

# Inrichting datagedreven projecten



Bert Dingemans en Yorrick Tillemans  
Mei 2026

<b>Inleiding.....</b>	<b>3</b>
<b>Waarde van data.....</b>	<b>3</b>
Samenvatting.....	3
Definitie van data.....	3
Aspecten van data.....	4
Data Stroom Model.....	4
Data, informatie en kennis.....	6
Data toepassingen.....	7
Data als productiemiddel.....	8
<b>Inrichting data gedreven werken.....</b>	<b>10</b>
Grondplaat voor data toepassingen.....	10
Data gedreven projectstappen.....	12
Betrokkenen in een data gedreven organisatie.....	15
Hiërarchie data gerelateerde rollen.....	18
Data-professional.....	21
Datagedreven capabilities.....	24
<b>Metadata bij data gedreven werken.....</b>	<b>36</b>
Metadata Data gedreven project proces.....	36
Matrices.....	44
Capability bedrijfsrol matrix.....	44
Bedrijfsrol proces matrix.....	45
<b>Over de auteurs.....</b>	<b>46</b>
<b>Bronnen.....</b>	<b>47</b>

# Inleiding

Steeds meer organisaties willen datagedreven werken introduceren in de organisatie. Er zijn al toepassingen bekend waar datagedreven werken met succes is geïntroduceerd. Echter voor organisaties waar datagedreven werken een nieuw paradigma is kan het succesvol introduceren van een datagedreven toepassing een uitdaging zijn. In dit whitepaper gaan we daarom een aantal hulpmiddelen bieden die bij het opzetten van een datagedreven project kunnen bijdragen aan een succesvolle introductie van een datagedreven toepassing.

Hoe ziet datagedreven werken eruit, welke domeinen kent datagedreven werken. Daarom een grondplaat voor datagedreven werken. Dit gaat helpen om de activiteiten van een datagedreven project in te delen en zo een logische beschrijving en volgorde van projectactiviteiten te bepalen.

Daarnaast gaan we kijken naar de bedrijfsrollen binnen het werkveld van datagedreven werken maar ook de rollen van de data-professionals uitwerking voor de bemensing van een datagedreven projectteam. Vervolgens beschrijven we een aantal belangrijke capabilities nodig bij het introduceren van datagedreven werken in een organisatie.

In het laatste hoofdstuk gaan we in op metadata benodigd per projectstap bij datagedreven werken. Bij iedere projectstap wordt metadata geconsumeerd en vervolgens andere metadata geproduceerd. Daarvan wordt hier een beschrijving gegeven.

Dit whitepaper is opgebouwd uit een aantal diagrammen die vervolgens tekstueel worden toegelicht. Deze diagrammen zijn uitgewerkt in de enterprise architectuur modelleertaal ArchiMate 3.2. Reden om voor deze aanpak te kiezen is dat ArchiMate een breed toegepaste modelleertaal is, een open standaard is en voldoende concepten heeft om de diverse aspecten van datagedreven te werken te modelleren. De modellen kunnen ook in een modelleertool geladen worden. Via deze link kun je een AMEF bestand downloaden via <https://metadata.data-docent.nl/upload/idgw.xml> .

# Waarde van data

## Samenvatting

*Veel organisaties onderkennen de waarde van data en willen deze data inzetten om meerwaarde te realiseren bijvoorbeeld door concurrentievoordeel te behalen, concurrerend te blijven of kosten te reduceren. Vaak wordt hiervoor de term 'data gedreven werken' gebruikt. In de praktijk blijkt dat niet iedereen dezelfde definitie heeft bij data. In dit whitepaper kijken we vanuit verschillende dimensies naar data.*

*We beginnen met de definitie van data en kijken naar de aspecten van data die helpen te ontdekken waar knelpunten kunnen ontstaan als we data op effectieve wijze willen inzetten. Vervolgens kijken we naar hoe we vanuit data via informatie kennis kunnen ontwikkelen die ons wijsheid brengt waarmee we uiteindelijk meer gewogen en betere beslissingen kunnen nemen.*

*We kijken naar data toepassingen waarin we vanuit data waarde creëren. Vervolgens onderzoeken we of data een productiemiddel is en hoe we op die wijze data kunnen inzetten. Als we data als een productiemiddel zien dan kunnen vervolgens kijken hoe we een productiemiddel dienen te managen om waarde te creëren.*

## Definitie van data

Data is overal aanwezig en kent daarbij vele verschijningsvormen. Echter, wat is data precies? Zoek je op internet bijvoorbeeld bij encyclo dan kom je bijna 20 definities tegen. Ook op wikipedia wordt een definitie gegeven. Deze definitie staat hieronder en gaan we uitdiepen.

**Een gegeven, ook wel datum:**

- Is een constructieve bijzonderheid: het is de vastgelegde uitdrukking van een feit. De woorden worden praktisch altijd in het **meervoud** geschreven; gegevens en data.
- Gegevens zijn de objectief waarneembare neerslag of registratie van feiten of op een bepaald medium, zodanig dat deze gegevens **uitgewisseld** en voor langere tijd **bewaard** kunnen worden. Data zijn dan ook steeds het resultaat van codering.
- Data zijn in die zin abstract, dat je ze steeds kan **hercoderen** in een ander medium of drager.
- Met deze data wordt er een **model** (een selectief deel dus) van de **werkelijkheid** vastgelegd in de tijd. Ofschoon de werkelijkheid nooit stilstaat, kan deze door het vastleggen van de gegevens toch worden bevroren.

Bron: wikipedia.nl

Het eerste deel van de definitie gaat in op het aspect dat het een uitdrukking is van een feit en dat data bijna altijd in meervoud voorkomt.

Het tweede deel van de definitie is vanuit het perspectief van de waarde van data interessanter. Een waarneembare neerslag of registratie van feiten. Hier zit een bijzonder interessant aspect in dat er blijkbaar een parallelle wereld bestaat binnen de datawereld die op enigerlei wijze een weerspiegeling is van feiten die in de werkelijkheid plaatsvinden. Dit is met name interessant omdat hiermee de data ontstaat rond feiten uit de werkelijkheid en door die data te verzamelen, te bewerken en te gebruiken kan we waarde uit data ontstaan.

Ik wil dat met een aardige metafoor inzichtelijk maken rond data. Stel ik ga met het openbaar vervoer naar de kapper om mijn haar te laten knippen. Ik heb al jarenlang dezelfde kapper en zij weet precies hoe ik het kapsel wil hebben.

Het feit uit de werkelijkheid is daarbij dat ik met het openbaar vervoer reis. In de parallelle wereld betekent dit dat ik bij het instappen bij de bus in- en bij aankomst uitcheck. Door die gebeurtenissen in de werkelijkheid ontstaat er data die een weerspiegeling is van het feit in de werkelijk in de data. Namelijk er vindt registratie plaats waar ik ben in- en uitgecheckt.

Het derde deel gaat in op het feit dat er data verzameld is over mijn reis namelijk twee data entiteiten rond in- en uitchecken. Echter om de data te kunnen gebruiken, bijvoorbeeld over welke kosten gemoeid zijn bij deze reis, moet de data getransformeerd of gehercodeerd worden naar een model op basis waarvan de vervoerder kan berekenen wat deze reis kost. Echter wil de vervoerder weten hoeveel reiziger er per maand op een bepaald traject zijn op welke dag en op welk tijdstip dan zal er wederom een hercodering plaats moeten vinden om dat inzicht te verkrijgen.

Het laatste deel van de definitie bevat een aantal zeer interessante karakteristieken rond data.

Namelijk in de parallelle wereld van data zal iedereen die de data gaat gebruiken een datamodel, dus

een vereenvoudiging van de werkelijk, maken. Echter dat datamodel zal voor iedere organisatie vanuit een eigen perspectief naar de data kijken. De openbaar vervoerder wil weten waar ik in- en uitgecheckt ben op welk moment van de dag en welke soort abonnement ik heb in verband met mogelijke kortingen. Echter de kapper wil weten hoe laat ik kom, of mijn haar eerst gewassen moet worden. Welk kapsel ik heb, misschien hoe lang geleden het is dat ik voor het laatst geweest ben. Ook hierbij bestaat mogelijk een model vanuit het perspectief bijvoorbeeld als een kapper een klantregistratie heeft waarover men een aantal kenmerken bijhoudt. Ook dit is wederom een datamodel maar vanwege het andere perspectief een compleet ander model.

## Aspecten van data

In de definitie in Wikipedia wordt nog iets dieper ingegaan op data, namelijk op de aspecten van data. Hieronder de vier aspecten van data:

- **Pragmatisch aspect** Het pragmatische aspect van gegevens (de gewenste uitwerking van gegevens) heeft betrekking op de uitwerking die een bepaald signaal heeft op gedragingen van de ontvanger.
- **Semantisch aspect** Het semantische aspect van gegevens heeft betrekking op de betekenis van de gegevens. Bij het uitwisselen van gegevens moeten zender en ontvanger dezelfde betekenis toekennen aan deze tekens.
- **Syntactisch aspect** De overdracht van gegevens gebeurt door middel van tekens of combinaties daarvan. De syntaxis legt vast volgens welke regels men tekens of combinaties van tekens mag of moet gebruiken.
- **Technisch aspect** Onder het technische aspect van gegevens wordt de wijze verstaan waarop gegevens worden vastgelegd, verwerkt of getransporteerd. Het technische aspect handelt over het fysieke medium, betrokken bij dergelijke gegevensmanipulaties.

In deze aspecten van data wordt gewerkt vanuit de informatietheorie van Shannon uit 1948. Namelijk er is een verzender van data naar een ontvanger. Deze zender heeft hierbij een bepaald doel namelijk dat de ontvanger van de data deze gebruikt bijvoorbeeld door begrip te vormen en eventueel door actie te ondernemen. Tussen de verzender en ontvanger kan onduidelijkheid ontstaan doordat er aan beide zijden transformatie plaatsvindt en er ruis kan ontstaan bij de overdracht.

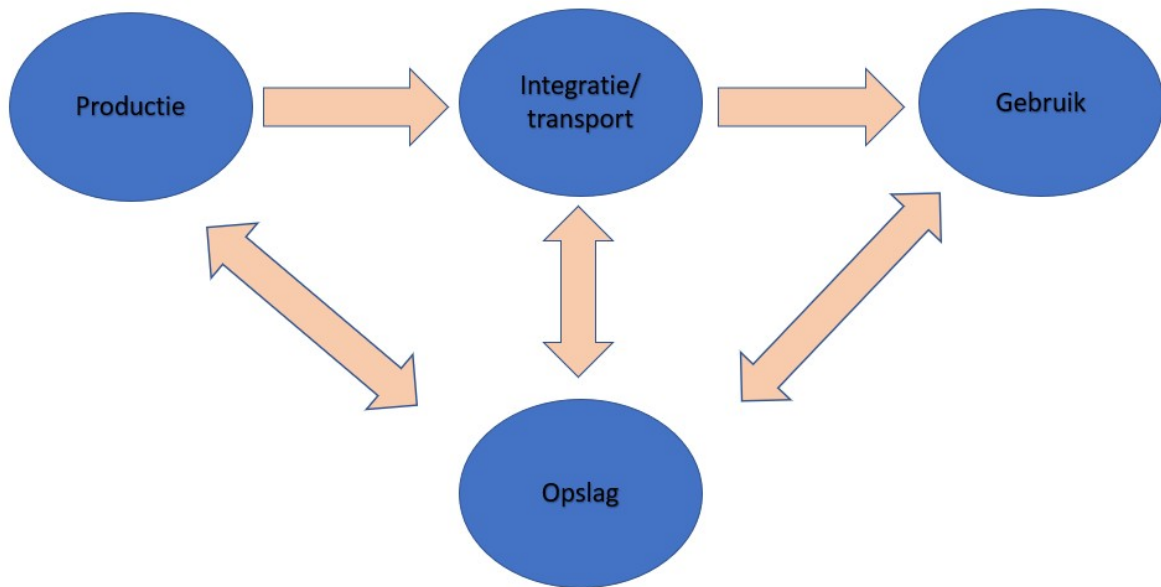
In de vier aspecten van data wordt dit opgesplitst in het gezichtspunt vanuit techniek, syntaxis, semantiek en pragmatiek. En als er in één of meerdere aspecten onduidelijkheid of verwarring is over de data tussen de verzender en ontvanger dan zullen er misverstanden of erger ontstaan.

Vanuit de waarde van dat is dit een belangrijk gegeven. Ook tussen de verzender en de ontvanger van data zal bij onduidelijkheid de waarde van data dalen. Er is tenslotte onduidelijkheid en het is maar de vraag of de ontvanger het juiste begrip krijgt of de juiste actie onderneemt.

## Data Stroom Model

Hierboven zijn we ingegaan op data die tussen verzender en ontvanger uitgewisseld wordt op basis van het model van Shannon. In dit hoofdstuk gaan we op een meer data gedreven wijze kijken naar die data tussen verzender en ontvanger.

Het datastroom model heeft net als het model van Shannon een verzender en ontvanger. Echter vanuit data perspectief spreken we in deze een data producent en een consument. Dat model zie je hieronder in de afbeelding weergegeven en het wordt kort toegelicht.



