. activiteit die uitgevoerd zou moeten worden. In die zin is het een verzameling van resultaten, processen en activiteiten. Dat is een bewuste, pragmatische keuze: de ene keer laat iets zich gemakkelijk duiden door een concreet resultaat te beschrijven, de andere keer gaat het om een benodigde activiteit. Herkenbaarheid en praktische bruikbaarheid is ook hier dus leidend geweest.

## Dataomgeving of Bedrijfsonderdeel

Praktijkervaring laat zien dat Data Management binnen een organisatie niet één inrichting kent. Er is bijna altijd sprake van meerdere dataomgevingen, bijvoorbeeld een operationele omgeving, een datawarehouse-omgeving of een data science pilot-omgeving. Daarnaast is veelal sprake van meerdere divisies of bedrijfsonderdelen die ieder hun eigen invulling van Data Management kennen. De kolommen zijn bedoeld om ieder van die omgevingen of bedrijfsonderdelen te vermelden. Zo wordt expliciet en inzichtelijk dat er verschillende contexten zijn, die ieder hun eigen inrichting kennen.

Per omgeving is de mogelijkheid opgenomen om deze van een kwalificatie te voorzien. De praktijk leert dat dit vaak gebeurt en dat het betrokkenen helpt om zo'n omgeving op de juiste manier te beschouwen.

Het Framework geeft twee mogelijke indelingen aan:

- Kwadrantindeling. Dit is een verwijzing naar het Kwadrantenmodel<sup>5</sup> dat een onderscheid maakt naar systematische en opportunistische voortbrenging van gegevens.
- Indeling naar Managed en Unmanaged. Dit is een variant op de categorisering naar kwadranten. Het geeft in algemene zin een antwoord op de vraag of het (data)management van een omgeving al dan niet systematisch is ingericht.

<sup>5</sup> [Data Management & Decision Support, The Data Quadrant Model](#)

Afhankelijk van de situatie kan er uiteraard gekozen worden voor een andere indeling. Het bestaan van meerdere omgevingen en/of bedrijfsonderdelen laat overigens onverlet dat centrale Data Governance - over omgevingen heen - nog steeds nuttig en nodig is.

## Samenhang

Er is uiteraard veel samenhang tussen de verschillende disciplines: beveiligingseisen vertalen zich naar standaarden voor anonimisering tijdens de gegevensverwerking; de definities van tabellen en attributen uit een fysiek datamodel (PDM) moeten terugkomen in een data dictionary, een datakwaliteitseis kan geïmplementeerd worden als een constraint in een logisch datamodel (LDM) en rolbeschrijvingen moeten in lijn zijn met bijbehorende activiteiten. Toch is er bewust voor gekozen om de samenhang tussen disciplines niet expliciet in het Framework terug te laten komen.

Deze keuze heeft te maken met de volgende overwegingen:

- Het streven naar eenvoud en overzicht boven volledige interne consistentie.
- Wat de samenhang precies is, is voor een deel arbitrair (horen datakwaliteitseisen bijvoorbeeld bij Data Quality of bij Data Requirements?) en afhankelijk van de reeds bestaande of beoogde inrichting. Ongetwijfeld zal er als het Framework wordt gebruikt discussie ontstaan over de samenhang tussen disciplines en activiteiten. Dat is juist goed en de bedoeling! Zo ontstaat meer inzicht en kan er een bewuste afweging worden gemaakt hoe daarmee om te gaan.

Wel is ernaar gestreefd de overlap tussen disciplines en met name aspecten onderling zoveel mogelijk te vermijden.

# Toelichting per discipline

Dit hoofdstuk geeft een toelichting op de verschillende disciplines en aspecten van het Framework. Het doel is de lezer op weg te helpen om het Framework effectief te kunnen gebruiken. De toelichting wordt zo beknopt mogelijk gehouden; verdere verdieping is terug te vinden in beschikbare literatuur, achterliggende raamwerken en bestaande referentiekaders.

## 1. Data Governance

Eerder is al de centrale positie van Data Governance toegelicht. Hier wordt volstaan met een toelichting op de onderliggende aspecten.

### **Strategie en beleid**

Dit heeft betrekking op de strategie en beleid voor de hele, integrale Data Management-functie (= alle disciplines). Op basis van beleid wordt onder andere bepaald welke disciplines nodig zijn, welke (generieke) kaders daarvoor gelden en hoe disciplines aan elkaar relateren.. Het betreft hier *Data Management*-strategie en beleid: wat er is nodig aan Data Management, hoe richten we dat in? Welke overwegingen spelen daarbij een rol?

Een Data Management-strategie wordt vaak verward met een *data*-strategie. In een data-strategie worden richtinggevende keuzes gemaakt ten aanzien van de *data* zelf en de toepassing ervan. Bijvoorbeeld het strategische belang van data voor de organisatie, de wijze waarop data ingezet wordt ter ondersteuning van primaire processen of bijvoorbeeld het stellen van doelen ten aanzien van 'datagedreven werken'. Dergelijke doelen en overwegingen hebben gevolgen voor het benodigde Data Management. Dat laatste is vervolgens weer onderdeel van de Data Management-strategie. Een datastrategie hangt daarentegen veel sterker samen met een bedrijfsstrategie. Uit de bedrijfsstrategie volgt de data-strategie, de Data Management strategie wordt gebaseerd op de data-strategie.

### **Organisatie, functies en verantwoordelijkheden**

Dit betreft de organisatorische inrichting van Data Management. Welke functies en verantwoordelijkheden zijn er nodig en waar worden die belegd in de organisatie. Is er een centrale stafafdeling nodig in de vorm van een Chief Data Office? Wat zijn de verantwoordelijkheden die daarbij horen? Welke andere ondersteunende afdelingen zijn nodig? Welke verantwoordelijkheden beleggen we in de lijn? Welke overlegstructuren en besluitvormingsorganen worden gebruikt?

Hoewel dit onderwerp 'maar' een aspect in het Framework is, is het wel een van de meest verstrekkende en besproken onderwerpen van Data Governance. Niet alleen beslaat de organisatorische inrichting alle andere disciplines en aspecten. Het gaat hier ook om namen en rugnummers: wie is waarvoor verantwoordelijk? Gebruikelijke rollen als 'data steward', 'gegevenseigenaar' of overlegstructuren als een Data Governance-board worden in het kader van dit aspect vormgegeven.

### **Roadmap en planning**

Het ontwikkelen en verbeteren van de *functie* Data Management is niet iets wat alleen tijdelijk in de vorm van een veranderprogramma plaatsvindt maar vanuit een staande inrichting periodiek aandacht vergt. Zo maakt iedere 'normale' afdeling vaak een jaarplan van wat ze het komend jaar van plan zijn: welke projecten pakken we op; welke investeringen zijn er nodig; welke verbeteringen willen doen? Dat is voor Data Management niet anders.

### **Opleidingen en communicatie**

Betreft het communiceren over de opzet en inrichting van de hele Data Management-functie en het coördineren, organiseren en aanbieden van (interne) opleidingen met betrekking tot de verschillende disciplines en aspecten. Voorbeelden zijn trainingen voor data stewards; bewustwording ten aanzien van datakwaliteit en trainingen voor specifieke tools.

### **Managementinformatie, finance en control**

Dit betreft de activiteiten om – al dan niet conform generieke organisatiestandaarden – te voorzien in managementinformatie over het functioneren van de Data Management-functie zelf. Ook het beheren van budgetten en het bewaken van investeringen valt hieronder. Denk hierbij aan overzichten over aantal lopende (data) projecten, investeringen per discipline, statistieken over data-incidenten en dergelijke.

Ook aanverwante aspecten die tot doel hebben de besturing en inrichting van de Data Management-functie te verbeteren vallen hieronder. Denk aan het laten uitvoeren van een periodieke audit of het uitvoeren van een volwassenheidsmeting.

### **Risicomanagement en business case**

Betreft standaard risicomanagement-processen en het uitvoeren van kosten/baten analyses van nieuwe initiatieven en projecten op het gebied van Data Management.

### **Data-administratie**

Voor effectief Data Management is de aanwezigheid van een volwaardige data-administratie onontbeerlijk: een integrale administratie van alle data die binnen een organisatie aanwezig is. Hoe kun je immers iets managen als je niet eens weet wat er te managen valt? Zie hier ook de analogie met Finance en HR: het equivalent voor Finance is de boekhouding (het 'boekhoudsysteem') en voor HR de personeelsadministratie.

De data-administratie is bewust als onderdeel van de (centrale) Data Governance-discipline opgenomen. Het opzetten en bijhouden van een dergelijke administratie dient volgens een welomschreven systematiek te gebeuren. Om dat te waarborgen ligt een centrale positionering voor de hand. De Data Governance-discipline vervult voor dit aspect een uitvoerende en ondersteunende rol. Of, om het heel concreet te maken: centrale stafmedewerkers die mutaties in de data-administratie beheren.

Hoe zo'n data-administratie kan worden opgezet is een vraagstuk op zich dat bovendien nog amper in bestaande literatuur of raamwerken terugkomt<sup>6</sup>.

### **Data-issuemanagement**

Het afhandelen van data-issues hangt samen met de data-administratie. Een data-issue kan meerdere vormen hebben, zoals een kwaliteitsprobleem, inconsistente definities, het overschrijden van een bewaartermijn of compliance issues. Ieder issue hangt samen met (concrete) data die als het goed is, is beschreven in de data-administratie. Deze twee aspecten zijn daarom nauw met elkaar verbonden. Naast het administratieve aspect, is het ook het issuemanagement-proces zelf van belang. Dat valt ook onder dit aspect.

Data-issuemanagement zorgt voor de inrichting en het beheer van een gemeenschappelijk administratie en de inrichting en uitvoering van bijbehorende afhandelprocessen. In de uitvoering hebben ook de andere disciplines een verantwoordelijkheid.

## **2. Data Architecture**

Data Architecture wordt gepositioneerd als de discipline die *inhoudelijk* sturing geeft aan de data zelf en de verwerking ervan in de meest brede zin van het woord. Daar waar de Data Governance-discipline de overige disciplines inricht en aanstuurt, doet Data Architecture dat voor de data en de verwerking ervan.

### **Beleid en spelregels**

Dit aspect betreft het opstellen en vastleggen van principes, spelregels, beleid en/of standaarden ten aanzien van de data architectuur-discipline zelf. Het geeft daarmee kaders voor de uitvoering van activiteiten die binnen deze discipline vallen.

Het aspect *Beleid en spelregels* komt ook bij de andere disciplines terug. De betekenis ervan is dus iedere keer specifiek voor die discipline: welke spelregels spreken we af voor datakwaliteit, voor datamodellering, voor gegevensbeveiliging et cetera?

<sup>6</sup> Een data-administratie is anders dan een metadata-management-oplossing. Bij dit soort technologie staat vaak het catalogiseren van fysieke data (databases, tabellen, folders, files) centraal. Een data-administratie manifesteert zich primair op een hoger abstractie-niveau: de essentie is het in kaart brengen van *gegevensgebieden* en *gegevensverzamelingen* en de relatie met fysieke objecten.

## **Data(model) architectuur, Gegevensbeheersysteem architectuur en Datalogistieke architectuur**

Deze drie architecturen vormen de kern van de Data Architecture discipline. Een logische vraag is wat architectuur nu precies behelst. Wij volgen hiervoor de benadering van Sturm en Van der Sanden<sup>7</sup> :

“Architecturen dienen voor de sturing van de kwaliteit van de informatievoorziening. [...] Een architectuur van een object beschrijft de componenten waaruit het is opgebouwd, de functie en het doel van die componenten en hun samenhang.”