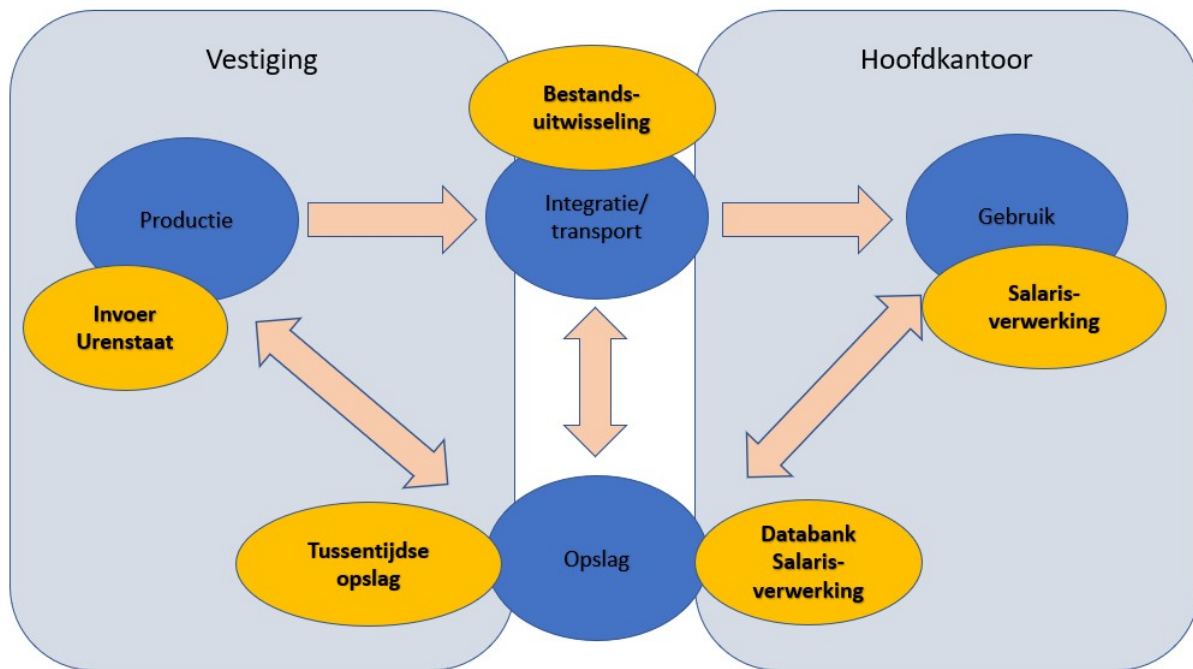
Data wordt geproduceerd en de producent kan in deze van alles zijn. Een persoon die iets intikt in een webformulier, een sensor in een apparaat die iets waarneemt, generatieve AI en dat omzet in data etc.

De data die geproduceerd is wordt ergens gebruikt om daarmee inzichten te krijgen en op basis daarvan actie te ondernemen. Een vertaling van begrip en actie vanuit de informatietheorie.

Echter tussen de productie en het gebruik gebeurt iets interessants vanuit het perspectief van de data en datamodellen.

- Namelijk het model van de data die geproduceerd is zal anders zijn dan de data bij het gebruik. Het voorbeeld van het openbaar vervoer laat dat zien. Het kastje in de bus is een sensor en zal data produceren produceert data in een datamodel gericht op registratie van het inchecken
- Bij de berekening van de reis en de bijbehorende prijs hoort weer data die er heel anders uitziet zoals een model van facturen en factuurregels.
- Daarvoor is integratie en transport noodzakelijk. De integratie zal ervoor zorgen dat de data op een dusdanige wijze getransformeerd wordt waarmee op basis van de geproduceerde data deze gebruikt kan worden. Daarbij kan er transport noodzakelijk bijvoorbeeld om geografische afstand te overbruggen, afstand in tijd of zelfs afstand in model te overbruggen.
- Onder alle stappen hierboven beschreven van productie -integratie/transport - gebruik kan opslag van data wenselijk zijn. Dat is afhankelijk van de situatie. Echter belangrijk is dat als er opslag plaatsvindt van data in bovengenoemde stappen er een datamodel is van de data zoals die binnen de stappen hierboven relevant is.

Ik zal dat toelichten met een eenvoudig voorbeeld.



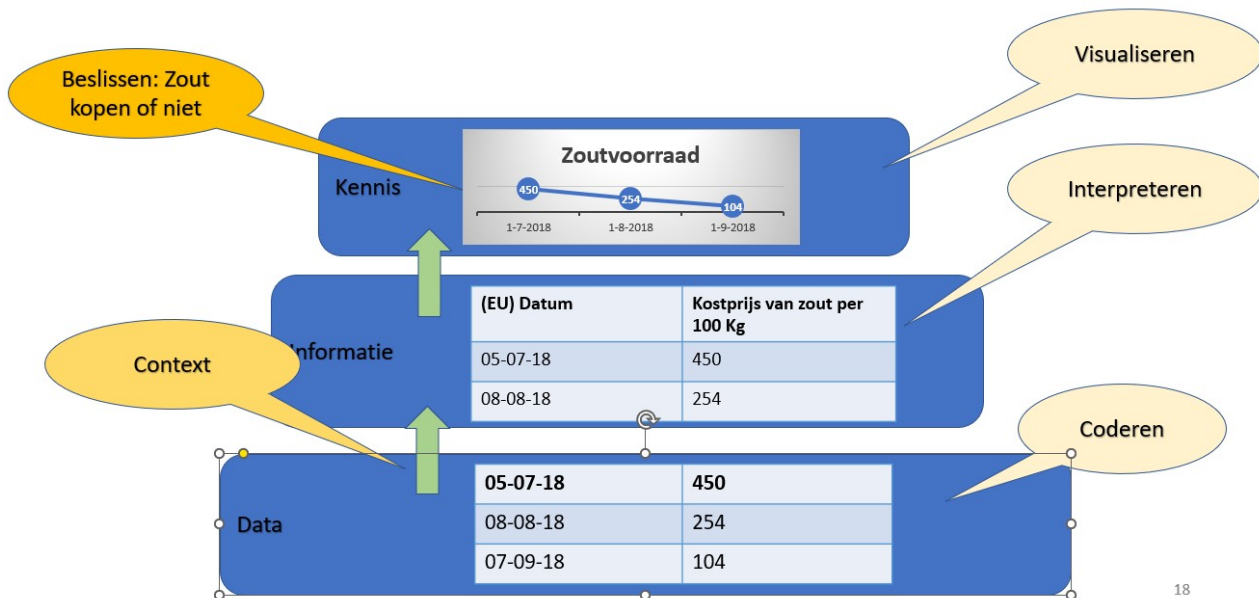
In het voorbeeld van tijdregistratie zal een medewerker uren invoeren in een urenstaat. Dat heeft een (data)model waarmee het eenvoudig is voor de medewerker om de urenstaat in te voeren. Echter wil men van de urenstaten van de verschillende medewerkers salaris- en verlofregistratie doen dan zal hierbij een ander datamodel relevant zijn.

Tussen de urenstaat en de salarisverwerking zit een afstand, bijvoorbeeld in tijd. De medewerker vult dagelijks de urenstaat en stuurt deze aan het einde van de maand op. Mogelijk is er een geografische afstand. De medewerker werkt in het noorden van het land en de afdeling HR werkt in de Randstad. Onder deze stappen zal opslag van data plaatsvinden. In ieder geval bij de urenstaat, want de medewerker zal dagelijks de urenstaat bijwerken op basis van het model van de urenstaat en pas aan het einde van de maand opsturen. Ook de salarisadministratie zal data opslaan in een bepaalde structuur of model. Dit omdat als er later onduidelijkheid kan zijn over het salaris of de verlofregistratie men kan kijken naar de data en bekijken wat daadwerkelijk heeft plaatsgevonden in de salarisadministratie data. Een mooie brug terug naar de parallele wereld in de data en de werkelijkheid.

## Data, informatie en kennis

Rond data zijn reeds een viertal aspecten behandeld vanuit de techniek uiteindelijk naar het pragmatisch aspect van de data en de invloed van de data op de ontvanger. We kunnen dit aspect nog iets verder uitdiepen vanuit het perspectief van de waarde van data. Namelijk met de data, informatie en kennis pyramide.

De data, informatie en kennis pyramide kan het best worden verklaard met een afbeelding die we van onder naar boven in de afbeelding wordt toegelicht.



18

In de afbeelding zien we een eenvoudig willekeurig voorbeeld. In de onderste laag zien we data. Echter deze data bestaan uit een tabelstructuur met waarschijnlijk een kalenderdatum en een reeks getallen. Dit is gecodeerd op enigerlei wijze. Echter alleen met de data kunnen we aan de gebruikszijde van de datastream niets. Er mist namelijk iets bij deze data.

Er mist context. Context is metadata die je aan de data kunt toevoegen waardoor de metadata iets verteld over deze data. Het blijkt inderdaad over een kalenderdatum te gaan en over een kostprijs. Deze metadata wordt in dit voorbeeld weergegeven als een kolomkop in de tabel. Hiermee is data in combinatie met metadata informatie geworden.

Op basis van deze informatie kunnen we dit gaan interpreteren. De kalenderdatums en de getallen vertellen ons iets en op basis van deze informatie, in combinatie met eerder verzamelde informatie, kunnen we kennis vergaren. Deze kennis gebaseerd op de interpretatie van data kunnen we desgewenst visualiseren. Met het visualiseren kunnen we uiteindelijk beslissingen nemen. In ons geval moet iemand met kennis beslissen: moeten we extra zout kopen of niet.

Dit voorbeeld laat zien dat hiermee de waarde van data in relatie tot het nemen van beslissingen een relatie heeft. Is de data van voldoende waarde, heeft de data voldoende kwaliteit, is er goede meta data aanwezig en kunnen we dit correct transformeren naar data. Dan zal het nemen van beslissingen en kennis op basis van de data en de informatie bijdragen aan het nemen van een gewogen of gefundeerde beslissing. Hiermee is een eerste toelichting gegeven op de waarde van data. In het volgende hoofdstuk gaan we dit analyseren vanuit het perspectief van data als productiemiddel.

## Data toepassingen

We begonnen dit whitepaper met de stelling dat data waarde heeft. Veel organisaties zijn daarom op zoek naar deze waarde van data en zoeken naar toepassingen voor data als productiemiddel. In dit hoofdstuk kijken we naar toepassingen.

Binnen organisatie wordt data al op allerlei manieren ingezet. Hieronder vind je een opsomming van een aantal toepassingen

- **Ondersteunen besluitvorming**, besluitvorming wordt gedaan op basis van data die verzameld wordt. Wil je bijvoorbeeld als openbaarvervoermaatschappij weten hoeveel reizigers op een bepaald moment de bus nemen kan op basis daarvan een besluit genomen om de dienstregeling aan te passen en een grote of een kleine bus in te zetten.
- **Verantwoording**, dit moet iedereen op enigerlei wijze afleggen. Een chauffeur van een bus moet verantwoording afleggen over wat hij of zij gedaan heeft, wat de resultaten van het ritten zijn aan de manager. Ook organisaties moeten verantwoording afleggen aan allerlei controlerende- en overheidsinstellingen zoals accountants, ministeries, CBS en DNB. Veelal heeft de verantwoording een verschijningsvorm in dashboards en rapportages (meestal in een vooraf gedefinieerd formaat en metamodel).
- **Inzet (en optimaliseren) in werkprocessen**, is data verzamelen over de werkprocessen in de organisatie. Bijvoorbeeld door de tijdregistratie maar ook door caseloads, wachtlijsten en efficiëntie van de werkprocessen in de organisatie te analyseren. In het geval dat deze data over de processen verzameld wordt dan kan deze data gebruikt worden om het werkproces te

optimaliseren en hiermee kosten besparen, een hogere doorlooptijd te realiseren of een betere dienst te leveren aan de uiteindelijke klanten.

- **Input voor andere informatiesystemen**, data wordt veelal geproduceerd door een samenstelling van sensoren, applicaties, social media en data die ingevoerd wordt door personen achter schermen etc. Deze geproduceerde data in deze samenstelling kan gebruikt worden door andere informatiesystemen die daarmee deze data gebruiken en een andere activiteit ondersteunen. Stel ik zoek op een website op hoe laat treinen en bussen vanaf een bepaalde plek aankomen en vertrekken. Deze websites worden gevoed door de planningsystemen van de afzonderlijke vervoermaatschappijen.
- **Extra dienstverlening en producten**, Organisatie verzamelen al heel veel data ter ondersteuning van de processen in de organisatie. Echter door deze data op een bepaalde wijze te transformeren en te combineren met andere data van andere dataproducten kunnen er nieuwe producten en diensten ontstaan. Bijvoorbeeld in ov-apps kun je zien hoe vol de eerstvolgende trein is kun je zelf besluiten later te gaan.
- **Signalering**, signalering rond allerlei gebeurtenissen die een optimaal proces verstoren voor zowel klanten als medewerkers zijn gebaseerd op data die op basis van patronen in data afwijkingen laten zijn. Door hier een signaal aan te verbinden kan een organisatie vroegtijdig ingrijpen en proberen een verstoring in het proces te voorkomen. Reis ik met het openbaar vervoer en signaleert de vervoerder een verstoring bijvoorbeeld door een kapotte trein dan kan er een signaal gegenereerd worden zodat de treindienstleiding op basis van dit signaal adequate actie kan ondernemen. Een ander type 'verstoring' is fraudedetectie. Hiermee detecteren bijvoorbeeld banken verdachte transacties.

Toepassingen van data zijn er in vele vormen en met name bij organisatie die data gedreven willen werken zijn voorbeelden van dergelijke toepassingen een inspiratiebron voor toepassingen waarmee een data gedreven organisatie meerwaarde kan bieden aan de omgeving zoals de klant, de medewerker en eventueel de controlerende instanties.

## Data als productiemiddel

Productiemiddelen zijn die zaken waarmee een organisatie probeert waarde te creëren. Productiemiddelen zijn bijvoorbeeld geld, grondstoffen, personeel, machines en installaties. Echter is data ook een productiemiddel of niet. In deze paragraaf kijken we naar een aantal voorbeelden van data als productiemiddel en benoemen we een aantal afwijkende kenmerken van data als productiemiddel.