De hiervoor genoemde inhoudelijke sturing heeft betrekking op drie objecten:

1. Data(model) architectuur<sup>8</sup>: dit betreft de architectuur van alle gegevens die een organisatie verwerkt. Hierbij gaat het om de grote lijnen: welke groepen van gegevens ('domeinen', 'gegevensgebieden') zijn te onderkennen? Waar raken deze elkaar? Wat is het karakter van ieder van deze groepen?
2. Gegevensbeheersysteem architectuur ('data at rest'): dit betreft de architectuur van de systemen die gegevens beheren. Dat kunnen operationele systemen zijn – een master-data -managementsysteem of de 'datakant' van een ERP-systeem, datawarehouse- of data lake oplossingen en dergelijke. Waar het om gaat is dat er sprake is van een bewuste indeling en ordening van systemen die gegevens beheren.
3. Datalogistieke architectuur ('data in transit'): heeft betrekking op alles wat met het uitwisselen van gegevens te maken heeft: welke koppelingen zijn wel/niet toegestaan; welke functionele componenten zijn vereist? Welke gelaagdheid in gegevenslogistiek is nodig?

## **Enterprise Alignment**

Data-architectuur staat niet op zichzelf maar heeft samenhang met andere architectuurdisciplines. Die samenhang moet gemanaged worden.

## **Master- en referentiedata aanpak**

Een specifieke aanpak voor master- en referentiedata past integraal in de drie hiervoor genoemde architecturen. Om die reden is het niet per definitie nodig master- en referentiedata apart op te nemen. Omdat niettemin master- en referentiedata vaak als apart verandervraagstuk wordt gezien komt het toch expliciet in het Framework terug. Onder dit aspect kunnen desgewenst specifieke uitwerkingen en/of aanwijzingen worden opgenomen.

<sup>7</sup> *Informatie-architectuur: de infrastructurele benadering*. Sanden, W. A. M. van der, en Sturm, B. (1997). Panfox Holding.

<sup>8</sup> Deze architectuur valt te beschouwen als de échte architectuur van de data: de data architectuur. Maar de term data architectuur wordt meervoudig gebruikt; ook om de hele discipline aan te duiden. Om verwarring te voorkomen spreken we daarom over de data(model) architectuur.

*De disciplines – Metadata, Data Quality en Data Security, Ethics, Privacy en Compliance – hebben een tweezijdig karakter. Enerzijds ligt het voor de hand deze disciplines apart te beschouwen zoals veel raamwerken doen. Ze reflecteren verschillende Data Management-vraagstukken waarvoor eigen benaderingen bestaan. De uitwerking van zo'n eigen benadering heeft echter vaak consequenties voor andere, bestaande disciplines. Het beheersen van datakwaliteit bijvoorbeeld kan ook (of zou moeten) verbeterd worden door aanpassingen in de manier waarop nieuwe data wordt ontwikkeld. Concreet betekent dit bijvoorbeeld dat bij het opstellen van data-requirements ook de kwaliteitseisen worden meegenomen of dat bijvoorbeeld in een logisch datamodel ook constraints worden gemodelleerd. Hetzelfde geldt in feite ook voor metadata, security en aanverwante aspecten. Met die bril moet dan ook naar deze aspecten worden gekeken. Omdat vraagstukken rondom deze disciplines in de praktijk vaak apart worden opgepakt, zijn ze hier ook separaat opgenomen.*

### 3. Metadata

Metadata(management) is een breed onderwerp en er zijn veel soorten metadata. We beperken ons hier tot wat vaak 'business metadata' wordt genoemd. Hieronder vallen verschillende vormen van definities die beschrijven wat de data betekent.

Andere vormen van metadata zijn uiteraard ook belangrijk. Denk aan de logging van de verwerkingsprocessen of 'technische metadata' waarin eigenschappen van programmatuur vastligt. Er is voor gekozen om deze soorten metadata hier niet als onderdeel van de discipline op te nemen. Dit is gebaseerd op het uitgangspunt dat de andere soorten metadata vaak (impliciet) worden meegenomen bij andere disciplines en aspecten, zoals Data Development en Data Operations.

#### **Begrippenwoordenboek**

Dit betreft het vastleggen en beheren van begrippen (termen en hun definities) zoals die in de primaire bedrijfsvoering worden gehanteerd, zoals 'Klant', 'Inkomen', 'Polis', 'Contract' en 'Omzet'. Deze begrippen hebben vaak een relatie met objecten in conceptuele (gegevens)modellen.

#### **Datadefinities**

Hieronder wordt verstaan het vastleggen en beheren van de definities van datapunten zoals deze voorkomen in logische gegevensmodellen, fysieke gegevensmodellen en de daadwerkelijke data. Wanneer in een database de tabel `ACCOUNT` voorkomt en attributen als `ACCOUNT_CATEGORY` is het zaak de betekenis ervan te kennen en indien nodig vast te leggen. Dit wordt steeds belangrijker naarmate labels steeds abstracter worden. Een kolomnaam als `V13_COL1_R2` is niet ongebruikelijk.

#### **Horizontale- en verticale lineage**

Dit behelst het inzichtelijk hebben en bewaken van de samenhang binnen de verschillende gegevensvoortbrengingen, respectievelijk de samenhang tussen het conceptuele, logische en

fysieke niveau. Horizontale lineage is gericht op de samenhang tussen de verschillende voorkomens van dezelfde data in verschillende gegevensbeheersystemen en datalogistiek. Dit is met name van belang wanneer data wordt hergebruikt, gekopieerd of gerepliceerd. Verticale lineage behelst de samenhang tussen begrippen en betekenissen op conceptueel niveau ('Klant', 'Transactie') en de weerslag daarvan in fysieke gegevensdragers als databases en bestanden.

### **Publicatie en collaboratie**

Het publiceren van de betekenis van gegevens met alle relevante belanghebbenden is essentieel om het gebruik ervan te bevorderen. Collaboratie is gericht op het creëren van mogelijkheden waarmee belanghebbenden zelf een bijdrage kunnen leveren aan het beheer van metadata. Bijvoorbeeld door definities toe te voegen of te verbeteren, relaties te leggen tussen definities of andere vormen van duiding toe te voegen.