- **Optimaliseren van werkprocessen door inzet van data**, zet je data in als productiemiddel in de werkprocessen dan wordt daar data geconsumeerd, geproduceerd en getransformeerd. Dit kan deze werkprocessen optimaliseren. Bijvoorbeeld zoals in een lopende band waar grondstoffen uiteindelijk worden omgezet tot een eindproduct. Over alle stappen die zorgdragen tot de totstandkoming van een eindproduct kan data bijdragen aan een optimale werkwijze.
- **Sturen van werkprocessen door data**, over data kan stuurinformatie verzameld worden om op die wijze de werkprocessen te sturen. Met name door het kiezen van die databronnen die de inzichten opleveren over het werkproces kunnen sturen en desgewenst kunnen ingrijpen als dit noodzakelijk blijkt. Zoals het reeds genoemde voorbeeld met storing van treinen en het aanpassen van de dienstregeling als dit noodzakelijk is.
- **Data inzetten als concurrentiemiddel**, gebruikt het productiemiddel data om voordeel te hebben ten opzichte van de concurrentie dan kan door het effectief inzetten van deze data om inzichten te krijgen waarmee je kennis vergaard die de concurrentie niet heeft. Bijvoorbeeld een kapper die rekening houdt met de wensen van de klant, bijvoorbeeld de voorkeur van het kapsel en het soort haarproducten dat gebruikt wordt. Bijvoorbeeld door het aanbieden van een nieuw product tegen grijze haren.
- **Verrijken van producten met data**, met name door extra data te geven over de producten die je aanbiedt kun je beter aansluiten op de behoeften van (toekomstige) klanten en daarmee nieuwe markten aanboren waarmee een product interessant wordt door klanten die anders niet in het product geïnteresseerd zouden zijn. Een kapsalon voorbeeld waar op haarproducten een QR code staat die meer informatie geeft gebaseerd op data over hoe dit product gebruikt kan worden voor optimaal resultaat of voor het voorkomen van huid problemen als er allergieën zijn

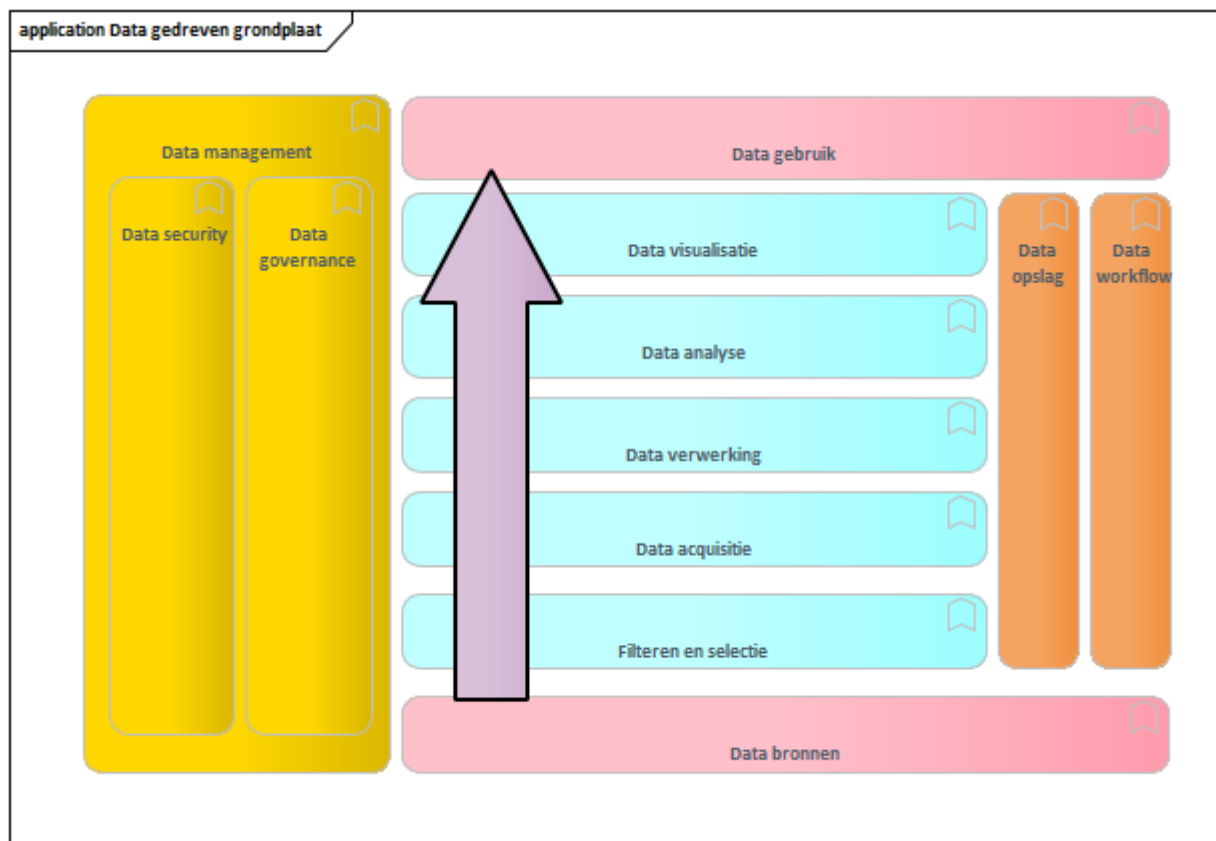
- **Data als product of dienst (disruptieve initiatieven)**, in de afgelopen jaren zijn er een aantal disruptieve initiatieven ontstaan die data verzamelen waarmee traditionele sectoren geraakt worden die daardoor gedwongen worden hun werkwijze en aanbod grondig aan te passen naar deze nieuwe initiatieven die op data transformeren naar relevante data gebaseerd zijn. Een OV voorbeeld is bijvoorbeeld de website Rome2Rio dat waar je je als klant bevindt kunt vinden hoe je met openbaar vervoer je je kunt verplaatsen waar je anders mogelijk een auto zou huren.
- **Business Intelligence en DWH**, deze schuingedrukte beschrijven zijn ondersteunend aan de bovenstaande items over data als productiemiddel. Business Intelligence en Data Warehouses zorgen ervoor dat je op basis van data uit allerhande databronnen een datamodel samenstelt waarmee je op effectieve wijze inzichten krijgt over bijvoorbeeld de organisatie. Dus uiteindelijk op basis van data gefundeerde kennis vergaren en op basis daarvan beslissingen nemen. Bijvoorbeeld een OV vervoerder die inzichten vergaard over de reisbewegingen van reizigers op welke dag en op welke tijdstippen. Om op basis daarvan het dienstrooster desgewenst te optimaliseren.
- **Kunstmatige intelligentie en Data Science**, dit zijn activiteiten die voortborduren op het transformeren en analyseren van data met moderne werkwijze zoals algoritmen en machinaal leren en andere moderne technieken. Met deze moderne methoden verkrijgen die met traditionele hulpmiddelen (nog)niet mogelijk zijn. Bijvoorbeeld de zitplaatszoeker die de klant informeert hoe vol de trein zal zitten en in welk treinstel nog wel zitplaatsen zijn.
- **Big Data oplossingen inzetten in bedrijfsprocessen**, zoekt een organisatie nieuwe databronnen om data als productiemiddel te gebruiken dan kan de productie van data exponentieel toenemen. Hiermee kunnen door te zoeken in deze exponentiele hoeveelheden data uitdagingen ontstaan over hoe gaan we die data allemaal opslaan hoe transformeren we de data tot de gewenste inzichten. Inzet van big data oplossingen faciliteren dit.

Data kan in deze gezien worden als een productiemiddel maar heeft een aantal andere kenmerken. Zo wordt data nooit moe (mensen), data slijt niet (machines), data wordt niet verbruikt (grondstoffen/geld) en data kan eenvoudig getransformeerd worden van de ene structuur naar de ander en desgewenst ook weer terug. Daarmee heeft data een aantal andere kenmerken. Echter data moeten wel onderhouden worden net als machines, personeel en installaties. Want onderhoud je data niet dan zal deze toch in waarde dalen. Met name veroorzaakt door de koppeling tussen de parallelle werelden. In de werkelijk zullen veranderingen plaatsvinden en worden deze veranderingen niet overgebracht in de data dan is de data geen weerspiegeling meer van de werkelijkheid. Bijvoorbeeld ga ik verhuizen en krijg ik een nieuw telefoonnummer en de kapper verwerkt dat niet in de klantregistratie dan is mijn data in waarde gedaald voor deze kapper. Om data als waardevol productiemiddel te beschouwen zullen we de organisatie anders dienen in te richten namelijk op basis van datagedreven werken. In het volgende hoofdstuk gaan we daartoe de inrichting van datagedreven introduceren.

# Inrichting data gedreven werken

Bij data gedreven werken zal de organisatie een inrichting zoeken voor de uitvoer van data gedreven projecten. Data gedreven werken projecten hebben een aantal kenmerkende stappen. Bijvoorbeeld is het kenmerkend dat allereerst de vraagzijde van een data gedreven project. Vervolgens wordt gekeken naar de aanbodzijde zoals de databronnen. Hier is daarom de projectuitwerking een belangrijk onderwerp. Echter ook gaan we in op de benodigde vaardigheden in een data gedreven organisatie. Vervolgens werken we daarbij een aantal benodigde bedrijfsrollen uit bij data gedreven werken.

## Grondplaat voor data toepassingen



Bron: Arcitura

Deze grondplaat toont hoe de data stroomt vanuit de databronnen naar uiteindelijk de data toepassing waar de data gebruikt wordt. Tussen bron en toepassing zal data getransformeerd en geanalyseerd worden om uiteindelijk de toepassing te realiseren.

Daarnaast worden deze data stromen gestuurd door data management en governance. Daarnaast worden de datastromen ondersteund door een aantal faciliterende functies.

Je kunt dit model beschouwen als een whitebox Input Proces en Output model met daaraan gerelateerd de besturing en de ondersteuning. Deze generieke blokken helpen voor duidelijkheid en fasering bij het initieel opzetten en later verbeteren van data oplossingen. Ook helpt het om inzichtelijk te maken waar bepaalde sub-functies in het geheel thuishoren en hoe de samenhang is.

### Toelichting van de dataflow

In het diagram gemodelleerd als applicatie functies vormgegeven als blauwe blokken

### Data bronnen

Databronnen, gegevensverzamelingen die gebruikt worden als grondstof voor de datagedreven toepassing.

**Data acquisitie**

Activiteiten die de verkrijging van de relevante databronnen bewerkstelligen. Activiteiten zijn zowel technisch als contractueel van aard.

**Filteren en selectie**

Bewerken van de gegevens vanuit de databronnen tot die datasets die relevant zijn voor de toepassing

**Data verwerking**

Transformatie, manipulatie en verrijking om het datamodel geschikt te maken voor een adequate data analyse, visualisatie en toepassing.

**Data analyse**

Activiteiten waarbij de getransformeerde data gebruikt wordt voor het zoeken naar verbanden, patronen of statistische verhoudingen.

**Data visualisatie**

Zichtbaar maken van de analyse resultaten ter ondersteuning van de analisten of voor presentatie aan andere stakeholders.

**Data gebruik**

Inzet (toepassing) van data in verschillende vormen van besluitvorming. Hier vindt de uiteindelijke creatie van waarde uit data plaats.

**Faciliterende functies**

Om de dataflow op een verantwoorde en beheersbare manier te faciliteren zijn de volgende blokken benodigd (geel, rode en roze blokken)

**Data governance**

Regelen van het eigenaarschap en de beslissingsbevoegdheden rond de data die gebruikt en getransformeerd wordt in de stappen binnen de grondplaat.

**Data management**

Data is een waardevol productiemiddel en daarom dient er management op plaats te vinden om de waarde van data te verhogen of behouden.

**Data opslag**

Opslag van de data en de tussenproducten in de verwerkingsstappen voor later gebruik in vervolgstappen of andere datapipes. Fysieke opslag van de gegevens in bijvoorbeeld een (relationele) database of een datalake.

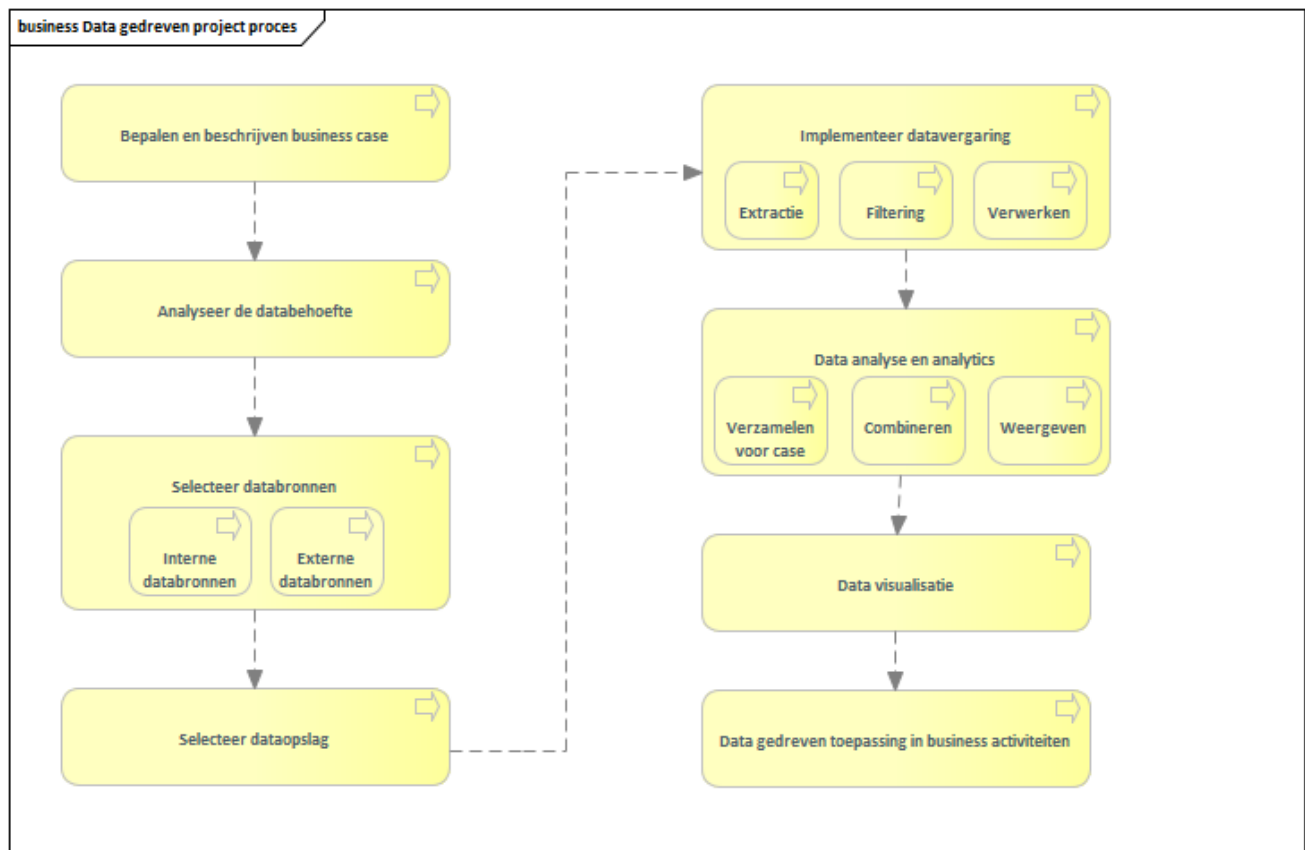
**Data security**

Beveiligings- en privacy aspecten van data.

**Data workflow**

Automatiseren en standaardiseren van bewerkingsstappen op de data en detectie en opvolging van errors in de verwerkingsstappen.

# Data gedreven projectstappen



Dit diagram toont de stappen die ingezet kunnen worden om een data gedreven project in projectstappen in te delen. Voor sommige van deze projectstappen worden ook deelprocessen beschreven.

## Analyseer de databehoefte

Op basis van de business case kan er een inschatting gemaakt worden van de data behoefte. Dat wordt veelal gedaan door een eenvoudig datamodel op te stellen van de databehoefte voor de business case.

Bijvoorbeeld een combinatie van een conceptueel en logisch datamodel van de databehoefte. Zie de whitepapers over datamodelleren

## Bepalen en beschrijven business case

Deze stap beschrijft de business case. Dit is de eerste stap in het proces en heeft een belangrijke relatie met de data gedreven toepassing die ontwikkeld gaat worden.

Meestal wordt de business case opgesteld door de data governance rollen in nauwe samenwerking met de data-professional.

## Selecteer databronnen

Databronnen maken feitelijk geen onderdeel uit van een Data gedreven toepassing. Databronnen zijn de grondstof waaruit de structurering tot informatie plaatsvindt. Daarom is het van belang om goed na te denken wat de juiste bron is. Dit wordt gebaseerd op de analyse van de databehoefte voor deze business case.

### Interne databronnen

Voor data gedreven toepassingen zal een deel van de data afkomstig zijn vanuit de eigen organisatie. Bijvoorbeeld over de activiteiten van de organisatie en hoe de organisatie de waarde levert binnen het vervaardigen van hun producten en diensten.

### Externe databronnen

Als er intern geen databronnen zijn of als de interne datasets onvoldoende kwaliteit hebben voor de databehoefte kan gekozen worden voor externe (betaalde) databronnen.

### **Selecteer dataopslag**

Kies voor een opslag medium voor de geacquireerde data (zoals rdbms, nosql, newsql of polyglot etc).

### **Implementeer datavergaring**

Vervolgens dient de data geacquireerd te worden. De verzamelde brondata wordt vervolgens eerst gefilterd voordat de verwerking plaats kan vinden. Verwerking is het omzetten van de ruwe data naar een structuur waarop data analyse mogelijk is. Vanwege het volume en de velocity is bij big data gedistribueerde verwerking veelal noodzakelijk. Er is een nauwe relatie met data opslag en data integratie.

#### **Extractie**

Extraheren van patronen in de data om inzichten te krijgen die wenselijk zijn om de toepassing succesvol te realiseren.

#### **Filtering**

Filteren van die data elementen die in de databronnen aanwezig zijn maar niet nodig zijn voor deze datagedreven toepassing.

#### **Verwerken**

Verwerken van de data door het uitbreiden van de data entiteiten zodat meer zeggingskracht ontstaat in de data. Ook transformeren van de data tot een analyseerbare structuur is in deze van belang. Context toevoegen, zie ook 'data modelleren'. Voor PoCs kan gekozen worden om nog niet aan te sluiten bij het datamodel van de organisatie, maar dit uit te stellen totdat de toegevoegde waarde is aangetoond. Dit versnelt het proces.

### **Data gedreven toepassing in business activiteiten**

Introduceren van een oplossing waarmee de data gedreven toepassing ingezet kan worden binnen de organisatie. Het introduceren en begeleiding van het gebruik van de toepassing van de business case binnen de organisatie. Dit gaat vaak gepaard met verandermanagement om de adoptie en impact ervan te begeleiden.

### **Data analyse en analytics**

Data-analyse is het gericht zoeken naar (statistische) verbanden in gegevensverzamelingen met als doel profielen op te stellen voor wetenschappelijk, journalistiek of commercieel gebruik. Zo'n verzameling gegevens kan gevormd worden door gebeurtenissen in een praktijksituatie te registreren (aankoopgedrag van consumenten, symptomen bij patiënten, et cetera) of door de resultaten van eerder uitgevoerde wetenschappelijke onderzoeken met elkaar te vergelijken en te herinterpreteren. Bij een AI-agent oplossing wordt op basis van de verschillende datasets stapsgewijs een probleem opgelost, zoals bijvoorbeeld IT-servicedesk support.

#### **Verzamelen voor case**

Verzamelen van verschillende datasets die aanwezig zijn en invulling kunnen geven aan de data analyse voor de desbetreffende business case.

#### **Combineren**

Combineren van verschillende datasets om analyse mogelijk te maken. Traditioneel gebeurt dit door het 'joinen' van tabulaire data, maar in agentic-AI oplossingen door bijvoorbeeld een multi-AI-agent oplossing.

#### **Weergeven**

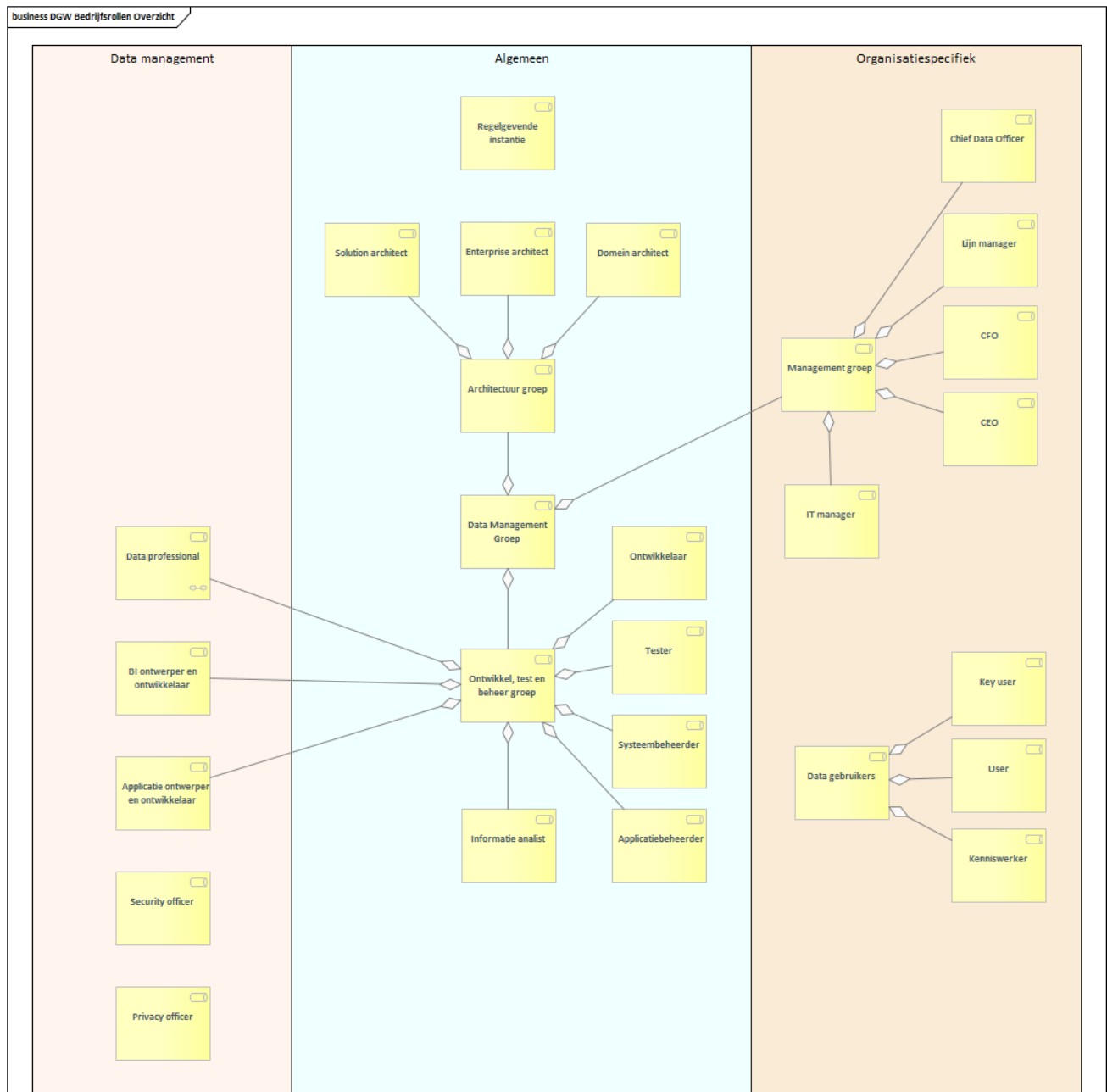
Visualisaties van resultaten, tussenresultaten en verkennende analyse visualisatie in relatie tot de datakwaliteiten. Weergeven is gericht op de doelgroep van data analisten. Denk bijvoorbeeld aan het plotten van locatie gerelateerde gegevens op een kaart om patronen te ontdekken of bij grafieken, hierbij wordt visualisatie gebeurt tijdens de analyse.

**Data visualisatie**

Visualiseren van de data analyse bijvoorbeeld in de vorm van rapportages, dashboards met grafieken en combinaties daarvan. Als laatste kunnen bij de visualisaties ook user interfaces horen voor de gebruikers van de data gedreven toepassing die in het project ontwikkeld wordt. Bijvoorbeeld in apps en websites ter ondersteuning van eindgebruikers.

# Betrokkenen in een data gedreven organisatie

Bij datagedreven werken worden meerdere betrokkenen onderkend. De ene betrokkene zal producent van data zijn, de ander zal data consumeren en er zullen betrokkenen zijn actief meewerken aan de totstandkoming van een datagedreven toepassing(en). Sommigen zullen op een combinatie betrokken zijn bij datagedreven werken. In dit hoofdstuk een overzicht van de meest voorkomende bedrijfsrollen bij datagedreven werken. Vermeldenswaardig hierbij is wel dat iedere organisatie een eigen naamgeving kan hebben voor de rollen in de organisatie. Het model is uitgewerkt op basis van het concept bedrijfsrollen en de aggregatie relatie in de ArchiMate modelleertaal.



Gezien de centrale rol van data in vrijwel alle bedrijfsprocessen, zijn alle bedrijfsrollen op enige wijze betrokken bij de datamanagement kennisgebieden. Met name binnen data gedreven werken projecten zullen zij vanuit hun rol bewuste wensen en eisen hebben. Daarnaast zullen zij een bijdrage leveren aan data gedreven oplossingen. Daarnaast kunnen een aantal rollen ook extern zijn vanuit de scope van de organisatie. We bedoelen hiermee data autoriteiten (bijv. autoriteit persoonsgegevens, nationaal cyber security center etc), maar ook ketenpartners.

## Applicatie ontwerper en ontwikkelaar

Ontwikkelt en ontwerpt technische implementaties binnen de administratieve line of business applicaties binnen de organisatie zoals CRM of ERP pakket.

### **Applicatiebeheerder**

Applicatiebeheerder verantwoordelijk voor de correcte werking van (data verwerkende) informatiesystemen. Er wordt vaak onderscheid gemaakt tussen functionele en technische applicatiemedewerkers. Technisch applicatiebeheer (TAB) is verantwoordelijk voor de correcte werking van een applicatie. De functionele applicatiebeheerder (FAB) zorgt voor een goede inrichting en gebruikersrechten.

### **Architectuur groep**

Groepering van rollen gericht op architectuur in de organisatie van generiek naar gedetailleerde architect rollen zoals domein (bedrijfsprocessen), IT en data.

### **BI ontwerper en ontwikkelaar**

Ontwikkelt en ontwerpt technische datatoepassingen en business intelligence oplossingen.

### **CEO**

Chief Executivel Officer, hoogst hiërarchisch verantwoordelijke voor de algemene organisatie.

### **CFO**

Chief Financial Officer, hoogst hiërarchisch verantwoordelijke voor de financiën.

### **Chief Data Officer**

Een Chief Data Officer (CDO) is de strategische leider die verantwoordelijk is voor het maximaliseren van de waarde van data binnen de organisatie, op een verantwoordelijke wijze. De CDO ontwikkelt en bewaakt de datastrategie, zorgt voor duidelijke governancestructuren en stimuleert datagedreven werken op alle niveaus.

Belangrijke verantwoordelijkheden zijn het definiëren van datastandaarden, het verbeteren van datakwaliteit, het aansturen van datamanagementprocessen en het waarborgen van compliance met wet- en regelgeving. De CDO fungeert als brug tussen business, IT en directie en zorgt dat datainitiatieven aansluiten op de organisatiedoelen.

Daarnaast bouwt de CDO een volwassen dataorganisatie op, inclusief rollen zoals dataeigenaren, data stewards en data engineers. De rol vraagt om visie, leiderschap en het vermogen om verandering te realiseren.