### **Technologie**

Hieronder valt de afweging, keuze en inrichting van technologie die gebruikt wordt om begrippen en datadefinities mee vast te leggen en te publiceren. Dit ter ondersteuning van het bewaken van horizontale en verticale lineage. Dat kan aparte metadata management tooling zijn maar bijvoorbeeld ook datamodelleer-tooling of andere oplossingen.

## **4. Data Quality**

Vaak vormt datakwaliteit één van de speerpunten van Data Management. Dat is ook niet verwonderlijk: gebrekkige datakwaliteit komt tijdens het gebruik van data naar boven met alle gevolgen van dien. De moeite die het vervolgens kost om datakwaliteit te verbeteren kan gezien worden als een indicator van de volwassenheid van de Data Management-functie in de organisatie.

Datakwaliteitsmanagement is vergelijkbaar met 'gewoon' kwaliteitsmanagement voor producten, diensten en processen. Dit betekent ook dat kwaliteit door de hele lifecycle van data heen gemanaged moet worden. Datakwaliteitsmanagement is daarom ook een integraal onderdeel van andere disciplines en aspecten die bepalend zijn voor de datakwaliteit. Denk aan de Data Architecture, Data Modeling en Data Development. Met andere woorden: de manier waarop binnen deze disciplines rekening wordt gehouden met kwaliteit, is sterk bepalend voor de uiteindelijke datakwaliteit.

Om die reden wordt de uitwerking van de Data Quality-discipline beperkt tot die aspecten die buiten het reguliere lifecycle management van data vallen.

### **Datakwaliteitsproces**

Dit betreft een integrale aanpak om datakwaliteit te meten, te rapporteren, oorzaken te analyseren, stakeholders te informeren en verbetermaatregelen te nemen.

Belangrijke onderdelen van het datakwaliteitsproces zijn:

- Data Profiling: een vorm van data-analyse om gegevens te inspecteren. Hierbij wordt gebruik gemaakt van statistische technieken om de daadwerkelijke structuur, inhoud en kwaliteit van gegevens te meten<sup>9</sup>. Praktisch gezien betreft het het gebruik van standaard (technische) controles op gegevenswaarden, zoals het voorkomen van lege velden of het bepalen van minimale en maximale waarden.
- Meting en terugkoppeling: aanvullend kunnen niet-technische controles worden uitgevoerd. Het betreft dan metingen van datakwaliteit die contextspecifiek zijn en expliciet moeten worden ontwikkeld. Denk aan functionele afhankelijkheden tussen een status-veld en een andere gegevenswaarde of bijvoorbeeld de controle op toegestane waarden. Het terugkoppelen van de meetresultaten is nodig om een kwaliteitscirkel tot stand te brengen.
- Oorzaakanalyse en verbetering: door de oorzaak te analyseren van kwaliteitsproblemen kunnen verbetermaatregelen worden bepaald. Deze kunnen betrekking hebben op het wegnemen van de oorzaak – bijvoorbeeld door het invoerproces van data te verbeteren – of het opschonen van vervuilde data.

*Hoe* datakwaliteitsprocessen uiteindelijk worden gerealiseerd en waar de *uitvoering* komt te liggen zijn belangrijke inrichtingsbeslissingen. Hier is sprake van samenhang met andere disciplines. Het meten van datakwaliteit kan bijvoorbeeld apart gebeuren (op kopieën van data, met aparte technologie) of worden geïntegreerd in gegevenslogistieke processen, zoals terugkerende dataaadprocessen. Het vastleggen en rapporteren van datakwaliteitsissues kan afzonderlijk – vaak in de tools die kwaliteitsmetingen ondersteunen – of integraal onderdeel zijn van een data-issueregister.

Het is dus belangrijk om bij de aanpak en inrichting van datakwaliteitsmanagement goed af te stemmen met andere relevante Data Management-disciplines.

### **Technologie**

Hieronder valt de afweging, keuze en inrichting ten aanzien van technologie die gebruikt gaat worden om datakwaliteitsmanagement te ondersteunen.

## **5. Data Security, Ethics, Privacy en Compliance**

Genoemde onderwerpen zijn gebundeld in één discipline. De gemene deler is dat er bij ieder onderwerp sprake is van een (extern) normenkader – al dan niet verankerd in wet- of regelgeving – waaraan de gegevensverwerking dient te voldoen. Dit vereist een vertaling naar specifiek beleid, eventuele specifieke implementaties en monitoring op de naleving ervan. Overigens dient er uiteraard bij voorkeur voor deze onderwerpen te worden aangesloten bij de generieke, doorgaans binnen een organisatie al bestaande richtlijnen, standaarden en hulpmiddelen. Binnen het Framework komen dan die componenten terug die in aanvullende zin als noodzakelijk worden beschouwd.

<sup>9</sup> Olson, J. E. (2003). *Data quality: the accuracy dimension*. Elsevier.

- Data Security behelst alle aspecten rondom de vertrouwelijkheid en beveiliging van gegevens. We beperken ons hier tot aspecten van beveiliging die hun rechtstreeks werking op de data zelf hebben. Informatiebeveiliging<sup>10</sup> is een veel breder vakgebied waaronder bijvoorbeeld ook aspecten als fysieke beveiliging, screening van medewerkers en netwerkbeveiliging vallen. Die aspecten hebben we bewust buiten beschouwing gelaten omdat deze doorgaans al voldoende worden meegenomen in generieke aanpakken voor IT-security. Zo kent de overheid de Baseline Informatiebeveiliging Overheid (BIO), één gezamenlijk uniform normenkader voor informatiebeveiliging binnen de gehele overheid.
- Ethics gaat in op het vraagstuk hoe verwerking en gebruik van gegevens in lijn kan worden gebracht met ethische principes.
- Privacy adresseert het vraagstuk van de bescherming van persoonsgegevens. De Algemene Verordening Gegevensbescherming (AVG of GDPR) vormt hier momenteel het belangrijkste uitgangspunt voor.
- Compliance is gericht op het voldoen aan alle overige geldende wet- en regelgeving die op het gebruik en verwerking van data van toepassing is. Hieronder valt ook regelgeving die branche- of contextspecifiek<sup>11</sup> is, zoals HIPAA<sup>12</sup> of Europese regelgeving met betrekking tot verwerking van gegevens over beleggingstransacties<sup>13</sup>.