### **Data gebruikers**

Alle data gebruikers binnen en buiten de organisatie hebben een consumerende rol in dit bedrijfsrollen landschap. Er is een onderverdeling gemaakt om in kaart te brengen welke specialisaties van data gebruikers een rol kunnen spelen bij het uitwerken van data gedreven projecten.

### **Data Management Groep**

Dit is de meest generieke groep rond data management. Feitelijk is dit iedere betrokkene binnen de organisatie die op enigerlei wijze betrokken is bij data. Wie is dit niet? In het betrokkenen model is dit daarmee het centraal verbindende punt van het betrokkenen model.

### **Data professional**

Algemene beschrijving van iedereen die als stakeholder vanuit zijn of haar professie betrokken is bij data gedreven werken.

### **Domein architect**

Domein is verantwoordelijk voor bedrijfsprocessen, inclusief data en informatiebeheer van een domein binnen de organisatie zoals bijvoorbeeld sales- of klantdomein etc.

### **Enterprise architect**

Een enterprise architect ontwikkelt en bewaakt de samenhangende structuur van processen, informatie, applicaties en technologie binnen de gehele organisatie. De rol richt zich op het verbinden van strategie en uitvoering, het maken van overkoepelende architectuurprincipes en -kaders, en het zorgen dat veranderingen passen binnen een toekomstbestendig geheel. Een enterprise architect helpt keuzes te onderbouwen, complexiteit te verminderen en investeringen te richten op maximale waarde voor de organisatie.

**Informatie analist**

Informatie analist beschrijft de behoeften van de organisatie rond informatievoorziening gericht op een bepaalde toepassing of organisatie verandering.

**IT manager**

Manager verantwoordelijk voor de IT en informatievoorziening.

**Kenniswerker**

Een kenniswerker is iemand die kennis en informatie tot zich neemt vanuit de binnen en buiten de aanwezige data en deze interpreteert, ontwikkelt en vervolgens gebruikt en distribueert en inzet voor informatie en kennis binnen de organisatie.

**Key user**

Gebruiker veelal afkomstig uit de primaire processen in de organisatie met vergaande detail kennis over een domein binnen de organisatie. Ook bekend als domein expert of Subject Matter Expert (SME).

**Lijn manager**

Manager voor de aansturing van primaire processen binnen Voorbeeld zoals sales, reparatie, voorraadbeheer etc. Inclusief secundaire processen zoals financiën, HRM, juridische zaken etc.

**Management groep**

Groepering van verschillende managers in de organisatie. Meestal gebaseerd op een bepaald primair werkveld.

**Ontwikkel, test en beheer groep**

Groep van rollen die zich bezig houdt met de implementatie van de informatievoorziening in informatiesystemen en technische infrastructuur. Implementatie bestaat uit ontwikkelaars, testen, en beheren van software, databases en IT infrastructuur etc.

**Ontwikkelaar**

Ontwikkelaar van IT voorzieningen in de breedste zin van het woord. Dus zowel software als IT infrastructuur etc.

**Privacy officer**

Een Data Privacy Officer is een onafhankelijke toezichthouder binnen een organisatie die erop toeziet dat persoonsgegevens op een rechtmatige, veilige en transparante manier worden verwerkt. De DPO bewaakt de naleving van privacywetgeving zoals de AVG, adviseert over privacyrisico's en fungeert als contactpunt voor zowel betrokkenen als de toezichthouder.

**Regelgevende instantie**

Organisaties die kaders of regels stellen aan de Voorbeeld organisaties. Denk aan overheidsinstellingen, inspectiediensten maar ook zorg verzekeraars etc.

**Security officer**

Een Data Security Officer (DSO) is de functionaris binnen een organisatie die verantwoordelijk is voor het ontwikkelen, implementeren en bewaken van het informatiebeveiligingsbeleid. De DSO richt zich op het beschermen van data tegen ongeautoriseerde toegang, verlies, misbruik of verstoring, en zorgt ervoor dat technische en organisatorische beveiligingsmaatregelen effectief functioneren. NB: de DSO is niet hetzelfde als de 'functionaris persoonsgegevens' (FPG) die vanuit de wetgever verplicht is bij bepaalde organisaties. De FPG is een toezichthouder en direct aanspreekpunt voor de autoriteit persoonsgegevens. De DSO is geen verplichte rol en is verantwoordelijk voor de uitvoering.

**Solution architect**

Solution architect introduceert veranderingen in de organisatie door het beschrijven van oplossingen om de organisatie te begeleiden van een Ist (huidige) naar de Soll (gewenste) situatie. Veelal zal dit IT gerelateerd zijn, maar dat is niet altijd noodzakelijk.

## Systeembeheerder

Beheerder van de ICT-infrastructuur en applicaties vanuit een technisch gezichtspunt.

## Tester

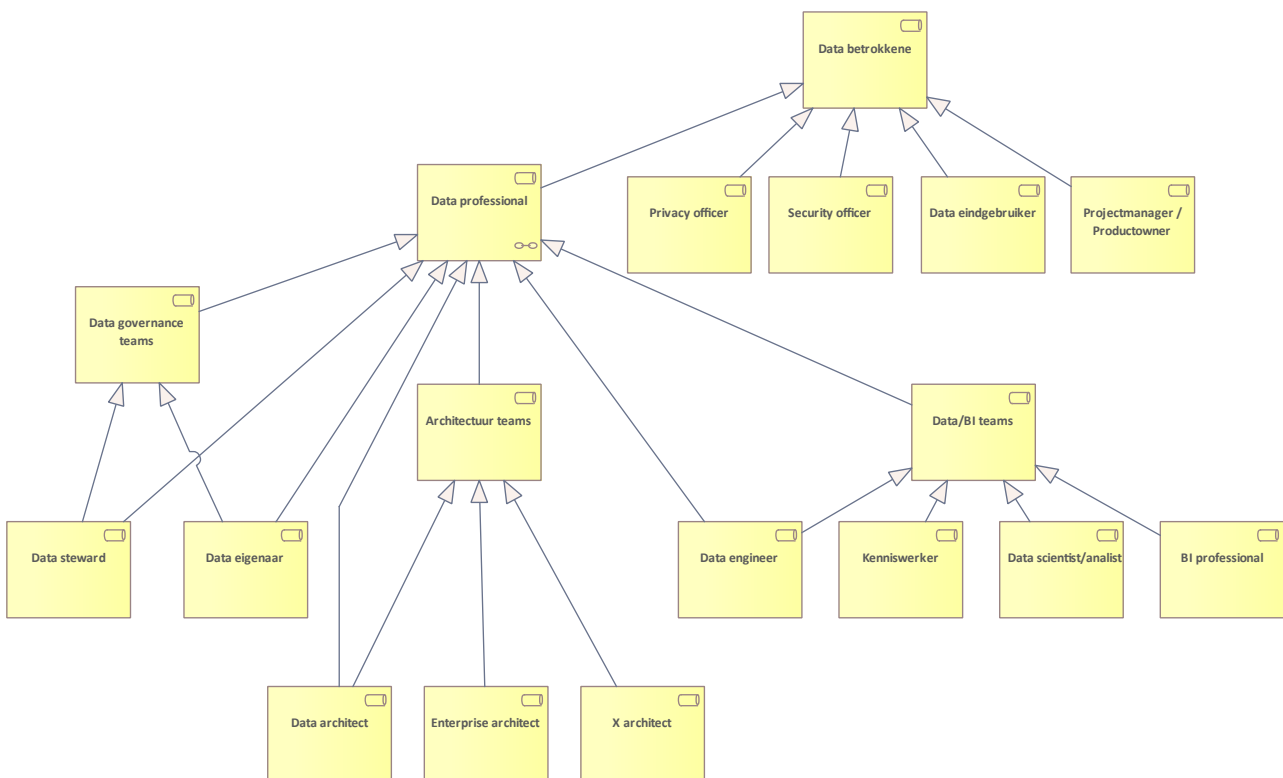
Rol die nieuw ontwikkelde toepassingen en solutions test op basis van de opgestelde specificaties bijvoorbeeld in een solution architectuur of een ontwerp- of designdocument.

## User

Gebruiker van data aanwezig binnen informatiesystemen binnen de organisatie.

## Hierarchie data gerelateerde rollen

Naast de indeling van de betrokkenen kunnen we ook een hiërarchisch model opstellen van de betrokkenen. Hierbij gebruiken we wederom de ArchiMate bedrijfsrollen notatiewijze als element maar nu de specialisatie als relatietype.



Hierarchische uitwerking van de bedrijfsrollen benodigd in een data gedreven organisatie.

### Architectuur teams

Verschillende architectuurteams. Veelal ingedeeld in domeinen of werkvelden waarin verschillende architecten met elkaar samenwerken.

### BI professional

Verschillende rollen samengevoegd tot een BI professional. Omzet van data naar Business Intelligence. Hierbij wordt data omgezet naar informatie, kennis en inzicht. De inzichten worden binnen BI meestal gerepresenteerd in dashboards, rapportages en bestandsformaten. Afhankelijk van het doel en de doelgroep van de BI toepassing. Een BI toepassing is een specifieke vorm van een datagedreven toepassing.

### Data architect

De data-architect vervult een rol in het uitwerken van de data-architectuur. De data-architectuur wordt bepaald door middel van analyse van de informatiebehoeften van een organisatie en wordt weergegeven met behulp van diverse modellen en technieken.

Verschillende gezichtspunten zijn mogelijk in de data-architectuur: conceptueel, logisch en fysiek, waarbij het fysieke niveau de daadwerkelijke implementatie beschrijft, in bijvoorbeeld databases.

De rol van data architect ondersteunt het introduceren van data gedreven werken door het opstellen van kaderstellende en beschrijvende data architectuur producten.

### **Data betrokkene**

Iedere persoon of groep van teams binnen en buiten de organisatie die op enigerlei wijze een relatie heeft met data gedreven werken in de organisatie.

### **Dataeigenaar**

Een dataeigenaar is verantwoordelijk (bij RACI: Accountable) voor de kwaliteit, betrouwbaarheid en juiste toepassing van een specifieke dataset binnen de organisatie. De rol richt zich niet op techniek, maar op inhoudelijke verantwoordelijkheid: de dataeigenaar bepaalt wat de data betekent, hoe deze gebruikt mag worden en welke kwaliteitseisen gelden.

Belangrijke taken zijn het vaststellen van definities, datastandaarden en validatieregels, het beoordelen van datakwaliteit en het nemen van besluiten over wijzigingen in de dataset. De dataeigenaar bewaakt ook de naleving van wet- en regelgeving, zoals privacy en beveiligingsrichtlijnen, en bepaalt wie toegang krijgt tot welke gegevens.

Daarnaast werkt de dataeigenaar nauw samen met data stewards, data engineers en business stakeholders om ervoor te zorgen dat data correct wordt verzameld, beheerd en toegepast. De rol vraagt om domeinkennis, besluitvaardigheid en een duidelijke visie op datagebruik.

### **Data eindgebruiker**

Eindgebruiker van de data, dat kan een interne maar ook een externe eindgebruiker zijn van de dataproducten die aangeboden worden in datagedreven toepassingen.

### **Data engineer**

Een data engineer ontwerpt, bouwt en beheert de technische infrastructuur die nodig is om grote hoeveelheden data betrouwbaar te verzamelen, verwerken en beschikbaar te maken. De kern van de rol ligt in het ontwikkelen van datapipelines die ruwe gegevens uit verschillende bronnen transformeren tot bruikbare, gestructureerde datasets voor analisten, data scientists en applicaties.

Een data engineer werkt met moderne cloudplatformen, databases, streamingtechnologie en automatiseringstools om schaalbare en efficiënte datastromen te realiseren. Belangrijke taken zijn het modelleren van data, het optimaliseren van opslag en performance, het bewaken van datakwaliteit en het implementeren van beveiligings- en governance richtlijnen.

### **Data governance teams**

Governance teams waarin een aantal data professionals met elkaar samen werken. Tenminste de data steward en dataeigenaar nemen deel in deze teams.

### **Data professional**

Algemene beschrijving van iedereen die als stakeholder vanuit zijn of haar professe betrokken is bij data gedreven werken zoals beschreven is in het voorgaande hoofdstuk.

### **Data scientist/analist**

Een data scientist/analist is een specialist die complexe datasets analyseert en modelleert om inzichten, voorspellingen en datagedreven beslissingen mogelijk te maken. De data scientist/analist combineert statistiek, machine learning, programmeervaardigheden en domeinkennis om patronen te ontdekken, modellen te bouwen en waardevolle informatie uit data te halen. Dit kunnen eenmalige of kortlopende analyses zijn, maar ook structurele datatoepassingen.

### **Data steward**

De data steward is een data governance rol die zich richt op tactisch en operationeel niveau van de data governance.

Data steward verricht namens de dataeigenaar de volgende activiteiten:

- Creëren en beheren van de metadata

- Opstellen en beheren van het Bedrijfs object model
- Beschrijven van beleidsrichtlijnen en procedures
- Beheren en oplossen van data kwaliteit issues
- Operationele data governance activiteiten

Binnen RACI is de dataeigenaar accountable (eindverantwoordelijk), maar meestal draagt deze het taken over aan de datasteward die responsible (verantwoordelijk) is.

### **Data/BI teams**

Teams gericht op het ontwikkelen van datagedreven toepassingen inclusief de daarbij horende Business Intelligence toepassingen.

### **Enterprise architect**

Een enterprise architect ontwikkelt en bewaakt de samenhangende structuur van processen, informatie, applicaties en technologie binnen een organisatie. De rol richt zich op het verbinden van strategie en uitvoering, het maken van architectuurprincipes en -kaders, en het zorgen dat veranderingen passen binnen een toekomstbestendig geheel. Een enterprise architect helpt keuzes te onderbouwen, complexiteit te verminderen en investeringen te richten op maximale waarde voor de organisatie.

### **Kenniswerker**

Een kenniswerker is iemand die kennis en informatie tot zich neemt vanuit de binnen en buiten de aanwezige data en deze interpreteert, ontwikkelt en vervolgens gebruikt en distribueert en inzet voor informatie en kennis binnen de organisatie.

### **Privacy officer**

Een Data Privacy Officer (DPO) is een onafhankelijke toezichthouder binnen een organisatie die erop toeziet dat persoonsgegevens op een rechtmatige, veilige en transparante manier worden verwerkt. De DPO bewaakt de naleving van privacywetgeving zoals de AVG, adviseert over privacyrisico's en fungeert als contactpunt voor zowel betrokkenen als de toezichthouder.

### **Security officer**

Een Data Security Officer is de functionaris binnen een organisatie die verantwoordelijk is voor het ontwikkelen, implementeren en bewaken van het informatiebeveiligingsbeleid. De DSO richt zich op het beschermen van data tegen ongeautoriseerde toegang, verlies, misbruik of verstoring, en zorgt ervoor dat technische en organisatorische beveiligingsmaatregelen effectief functioneren.

### **Architecten (Groep)**

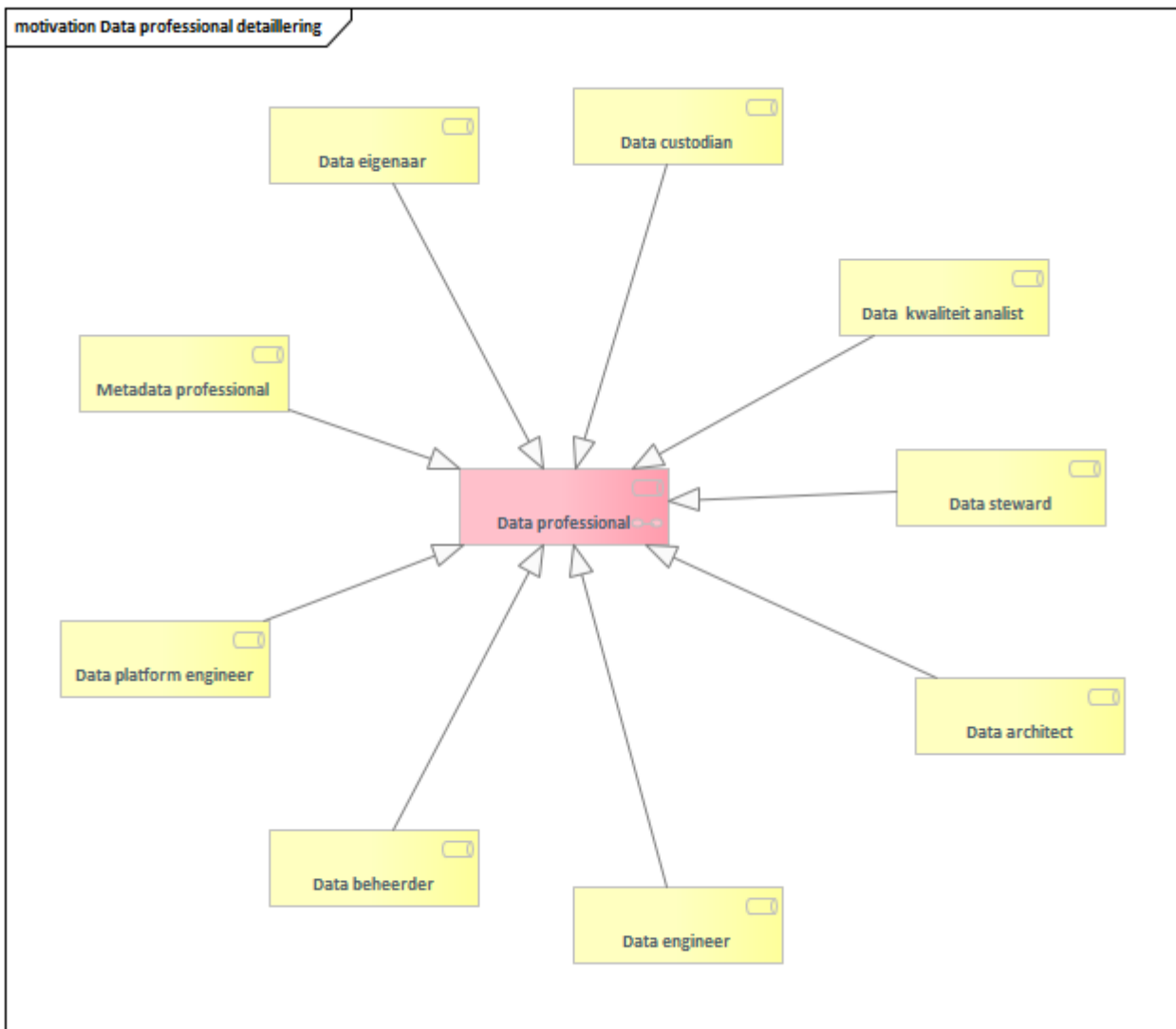
Architecten zijn er in vele vormen en benamingen. Hier worden alleen de data- en enterprise architect benoemd. Maar denk ook aan domein-, solution-, infrastructuur-, proces- en applicatiearchitecten.

# Data-professional

In het voorgaande hoofdstuk zijn we ingegaan op de betrokkenen bij datagedreven toepassingen in een organisatie. In dat voorgaande model hebben we de data-professional uitgewerkt. Data-professionals zijn diegene die bij het introduceren van een datagedreven toepassing betrokken zijn. Met andere woorden deze data-professionals zullen deelnemen in de projectorganisatie van een datagedreven toepassing.

Hieronder wordt een overzicht gegeven van de meest kenmerkende rollen binnen een datagedreven project. Hieronder in deze afbeelding uitgewerkt als specialisaties van de data-professional.

Afhankelijk van het soort datagedreven toepassing die geïntroduceerd wordt zal er een combinatie van rollen geselecteerd worden. Hierbij wordt wederom een ArchiMate diagram uitgewerkt op basis van bedrijfsrollen en specialisaties.



Detailuitwerking van de verschillende rollen van de data professional. Iedere organisatie zal een andere detaillering kennen. Echter in deze uitwerking worden de meest voorkomende specialisaties beschreven en kort toegelicht in relatie tot datagedreven projecten. Dit is een gedetailleerdere uitwerking dan in de hiërarchische indeling. Reden om dit toe te voegen is dat deze data-professionals een belangrijke rol spelen bij datagedreven projecten.

## Data kwaliteit analist

Data professional die zich bezighoudt met het analyseren van data kwaliteits issues en vervolgens het zorgdragen voor het oplossen van deze kwaliteitsissues. Daarnaast is een belangrijke taak van de kwaliteitsanalist dat er impact analyse gedaan worden, er prioritering wordt gegeven aan het oplossen van kwaliteitsissues. Daarnaast zal de kwaliteits analist issues en maatregelen registreren in een informatiesysteem en deze uitwerken tot een kwaliteitsregister.

## Data architect

De data-architect vervult een rol in het uitwerken van de data-architectuur. De data-architectuur wordt bepaald door middel van analyse van de informatiebehoeften van een organisatie en wordt weergegeven met behulp van diverse modellen en technieken.

Verskillende gezichtspunten zijn mogelijk in de data-architectuur: conceptueel, logisch en fysiek, waarbij het fysieke niveau de daadwerkelijke implementatie beschrijft, in bijvoorbeeld databases.

De rol van data architect ondersteunt het introduceren van data gedreven werken door het opstellen van kaderstellende en beschrijvende data architectuur producten.

### **Dataplatform engineer**

Een dataplatform engineer is verantwoordelijk voor het verzamelen, structureren, beveiligen en onderhouden van data binnen een organisatie. De rol richt zich op datakwaliteit, correcte invoer en verwerking, databasebeheer en naleving van privacy en beveiligingsregels. Het doel is ervoor te zorgen dat data betrouwbaar, toegankelijk en veilig is voor gebruik in analyses en bedrijfsprocessen.

### **Data custodian**

Een data custodian is verantwoordelijk voor het operationele beheer van data binnen de technische omgeving. Waar de data-eigenaar bepaalt wat de data betekent en hoe deze gebruikt mag worden, zorgt de data custodian voor hoe de data technisch wordt opgeslagen, beschermd en beschikbaar gesteld.

Belangrijke taken zijn het inrichten en onderhouden van de technische infrastructuur waarin data wordt opgeslagen, het toepassen van beveiligingsmaatregelen, het uitvoeren van backups en het bewaken van toegangsrechten. De data custodian zorgt ervoor dat systemen voldoen aan de eisen van de dataeigenaar en aan interne en externe regelgeving.

Daarnaast werkt de data custodian nauw samen met data engineers, database administrators en security-specialisten om datastromen veilig en betrouwbaar te laten functioneren. De rol vraagt om technische expertise, nauwkeurigheid en een sterke focus op beveiliging en continuïteit.

### **Data eigenaar**

Een data-eigenaar is verantwoordelijk voor de kwaliteit, betrouwbaarheid en juiste toepassing van een specifieke dataset binnen de organisatie. De rol richt zich niet op techniek, maar op inhoudelijke verantwoordelijkheid: de data-eigenaar bepaalt wat de data betekent, hoe deze gebruikt mag worden en welke kwaliteitseisen gelden.

Belangrijke taken zijn het vaststellen van definities, datastandaarden en validatieregels, het beoordelen van datakwaliteit en het nemen van besluiten over wijzigingen in de dataset. De data-eigenaar bewaakt ook de naleving van wet en regelgeving, zoals privacy en beveiligingsrichtlijnen, en bepaalt wie toegang krijgt tot welke gegevens.

Daarnaast werkt de data-eigenaar nauw samen met data stewards, data-engineers en business stakeholders om ervoor te zorgen dat data correct wordt verzameld, beheerd en toegepast. De rol vraagt om domeinkennis, besluitvaardigheid en een duidelijke visie op datagebruik.

### **Data-engineer**

Een data-engineer ontwerpt, bouwt en beheert de technische infrastructuur die nodig is om grote hoeveelheden data betrouwbaar te verzamelen, verwerken en beschikbaar te maken. De kern van de rol ligt in het ontwikkelen van data pipelines die ruwe gegevens uit verschillende bronnen transformeren tot bruikbare, gestructureerde datasets voor analisten, data scientists en applicaties.

Een data engineer werkt met moderne cloudplatformen, databases, streamingtechnologie en automatiseringstools om schaalbare en efficiënte datastromen te realiseren. Belangrijke taken zijn het modelleren van data, het optimaliseren van opslag en performance, het bewaken van datakwaliteit en het implementeren van beveiligings- en governance richtlijnen.

### **Data professional**

Algemene beschrijving van iedereen die als stakeholder vanuit zijn of haar professie betrokken is bij data gedreven projecten.

### **Data steward**

De data steward is een data governance rol die zich richt op tactisch en operationeel niveau van de data governance.

Activiteiten van de data steward zijn:

- Creëren en beheren van de metadata
- Opstellen en beheren van het Bedrijfs object model
- Beschrijven van beleidsrichtlijnen en procedures
- Beheren en oplossen van data kwaliteit issues
- Operationele data governance activiteiten

### **Database administrator**

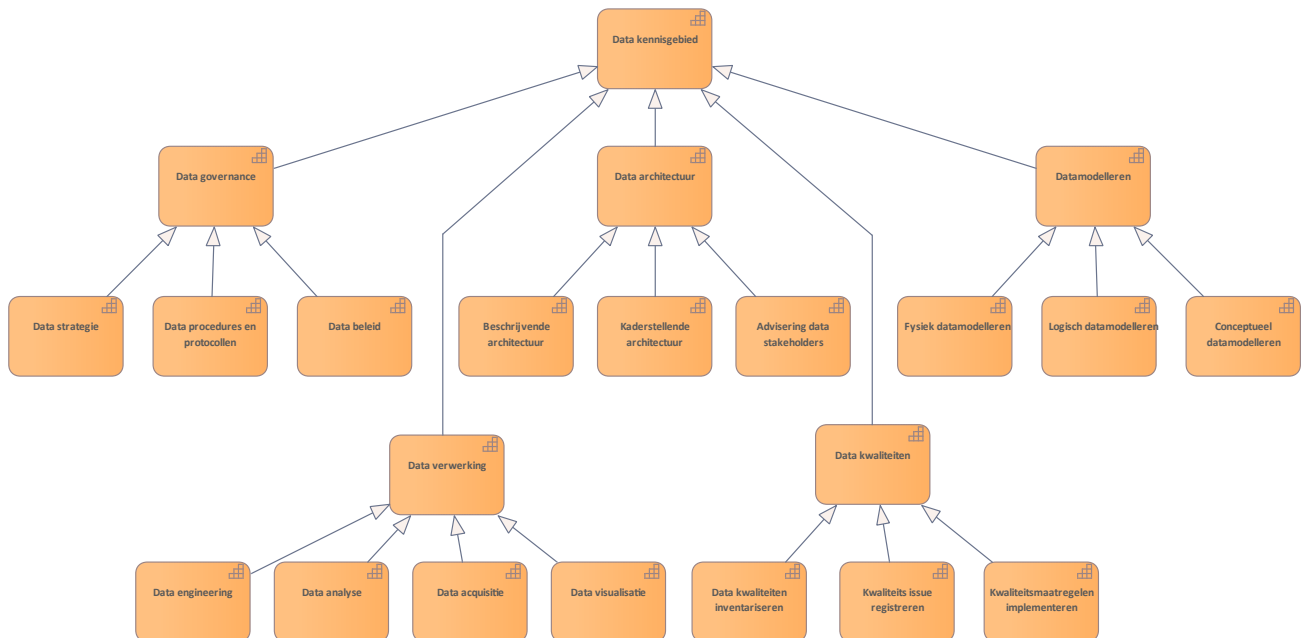
Platform engineer: dataengineer en datascientist hebben opslagmedia (storage) en 'data engines' (compute) nodig om data te kunnen verwerken en analyseren. Een platform biedt hen deze faciliteiten. Door het opzetten en de beheertaak van een dataplatform te centraliseren ontstaat schaalvoordeel en bevordert de compliance. Ook kunnen centrale diensten worden ontwikkeld die veelvoorkomende data taken vereenvoudigen, zoals bijvoorbeeld integraties. Het opzetten en beheren van een dataplatform is de taak van platform engineers.