Voor ieder van de genoemde onderwerpen is beleid nuttig en nodig: dit maakt duidelijk hoe met ieder van deze eigenschappen omgegaan moet worden. Analoog aan Data Quality geldt echter ook hier dat genoemde componenten – beveiliging, privacy, ethics en compliance – door de hele lifecycle van data heen gemanaged moeten worden. Dit impliceert dat het management ervan ook een integraal onderdeel van andere disciplines en aspecten moet zijn. Dit belang kan ook vanuit een andere hoek worden gezien: de AVG schrijft het principe van Privacy by Design<sup>14</sup> voor. Dit betekent dat zowel tijdens het ontwerp van de gegevensverwerking als tijdens de daadwerkelijke gegevensverwerking zelf passende maatregelen worden getroffen om de beginselen van gegevensbescherming na te leven. Dit 'by Design' principe is ook toe te passen op datakwaliteit, gegevensbeveiliging, ethics en compliance. Ook deze eigenschappen dienen op een vergelijkbare manier te worden meegenomen tijdens ontwerp en verwerking van gegevens teneinde aan de corresponderende vereisten te voldoen.

Deze discipline brengt dus vooral eisen met zich mee voor andere disciplines. Desalniettemin zijn er een aantal eigenschappen die expliciete aandacht behoeven, ofwel omdat ze buiten

<sup>10</sup> Zie verder bijvoorbeeld *NEN-ISO/IEC 27002+C1+C2 (nl) -Informatietechnologie - Beveiligingstechnieken - Praktijkrichtlijn met beheersmaatregelen op het gebied van informatiebeveiliging*

<sup>11</sup> Vanzelfsprekend is deze regelgeving alleen van toepassing voor organisaties die gegevens verwerken die onderworpen zijn aan dergelijke specifieke regelgeving.

<sup>12</sup> [Summary of the HIPAA Security Rule](#)

<sup>13</sup> Journal of the European Union, [COMMISSION DELEGATED REGULATION \(EU\) 2017/590](#) of 28 July 2016

<sup>14</sup> Zie artikel 25 [AVG](#). In artikel 25 AVG staat ook privacy by default.

de lifecycle van data om gemanaged moeten worden, ofwel omdat ze zo belangrijk worden gevonden dat aparte vermelding op zijn plaats is.

### ***Issuemanagement***

Geconstateerde issues ten aanzien van één van de genoemde eigenschappen moeten geregistreerd en afgehandeld worden. In beginsel is hier de issue-administratie die onder de Data Governance-discipline is benoemd voor bedoeld. Door één centrale administratie te gebruiken kan immers een totaal en integraal beeld<sup>15</sup> worden opgebouwd van alle issues rondom data.

### ***Meting, monitoring en control***

Dit betreft het invullen van maatregelen om naleving van vereisten ten aanzien van gegevensbeveiliging, privacy, ethics en compliance te meten, te rapporteren en te beheersen. Hier ligt ook een relatie met het aspect 'Managementinformatie, finance en control' binnen de Data Governance-discipline. Bij deze laatste is voorstelbaar dat op een geaggregeerd niveau wordt gerapporteerd, waarbij de monitoring binnen deze discipline meer details bevat. Onder water c.q. achter de schermen kan van dezelfde oplossingen gebruik gemaakt worden om deze monitoring concreet te realiseren.

### ***Data-autorisatie***

Data-autorisatie is het vraagstuk wie welke gegevens op grond van welke criteria mag raadplegen of muteren. Om hieraan invulling te kunnen geven moet het in eerste instantie mogelijk zijn op een eenduidige manier verschillende gegevensverzamelingen te identificeren. Dit is onderwerp van de Data Architectuur-discipline. Daarnaast zijn spelregels nodig die de condities voorschrijven waaraan moet worden voldaan om gegevens te mogen raadplegen. In eerste instantie dienen deze spelregels op functioneel-inhoudelijk niveau te worden uitgewerkt en vastgelegd. Daarna moeten deze uiteraard technisch worden geïmplementeerd en nageleefd.

### ***Dataclassificatie***

Dit betreft primair het classificeren van data ten aanzien van vertrouwelijkheid. Waar dit vaak gebruikelijk en voorgeschreven<sup>16</sup> is voor documenten, wordt dit niet altijd automatisch geassocieerd met gegevens(verzamelingen). Effectief gebruik van data vereist echter ook dat duidelijk is hoe data behandeld moet worden. Verschillende classificaties brengen ook verschillende beschermingsmaatregelen (by design) met zich mee.

### ***Dataretentie***

Niet alle gegevens mogen onbeperkt worden bewaard. Dataretentie is het aandachtsgebied waarbij bewaartermijnen voor gegevens worden bepaald en maatregelen worden genomen

<sup>15</sup> Vergelijkbaar met het concept van een integraal klantbeeld.

<sup>16</sup> Voorgeschreven in de BIO en ISO27002

om tijdig data te archiveren en te vernietigen. De daadwerkelijke uitvoering ervan is onderdeel van Data Operations.

### ***Data Masking / encryptie / pseudo- en anonimisering***

Omwille van eisen vanuit beveiliging, compliance of privacy kan het nodig zijn gegevens te pseudonimiseren of anonimiseren. Dit brengt extra eisen met zich mee aan zowel het ontwerp van de gegevensverwerking als aan de gegevensverwerking zelf. Wanneer dit voor verschillende soorten gegevens herhaaldelijk nodig is, is het zaak te komen tot een set van standaard maatregelen om hierin te voorzien.

### ***Identificatie van persoonsgegevens***

Dit aspect is opgenomen om expliciet aan te geven dat, om te voldoen aan de vereisten vanuit de AVG, duidelijk moet zijn welke gegevens zich kwalificeren als tot persoonsgegevens te herleiden data en wanneer deze kwalificatie (bijvoorbeeld bij het ontwerpen van een nieuwe gegevensverwerking) moet plaatsvinden. Immers, maatregelen kunnen pas worden genomen wanneer duidelijk is om welke gegevens het precies gaat. Dit betekent enerzijds dat er (procedurele) maatregelen moeten worden getroffen om tijdens het ontwerpen van nieuwe gegevens, expliciet te maken welke gegevens het betreft, en anderzijds dat van reeds bestaande en opgeslagen gegevens wordt vastgelegd welke gegevens persoonsgegeven zijn. De data-administratie - onderdeel van Data Governance - is de aangewezen administratie om deze informatie in op te slaan. Organisaties zijn in bepaalde situaties verplicht een functionaris gegevensbescherming (FG) aan te stellen. Deze houdt binnen de organisatie toezicht op de toepassing en naleving van de AVG. Is een FG aanwezig dan kan gebruik worden gemaakt van door deze functionaris voorgeschreven kaders en richtlijnen en hoeft hier niet opnieuw het wiel te worden uitgevonden.

### ***Verwerkingsbeoordeling / Data Protection Impact Assessment***

Op de *beoogde* verwerking van persoonsgegevens dient een beoordeling van deze verwerking plaats te vinden. Zoals een toets op rechtmatigheid, dataminimalisatie en doelbinding. Een dergelijke beoordeling moet altijd plaatsvinden. Wanneer de beoogde verwerking een hoog privacy risico met zich meebrengt, is het in veel gevallen nodig om een Data Protection Impact Assessment (DPIA)<sup>17</sup> uit te voeren.

### ***Verwerkingsregister***

De AVG schrijft voor dat iedere verwerking van persoonsgegevens op hoofdprocesniveau wordt geadmineistreerd in een verwerkingsregister. Een dergelijke registratie kan gecombineerd worden met - of ten minste gerelateerd worden aan - de integrale data-administratie.

<sup>17</sup> In de AVG heet dit een Gegevensbeschermingseffectbeoordeling (GEB)

### **Verwerkersovereenkomst**

Wanneer een deel van de gegevensverwerking door derden wordt uitgevoerd, moet hiermee een verwerkersovereenkomst worden afgesloten dan wel moeten alternatieve formele afspraken worden gemaakt waarmee naleving van vereisten uit de AVG worden gewaarborgd. Dat kan overigens complexer uitpakken dan initieel gedacht, afhankelijk van de vraag bij wie de verwerkersverantwoordelijkheid ligt. Dit kan ook meerdere partijen betreffen en tevens kan er sprake zijn van een gezamenlijke verantwoordelijkheid.

## **6. Data Technology**

Dit bevat alle (basis)technologieën die nodig zijn om gegevens te verwerken en te gebruiken. Zoals eerder aangegeven komen hier niet alle soorten data-technologie terug. Zo is bijvoorbeeld technologie voor datakwaliteit of metadata opgenomen onder de betreffende disciplines. De opsomming die hier gegeven wordt – Opslag t/m Visualisatie – en overige verdeling over de disciplines is dus enigszins arbitrair.

In het kader van deze discipline is het in ieder geval nodig keuzes te maken over welke technologieën voor welk doel in welke dataomgeving gebruikt gaan worden.

### **Voortbrenging en versiebeheer**

De aanschaf van technologie alleen is niet voldoende. Eenmaal beschikbaar dient de technologie – net zoals de toepassingen die met de technologie worden gemaakt – zelf te worden voortgebracht. Wanneer er nieuwe versies van ingekochte technologie beschikbaar komen – bijvoorbeeld een upgrade van het database platform of een nieuwe versie van gebruikte data integratietooling – moet worden nagedacht hoe deze nieuwe versies geïnstalleerd gaan worden. Een voortbrenging van *ontwikkeling – test – acceptatie – productie* (OTAP) van de tooling zelf kan hiervoor nodig<sup>18</sup> zijn.

### **Infrastructuur**

Nieuwe data-technologie brengt soms ook nieuwe eisen met zich mee met betrekking tot de infrastructuur waarop deze technologie moet draaien. Het specifiek aanpassen, configureren en/of parametriseren van onderliggende infrastructuur voor gegevensverwerking is dan nodig.

## **7. Data Modeling**

Voor gedegen Data Management is data-modellering onontbeerlijk. Punt. Het belang van data-modellering wordt helaas lang niet altijd onderkend. Ten onrechte wordt modellering geassocieerd met het maken van schema's ('plaatjes'), vertraging ('wanneer kunnen we beginnen met ontwikkelen?') of oeverloos nadenken over wat iets is of betekent. Data-modellering heeft zeker zijn uitschieters gekend waarin de modelleur in de spreekwoordelijke ivoren toren zijn model aan het maken was terwijl de rest van het ontwikkelteam gewoon aan het bouwen was. Dat dit niet effectief en onwenselijk is, behoeft

<sup>18</sup> On-premise omgevingen kunnen hierin verschillen van cloud-omgevingen.

geen nadere uitleg. Maar het omgekeerde is eveneens onwenselijk: nieuwe gegevensverwerkingen realiseren zonder enig besef van structuur, samenhang en betekenis van gegevens. Niet zelden valt te lezen dat resultaten van analytics of data science-initiatieven tegenvallen of simpelweg niet van de grond komen. Dit is terug te voeren op gebrekkige data in termen van betekenis, kwaliteit en/of samenhang. Het toepassen van data-modellering had dit voor een groot deel kunnen voorkomen.

Data-modellering speelt een rol aan het begin van de data lifecycle en levert een substantiële bijdrage aan de bruikbaarheid van de data<sup>19</sup>. Bovendien zijn de modellen zelf een logische plek om aanvullende eisen en eigenschappen met betrekking tot kwaliteit, beveiliging, privacy en compliance vast te leggen. Zie de opmerkingen die hierover eerder zijn gemaakt bij de betreffende disciplines.