### **Metadata professional**

Een bijzondere rol meestal gerelateerd aan data modellers, data stewards en data architecten. Zij houden zich bezig met het bepalen welke metadata in welke situatie en welke fase van het data gedreven werken proces noodzakelijk is. Vervolgens dragen zij er zorg voor dat deze metadata wordt verzameld en geregistreerd. Ook het meten van de datakwaliteit van de metadata valt onder deze rol.

# Datagedreven capabilities

Binnen datagedreven werken zijn een aantal capabilities essentieel om een datagedreven toepassing met succes te introduceren. In dit hoofdstuk geven we een overzicht van de capabilities ingedeeld op de belangrijkste kennisgebieden. Vervolgens gaan we in op de capabilities voor de data-professional die de capabilities zal inzetten om een datagedreven toepassing te introduceren. Voor de capabilities maken we gebruik van capabilities elementen uit de strategielag van ArchiMate als relatietype gebruiken we specialisaties en aggregaties.



Dit is een eenvoudig en globaal overzicht van capabilities die wenselijk zijn als je data gedreven werken wilt introduceren in een organisatie. Hier zijn ook andere delen van datagedreven werken relevant denk aan data-integratie toepassingen, APIs en master en reference datamanagement. Echter in dit whitepaper gaan we alleen in op de capabilities in het datafundament van datagedreven werken.

## Advisering data stakeholders

Adviseren van de diverse stakeholders, met name binnen de eigen organisatie over de verschillende aspecten van data gedreven werken.

## Beschrijvende architectuur

Beschrijven van inrichting van het data gedreven landschap. Veelal uitgewerkt in modellen en beschrijvingen van het applicatie en bedrijfsproces landschap in een organisatie.

## Conceptueel datamodelleren

Het conceptuele datamodel beschrijft de structuur van en de relaties tussen de conceptuele gegevensobjecten, genaamd entiteiten. De focus ligt bij deze modellering ligt op semantisch en pragmatisch aspect.

## Data acquisitie

Competencies rond de data acquisitie. Dat houdt in dat er technische toegang geregeld dient te worden. Daarnaast zullen er afspraken en contracten opgesteld dienen te worden in samenspraak met de bronhouders en de afnemers van de data. In een aantal geval zijn ook afspraken van financiële aard onderdeel van de data acquisitie onderhandelingen.

## Data analyse

Analysten van de data voor de data gedreven toepassing. Met name is het het selecteren van de beste of meest adequate data analyse technieken algoritmen. Dit is mede afhankelijk van de aard van de dataset, de te beantwoorden vragen maar ook de datakwaliteiten van de gebruikte dataset.

## **Data architectuur**

De data-architectuur is een onderdeel van de enterprise architectuur en geeft een overzicht van de aanwezige en benodigde gegevens in een organisatie.

De data-architectuur wordt bepaald door middel van analyse van de informatiebehoeften gerealiseerd in data toepassingen van een organisatie en wordt weergegeven met behulp van diverse modellen en technieken. Bij het beschrijven van de data-architectuur worden bijvoorbeeld de volgende middelen gebruikt:

- Overzicht van databases en hun beheer, te denken is aan een datawarehouse
- Data-definities, beschrijving van de data
- Datamodel, logisch en fysiek
- Data flow diagrammen

Verschillende gezichtspunten zijn mogelijk in de data-architectuur: conceptueel, logisch en fysiek, waarbij het fysieke niveau de daadwerkelijke implementatie beschrijft, in bijvoorbeeld databases.

Voor het gegevensbeheer binnen een organisatie zijn ook van belang de

- Applicatie-architectuur
- Informatie-architectuur
- Bedrijfsproces-architectuur.

## **Data beleid**

Het uitwerken van het data beleid door doelen, kaders en een data strategie uit te werken en daadwerkelijk te implementeren in de staande organisatie.

## **Data engineering**

Met name gericht op de transformatie van de data afkomstig vanuit de databronnen naar datasets uitgewerkt naar een analyseerbare structuur van de dataset. Met name als deze transformaties een repeterend karakter krijgen zal data engineering toenemen in belang.

## **Data governance**

Data governance bestaat uit een aantal activiteiten zoals

- Opstellen van governance en management procedures
- Definiëren van data beleid (Visie en Missie voor Data)
- Opstellen van een business glossary en de definitie van standaarden
- Uitwerken van governance rollen in de organisatie
- Coördinatie met Enterprise- en Data architectuur (teams)
- (Economische) waardebeoordeling van data

## **Data kennisgebied**

Containerbegrip voor data management kennisgebieden. Van hieruit worden de DMBok kennisgebieden benoemd met verdere specialisaties relevant voor het beschrijven van de onderliggende bedrijfsprocessen.

## **Data kwaliteiten**

Kwaliteit van data is een kern competentie van organisaties. De kwaliteit van dat is voor veel organisaties een probleem. Daarom dient voor data kwaliteit een organisatie inrichting in de vorm van maatregelen, kaders, processen en bewustwording ingericht te worden.

## **Data kwaliteiten inventariseren**

Voor data kwaliteiten dient op regelmatige basis de kwaliteitsniveaus op basis van kwaliteits dimensies bepaald te worden. Deze inventarisatie is daarbij gericht op de huidige en de gewenste data kwaliteits dimensies.

## **Data procedures en protocollen**

Opstellen van procedures en protocollen op het gebied van data management. Dit is inclusief de data kwaliteits aspecten en data security, privacy en data architectuur.

## **Data strategie**

Een datastrategie is een samenhangend plan waarmee een organisatie bepaalt hoe zij data verzamelt, beheert, beveiligt en inzet om bedrijfsdoelen te realiseren. Het beschrijft de gewenste datavolwassenheid, de benodigde governance, kwaliteit, architectuur en vaardigheden. Een goede datastrategie zorgt dat data betrouwbaar, toegankelijk en waardevol wordt, zodat teams beter kunnen sturen, innoveren en automatiseren.

### **Data verwerking**

Verwerken van de data uit de databronnen naar een datagedreven toepassing. Dit bestaat uit een aantal nauw gerelateerde competencies.

### **Data visualisatie**

Omzetten van de uitkomsten van de data analyse naar een visualisatie voor de doelgroep of het doel van de visualisatie. Denk hierbij aan rapportages, dashboards maar ook datasets die via APIs aan afnemende toepassingen worden aangeboden (via open standaarden).

### **Datamodelleren**

Uitwerken van modellen die de data vanuit verschillende dimensies beschrijven. Hoe wordt de data geproduceerd, gebruikt, verwerkt en opgeslagen.

### **Fysiek datamodelleren**

Data modellering hoe data in technische systemen wordt opgeslagen, verwerkt en getransporteerd. Veelal gericht op het modelleren van databases en API's.

### **Kaderstellende architectuur**

Vanuit diverse werkvelden kunnen kaders gesteld worden aan de productie en het gebruik van data. Denk bijvoorbeeld aan data architectuur, security en privacy maar ook aan het stellen van kwaliteitseisen aan data. Kenmerkend hierbij is, dat de kaders richting geven aan het gebruik van de data, maar ook aan andere aspecten zoals de opslag, het gebruik en incidenteel ook aan de wijze waarop de data geproduceerd wordt. Kaders kunnen op meerdere wijzen beschreven worden bijvoorbeeld als risico, eis of beperking. Hier sluiten we aan op de modelleerwijze van ArchiMate en werken we de kaders uit op basis van requirements en principes. Waarbij we principes beschouwen als generalisaties van de requirements. Naast de kaders zul je binnen een metadatamanagement model ook een uitwerking zien van de maatregelen die genomen kunnen worden om de gestelde kaders te realiseren.

### **Kwaliteits issue registreren**

Registreren van data kwaliteits issues voor het inventariseren van de kwaliteitstoestand van data in informatiesystemen en bij het data gebruik.

### **Kwaliteitsmaatregelen implementeren**

Implementeren van maatregelen om data kwaliteits issues op te lossen. Dit kunnen zowel fysieke, technische, infrastructurele als ook organisatorische en bewustwordings maatregelen zijn.

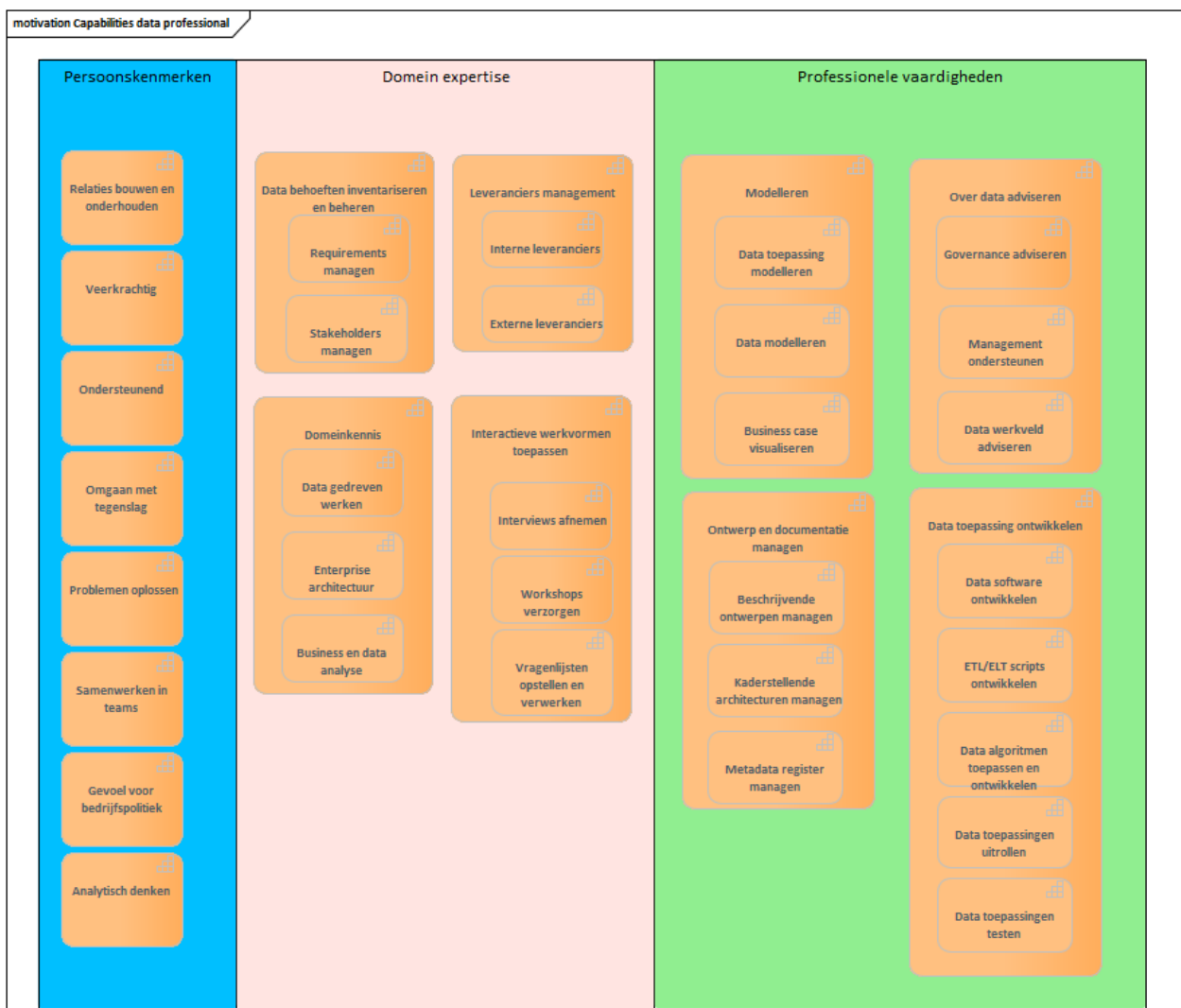
### **Logisch datamodelleren**

Detailmodellering van hoe data wordt gebruikt in een organisatie. Dit is echter zonder de technische implementatie van de data. Het is het startpunt voor de fysieke modellering waar wel rekening wordt gehouden met de implementatie aspecten.

# Capabilities data professional

In een datagedreven projectteam zijn zoals reeds beschreven meerdere specialistische professionals betrokken. Deze data-professionals zullen gezamenlijk een aantal capabilities in dienen te brengen om een datagedreven toepassing met succes te introduceren in een organisatie.

Hieronder daarom een overzicht van de belangrijkste capabilities die je in een datagedreven projectteam zult terugzien.



In deze afbeelding krijg je een overzicht van een aantal generieke capabilities voor de rol van data-professional. In deze afdeling wordt de notatie van ArchiMate gebruikt waarin capabilities in de strategielaaag worden gemodelleerd gebruik makend van de aggregatierelatie om de verbanden tussen de capabilities te modelleren. De nesting van de elementen in de afbeelding is dus een aggregatie.

## Analytisch denken

Analytische vaardigheden vormen een belangrijk onderdeel van de rol van data-professional. Het betekent dieper en dieper graven totdat de ware situatie aan het licht komt en het echte probleem is gedefinieerd. Het gaat om het doorzoeken van vaak tegenstrijdige gegevens en bepalen waar gegevens relevante inzichten opleveren en waar dit niet het geval is. Het betekent geen genoegen nemen met het voor de hand liggende, dingen niet voor waar aannemen en geen overhaaste conclusies trekken. Analytisch denken houdt ook in dat de resultaten van de analyse worden gepresenteerd in een vorm die geschikt is voor de relevante belanghebbenden. En het houdt in dat de ontvangen wijsheid bij elke beurt wordt uitgedaagd: vraag 'Waarom doe je dit?' 'Wat heeft het voor meerwaarde?' 'Waar wordt het gedaan?' 'Hoe is het gedaan?' 'Wie is of moet verantwoordelijk zijn?' 'Wanneer moet het gebeuren?' 'Is er een andere manier om dit te doen?' Sommige analisten lijken te geloven dat de rol van de data-professional eenvoudig is en bestaat uit het vastleggen van wat het bedrijfs personeel zegt

dat ze willen en het doorgeven van deze 'vereisten' aan een ander team voor levering. Deze aanpak kan echter niet de vruchten plukken van een goede analyse.