Om te voorkomen dat één van de geschetste uitersten optreed, is het belangrijk bewuste keuzes te maken ten aanzien van gegevensmodellering. Dus geen alles-of-niets benadering maar een weloverwogen aanpak die recht doet aan de eisen van de betreffende context. Een voorbeeld is het (verplicht) modelleren en voortbrengen van data op basis van een model van organisatiebreed in gebruik zijnde kerngegevens en een lossere benadering van gegevens die heel use-case specifiek zijn. Elementaire keuzes ten aanzien van modellering kunnen al binnen de Data Architecture-discipline worden genomen. Binnen de Data Modeling discipline worden die dan verder ingevuld.

Onderdeel van deze inrichting betekent meer dan alleen keuzes ten aanzien van te gebruiken modellen of te hanteren methodieken. Ook impliceert dit dat afspraken worden gemaakt over de samenhang tussen de Data Modeling-discipline en andere discipline, waaronder primair Data Development.

### **Modelleerwijze en -standaarden**

Hieronder vallen verschillende soorten afspraken en standaarden, zoals naamgevingsconventies, afspraken over te hanteren datatypes en het gebruik van naamloze sleutels. Ook de keuzes ten aanzien van de methodieken valt hieronder: zoals 3NF (derde normaalvorm), Anchor-modeling, Data Vault of Dimensioneel. Wat wordt waarvoor en waarom gebruikt.

### **Modelleertools en schematechnieken**

Voor data modellering zijn professionele tools nodig. Een PowerPoint of Visio-tekening volstaat in deze niet, net zomin als een architect een ontwerp van een huis in Visio maakt. Naast de keus voor de juiste tooling, is het raadzaam ook afspraken te maken over te hanteren schematechnieken: wordt bijvoorbeeld de UML-schematechniek gebruikt, Chen of Bachmann.

<sup>19</sup> *Data modeling essentials*. Elsevier. Simson, G., en Witt, G. (2004).

### ***Voortbrenging en versiebeheer***

Net als software, dienen de modellen ook middels OTAP te worden voorgebracht. Terwijl één versie van het model ('productie') wordt gebruikt om concreet op te ontwikkelen, kan ondertussen alweer aan een nieuwe versie van het model worden gewerkt ('ontwikkeling').

### ***Conceptuele, logische- en fysieke modellering***

Ieder van deze drie modellen heeft zijn eigen toepassingsgebied. Het conceptuele model is een technologie-onafhankelijk model en bedoeld voor afstemming tussen modelleur en business stakeholder. FCO-IM (Fully Communication Oriented Information Modeling) is hiervoor uitermate geschikt<sup>20</sup>. Het logisch model is een vertaling van het conceptuele model naar een structuur die technisch implementeerbaar is maar nog steeds (database)platform onafhankelijk is. Het fysieke model is daarop weer vertaling naar een implementeerbaar model voor een specifiek platform (bijvoorbeeld Oracle, PostgreSQL of Microsoft SQL Server).

## **8. Data Development**

Onder deze discipline vallen alle werkzaamheden die tot doel hebben concrete gegevensverwerkingen te realiseren. De aard van deze werkzaamheden kunnen verschillen per dataomgeving. Voor een operationele (maatwerk)toepassing bijvoorbeeld kan dit beperkt zijn tot het implementeren van een fysiek datamodel in een database waarop vervolgens schermfunctionaliteit gebouwd kan worden. In het geval van een analytische toepassing kan de verwerking bestaan uit het realiseren verschillende gegevensontsluitingen, gegevensintegraties en verrijkingen.

### ***Ontwikkelstandaarden en -patronen***

In alle gevallen is het nuttig ontwikkelstandaarden en -patronen af te spreken en toe te passen. Onder een patroon wordt hier verstaan een "reusable solution to a commonly occurring problem within a given context in software design."<sup>21</sup>

### ***Data-requirements-analyse***

Vereisten ten aanzien van data zijn al bij andere disciplines aan de orde gekomen (o.a. Data Quality, Data Security en Data Modeling). Hier is dus sprake van overlap die hopelijk op basis van de gegeven beschrijvingen duidelijk is. Wanneer eenmaal concreet ontwikkeld gaat worden, is requirements analyse nodig, net als in 'normale' softwareontwikkeling. Naast het opstellen van datamodellen en het meenemen van (niet-functionele) eisen ten aanzien van datakwaliteit, beveiliging en andere eigenschappen, is het ook belangrijk om te analyseren en vast te leggen:

- wie de belanghebbenden zijn bij bepaalde data;
- waarvoor zij de gegevens wensen te gebruiken;
- waarom dit voor hen nodig is.

<sup>20</sup> FCO-IM, wikipedia

<sup>21</sup> [Software design pattern](#), wikipedia

Het stellen van de eenvoudige maar doeltreffende vraag: “Voor welk probleem is deze data een oplossing” is hierbij een krachtig hulpmiddel.

### ***GLO-management***

Wanneer sprake is van gegevenslogistieke processen, is het raadzaam – als onderdeel dus van het ontwikkelproces – gegevensleveringsovereenkomsten (GLO) af te sluiten. De GLO is een instrument om afspraken vast leggen tussen de leverancier van gegevens en de ontvanger/gebruiker van gegevens. Die afspraken kunnen betrekking hebben op technische aspecten (bestandsformaat, naamgeving) maar bijvoorbeeld ook op organisatorische eigenschappen (wie is contactpersoon bij de leverancier in geval van issues). GLO's kunnen zowel voor interne als externe gegevensleveringen gebruikt worden en zowel voor inkomende ('inbound') als uitgaande ('outbound') stromen. Een dergelijke GLO registratie kan gecombineerd worden met de integrale data-administratie.

### ***Voortbrenging, testen en versiebeheer***

Hieronder valt de gebruikelijke voortbrenging via OTAP en versiebeheer van alle software die wordt gerealiseerd voor gegevensverwerking. Extra aandacht is vereist voor testen. Ook voor gegevensvoortbrengingen gelden gebruikelijke testprincipes en methodieken waaronder unit-, systeem-, integratie en acceptatietesten.

### ***Data Services***

Dit is de placeholder voor het ontwikkelen van functionaliteit voor gegevensbeheersystemen: de opslag van gegevens, het beschikbaar stellen van gegevens aan andere services en applicaties, en het doorvoeren van mutaties die worden aangevraagd of aangeboden. Bijvoorbeeld de database die onder een kritische toepassing voor het primaire bedrijfsproces ligt.

### ***Datalogistiek***

Hieronder valt het ontwikkelen van gegevenslogistieke verwerkingen: zoals het ontsluiten van brondata; uitlezen van een (externe) API; het integreren van verschillende gegevensbronnen of het genereren van een datamart. Zo wordt bijvoorbeeld in datawarehouse-omgevingen in dit kader gesproken over Extract, Transform en Load (ETL), in de context van data science en data lakes over 'ingestion'. In essentie komt het op hetzelfde neer: data wordt ingelezen, er vindt een bepaalde gegevensverwerking plaats, en de data wordt weer ergens opgeslagen<sup>22</sup>.

### ***Data in OTAP***

Een apart vraagstuk is het gebruik van (test)data bij het ontwikkelen van nieuwe gegevensverwerkingen. Ervaring leert dat het wenselijk is productiedata te gebruiken in ontwikkeling. Maar niet altijd zal dit mogelijk of toegestaan zijn. De discipline Data Security

<sup>22</sup> Een dergelijke verwerking kan ook gebeuren zonder dat er fysieke replicatie van de data plaatsvindt. Welke technische manier gebruikt wordt om een verwerking concreet fysiek te implementeren is een vraagstuk dat bij de realisatie moet worden beantwoord.

dient hiervoor richtinggevend beleid te maken. Op basis daarvan moeten al dan niet maatregelen of hulpmiddelen worden gerealiseerd om de juiste data in de verschillende omgevingen beschikbaar te kunnen stellen. Een voorbeeld is een copier-mechanisme dat een steekproef aan data uit productie haalt, pseudonimiseert / anonimiseert en beschikbaar stelt in de ontwikkelomgeving.

### ***Datamigratie***

Wanneer een gegevensverwerking wordt ontwikkeld die een bestaande vervangt kan het nodig zijn data migraties uit te voeren. Dit is vaak het geval bij het realiseren van nieuwe gegevensbeheersystemen, die al dan niet onderdeel zijn van een pakketimplementatie zoals een ERP-systeem. De migratie zelf kan gezien worden als een gegevenslogistiek vraagstuk, waarbij het oude systeem fungeert als bron en nieuwe systeem als doel.

## **9. Data Operations**

De laatste discipline representeert het operationeel gebruik van data en de ondersteuning die daarbij nodig is. Kort en goed: voor gegevenstoepassingen, of dit nu gegevensbeheersystemen zijn of analytische toepassingen is beheer in de klassieke zin van het woord nodig. Zeker in een analytische context waarin er soms sprake is van een dataplatform dat 'levert' aan data analisten is inrichting van dit beheer aspect vaak een ondergeschoven kindje.

### ***Gebruikersondersteuning***

Gebruikersondersteuning in een data context betekent dat er voorzieningen zijn om afnemers van data te ondersteunen, bijvoorbeeld in geval van vragen over de betekenis of kwaliteit van gegevens. Dit wil overigens geenszins zeggen dat dit ook centraal moet worden belegd. Waar het om gaat is dat gebruikersondersteuning bewust wordt ingericht. Lokale materiedeskundigen voor verschillende gegevensgebieden kunnen hier bijvoorbeeld een prima rol in vervullen.

### ***Incident & Problemmanagement***

Hiervoor gelden in beginsel standaard principes en maatregelen met betrekking tot Incident & Problemmanagement. In geval van incidenten of problemen met de gegevensverwerking – zoals een data-laadproces dat stuk loopt of autorisatieproblemen op een database – moeten deze adequaat worden afgehandeld. Aanvullend heeft dit aspect ook betrekking op de data zelf – het technische proces kan goed verlopen zijn terwijl er problemen in de data zijn ontstaan die vervolgens als incident of probleem beschouwd en afgehandeld moeten worden.

### ***Service Delivery Management***

In de situatie waarin een gegevensvoorziening als leverancier voor andere applicaties of gebruikersgroepen optreedt, ligt het voor de hand afspraken te maken over het niveau van de dienstverlening. Wat mogen gebruikers c.q. afnemers bijvoorbeeld verwachten ten aanzien van gebruikersondersteuning, reactietijden of de frequentie waarin wijzigingen kunnen worden doorgevoerd.

### ***Data Processing Management***

Onder deze noemer worden beheerwerkzaamheden geschaard die tot doel hebben de operationele gegevensverwerking optimaler te laten verlopen. Een typisch voorbeeld hiervan is het verbeteren van de prestaties van databases of ETL-processen (performance tuning). Ook het periodiek uitvoeren van backups of testen van recovery-procedures vallen hieronder.

### ***Data-archivering***

Betreft het daadwerkelijk archiveren of verwijderen van gegevens conform dataretentie richtlijnen.

### ***Beheer datakwaliteit***

Hieronder valt operationeel oplossen van datakwaliteitsproblemen, bijvoorbeeld door het doorvoeren van wijzigingen in gegevensbeheersystemen.

### ***Beheer begrippen en datadefinities***

Het operationeel gebruik van data leidt vaak tot nader inzicht in de betekenis van de data. De kennis over de data wordt door het gebruik ervan verder vergroot. Het is zaak deze kennis vast te houden door deze toe te voegen aan bestaande begrippen en datadefinities.

### ***Beheer master- en referentiedata***

Het (centraal) beheren van master- en referentiedata is vaak een beheertaak die expliciet wordt ingericht. Hieronder valt het opvoeren, muteren en verwijderen van gegevens.