Analytisch denken betekent echter niet dat elke situatie eindeloos wordt geanalyseerd, wat aanzienlijke middelen en inspanningen vergt, aangezien een ander belangrijk element van deze competentie is dat degenen die analytisch kunnen denken, in staat zijn om het analyseniveau te beoordelen dat nodig is voor een specifieke situatie. Dit betekent niet dat u kortere wegen moet nemen bij de analyse; Het betekent wel dat we de belangrijkste factoren en de contextuele beperkingen moeten herkennen en ervoor moeten zorgen dat de analyse voldoende is om de zakelijke problemen aan te pakken zonder te proberen alles te analyseren.

### **Gevoel voor bedrijfspolitik**

Politiek bewustzijn is moeilijk te definiëren, maar zeer duidelijk wanneer het wordt waargenomen of aangetroffen. Alternatieve termen zoals 'nous' of 'streetwise' worden vaak gebruikt om de essentie van deze competentie weer te geven. In wezen betekent politiek bewustzijn het vermogen om uit te zoeken wat wel en niet politiek acceptabel is bij het werken binnen een organisatie en met bepaalde individuen - en in staat zijn om de juiste organisatorische hefboomen te gebruiken om dingen voor elkaar te krijgen.

Dit vereist dat analisten de bronnen van macht en informatie binnen de organisatie kennen, begrijpen wat wel of niet acceptabel is en hun aanpak daarop afstemmen. Politiek bewustzijn hebben betekent niet dat je het met iedereen eens bent of de status quo accepteert; Maar het betekent wel vindingrijkheid gebruiken en scherpzinnig zijn om resultaten te behalen, zelfs bij tegenstand.

### **Omgaan met tegenslag**

Business analisten werken vaak binnen contexten waar de op te lossen problemen en de gewenste resultaten onduidelijk of zelfs nog niet bekend zijn. Dit vereist een mindset die kan omgaan met ambiguïteit en aanpasbaar en responsief is.

### **Ondersteunend**

Effectieve ondersteuning vereist uitgebreide interpersoonlijke en organisatorische vaardigheden en is meestal het resultaat van een combinatie van een goede voorbereiding, een duidelijk begrip van de doelstellingen, 'buy-in' van senior stakeholders en het gebruik van nuttige technieken gezien de taak, de deelnemers en de organisatorische context. Een goede facilitator moet beschikken over een aantal van de andere persoonlijke kwaliteiten die elders worden besproken. Er kunnen echter enkele specifieke eigenschappen worden geïdentificeerd die bijzonder relevant zijn bij het faciliteren. Facilitators moeten het vermogen hebben om snel de persoonlijkheidstypes en de voorkeursleerstijl te evalueren van een groep mensen die ze al dan niet eerder hebben ontmoet.

Aanpassingsvermogen is een andere belangrijke kwaliteit, aangezien facilitators in staat moeten zijn om de voortgang van workshops te beoordelen en de aanpak waar nodig aan te passen om op koers te blijven. Het vermogen om snel te reageren en zich aan te passen is vereist omdat een facilitator voortdurend verwerkt wat de deelnemers zeggen, de vragen formuleert die helpen om de discussie vooruit te helpen en beoordeelt of de huidige technieken werken en wat er nog meer in plaats daarvan kan worden geprobeerd. Facilitators moeten ook assertief zijn en in staat zijn om hun autoriteit tijdens de workshop te vestigen.

Een kwestie die vaak naar voren komt in verband met facilitering is of de facilitator een materiedeskundige (SME) moet zijn in het te bespreken onderwerp. Enige kennis is vereist om de discussie te volgen, maar het is niet nodig dat een facilitator een expert is in het onderwerp. In veel situaties kan dit de facilitator helpen om objectief te zijn en vragen te stellen. De rol van de

facilitator is immers om anderen te helpen tot een conclusie te komen, niet om een eigen beslissing op te leggen.

### **Problemen oplossen**

Te vaak klagen data-professionals dat er een oplossing wordt gekozen zonder dat er een waardering is voor het probleem dat moet worden aangepakt. De focus op het begrijpen van het probleem voordat u zich naar een oplossing haast, is een belangrijk principe en waarde propositie van bedrijfsanalyse.

De meerderheid van de data-professionals willen graag zakelijke problemen oplossen, en er zijn veel probleemoplossende technieken en kaders binnen de toolkit voor data-professionals.

### **Relaties bouwen en onderhouden**

Deze competentie betreft het vermogen om goed met mensen om te gaan, binnen een zakelijke context. Sommige mensen bezitten dit vermogen van nature, terwijl anderen zich moeten inspannen om het te ontwikkelen. Hoe dan ook, het is een essentiële vaardigheid voor een data-professional. Data-professionals moeten mensen aanmoedigen om informatie te verstrekken en meningen te delen, en ook om ideeën voor verandering te bespreken. Dit is veel gemakkelijker als de discussies worden gevoerd binnen een goede werkrelatie.

### **Veerkrachtig**

Deze kwaliteit wordt vaak over het hoofd gezien, maar is uiterst belangrijk. Het betekent voldoende zelfvertrouwen hebben om druk te weerstaan, voorstellen uit te dagen, effecten te analyseren en argumenten te ondersteunen. Veerkracht is een kerncompetentie om effectief samen te werken met belanghebbenden en om het brede scala aan situaties te beheren waarmee data-professionals te maken kunnen krijgen.

### **Samenwerken in teams**

Data- en bedrijfsanalyse houdt in dat wordt samengewerkt om informatie te verkrijgen van verschillende belanghebbenden, waaronder collega's uit IT- en veranderdisciplines, leveranciers, bedrijfsmedewerkers en bedrijfsmanagers. Het vermogen om in een team te werken is een essentiële vaardigheid voor een data-professional.

Data-professionals hebben baat bij het begrijpen van de aard van teams en hoe ze zich ontwikkelen. Waardering voor wat een team succesvol maakt, stelt data-professionals in staat om goed samen te werken met collega's en hun analytische vaardigheden te gebruiken om eventuele problemen te identificeren en kansen voor te stellen voor de ontwikkeling van het team.

### **Data behoeften inventariseren en beheren**

Voor de data-professionals zijn stakeholders en concerns essentiële begrippen die bepalend zijn voor de data-architectuur.

Het inventariseren en beschrijven van welke stakeholders en concerns van invloed zijn op de data-architectuur is daarmee een belangrijke activiteit voor de data-professional.

#### **Stakeholders managen**

Stakeholdermanagement is een belangrijk onderdeel van bedrijfsanalyse. Het omvat het vermogen om belanghebbenden te identificeren, het belang van de belanghebbenden te beoordelen, de perspectieven van belanghebbenden te analyseren en strategieën voor stakeholdermanagement te ontwikkelen.

#### **Requirements managen**

Requirements engineering (RE) is een raamwerk van activiteiten die worden toegepast bij het definiëren van vereisten en omvat een scala aan elicatie-, analyse- en modelleringsvaardigheden. Requirements vormen de basis van waaruit bedrijfs- en IT-oplossingen worden ontworpen en ontwikkeld, dus dit is een belangrijk competentiegebied voor data-professionals.

### **Domeinkennis**

De data-professional dient rekening te houden met meerdere domeinen. Binnen deze domeinen worden andere technieken, methoden, tools en frameworks gebruikt. De data-professional dient bekend te zijn met de hulpmiddelen binnen deze domeinen en deze effectief te kunnen toepassen.

#### **Data gedreven werken**

Steeds meer organisaties introduceren data gedreven werken als paradigma voor verandering. Het begeleiden van data gedreven werken is daarmee een centraal werkveld voor de data-professional.

De data-professional dient daarom de ontwikkelingen in het data gedreven werkveld nauw te volgen en deze ontwikkelingen te vertalen naar het werkveld binnen de context van de eigen organisatie.

#### **Business en data analyse**

Data gedreven werken is gericht op veranderingen in de business van een organisatie. Data en business zijn daardoor nauw aan elkaar gerelateerd. Iets wat gechargeerd maar vanuit data

perspectief kan gesteld worden dat een verandering in de business niet mogelijk is zonder veranderingen in de data.

Daardoor is er een nauwe relatie tussen business en data en daardoor een nauwe samenwerking tussen business analisten en data-professionals.

Met name het beschrijven van de data op conceptueel niveau is in deze een data-ontwerpproduct dat de verbinding legt tussen business en data. Daarnaast ontstaan vanuit de business doelen en concerns die in de data-documentatie worden uitgewerkt in kaders zoals data-principes.

### **Enterprise architectuur**

Organisaties die onder architectuur werken noemen dat veelal enterprise architectuur. Hiermee wordt enterprise architectuur als een container wordt beschouwd. Binnen enterprise architectuur worden meerdere domeinen onderkend zoals infrastructuur, applicatie- bedrijfs en zeker ook data-architectuur.

Daarmee ontstaat een behoefte dat een aantal aspecten binnen deze aspectarchitecturen te standaardiseren en generaliseren.

Vanuit de verschillende aspecten in de enterprise architectuur dienen de architecten gezamenlijk de producten waaronder kaders te definiëren.

Zo zal er nauw samengewerkt worden in het bepalen van de architectuurtalen en het metamodel van de enterprise architectuur. Dit is inclusief het gebruik van de tooling en het inzetten van bijvoorbeeld architectuur repositories voor een geïntegreerde inrichting van de aspectarchitectuur.

### **Interactieve werkvormen toepassen**

Data-professionals werken nauw samen met veel verschillende soorten stakeholders in de organisatie. Gezamenlijk werken zij producten en modellen uit rond data gedreven werken.

Deze stakeholders leveren niet alleen input aan de datagedreven toepassing zij dienen ook betrokken te worden bij de uitwerking van de business case. Daarnaast zullen zij in een later stadium de rol van ambassadeur gaan vervullen bij het introduceren van data gedreven werken.

Om die samenwerking te optimaliseren en de stakeholders te enthousiasmeren voor het bijdragen aan de data toepassingen kunnen verschillende interactieve werkvormen ingezet worden. Deze interactieve werkvormen dienen aan te sluiten bij waar deze stakeholders het gevoel hebben een zinvolle bijdrage te leveren aan de datagedreven toepassingen. De keuze van de juiste werkvormen dienen hieraan bij te dragen.

### **Workshops verzorgen**

Naast het afnemen van interviews en het werken met vragenlijsten zijn interactieve workshops met stakeholders een krachtige werkwijze om de concerns en requirements van stakeholders in beeld te krijgen

Met name in situaties waar de stakeholders nog weinig ervaring hebben met data-management en/of data-gedreven werken is de inzet van intensievere werkvormen in de vorm van verschillende soorten workshops een hulpmiddel dat de data-professional kan inzetten.

Kennis van de organisatiestructuren -cultuur en bekend met de data context kan de data-professional helpen in het bepalen welke interactieve workshops succesvol ingezet kunnen worden bij het uitwerken van datagedreven producten.

### **Interviews afnemen**

Bij het introduceren van data gedreven werken zullen vele soorten stakeholders betrokken zijn bij het introduceren van data gedreven toepassingen onder architectuur.

Om hun wensen en behoeften op gestructureerde wijze in kaart te brengen dienen adequate interviewtechnieken te worden toegepast. Vormen als groepsinterviews, een op een interviews en eventueel weg gebaseerde schriftelijke interviews zijn hiervoor beschikbaar.

Echter welke interview vorm past het beste in welke context is een criterium dat een data-professional bij het afnemen van interviews dient toe te passen. De data-professional dient daarom bekend te zijn welke interviewvormen beschikbaar zijn en welke interviewvorm in welke situatie het meest geschikt is.

### **Vragenlijsten opstellen en verwerken**

Bij het uitwerken van de data-toepassingen, met name bij het in kaart brengen van de requirements en concerns van stakeholders kunnen op meerdere wijzen de behoeften van betrokkenen in kaart worden gebracht.

Het opstellen en verwerken van vragenlijsten is hierbij een veel toegepaste werkwijze. Daartoe dient de data-professional te beschikken over hulpmiddelen en vaardigheden om dergelijke vragenlijsten te kunnen opstellen en te verwerken. Hiermee heeft de data-professional een krachtig hulpmiddel om een duidelijk beeld te krijgen van de concerns en op basis hiervan een datagedreven toepassing uit te werken.

### **Leveranciers management**

Aangezien de meeste dataveranderingsprojecten de ontwikkeling of aanschaf van

softwaretoepassingen vereisen, is een algemeen begrip van technologie, technologische ontwikkelingen en benaderingen van softwareontwikkeling noodzakelijk, zodat data-professionals zinvol kunnen communiceren met hun technologiegerichte collega's en hun rol en bijdrage aan de oplossing en het ontwikkelingsproces kunnen waarderen.

De mate waarin data-professionals technische kennis nodig hebben, hangt af van de aard van het analysewerk dat wordt uitgevoerd. De belangrijkste vereiste is dat de data-professional het potentieel van technologie en de benaderingen en termen die door technische specialisten worden gebruikt, begrijpt. Enkele van de belangrijkste gebieden die data-professionals moeten begrijpen, worden hieronder opgesomd:

- Trends en ontwikkelingen zoals AI, robotic process automation (RPA), big data, software as a service (SaaS), visualisatie, mobiele technologieën, en hoe deze impact hebben op organisaties en de potentie die ze bieden voor nieuwe of verbeterde producten of diensten.
- Technische infrastructuurcomponenten zoals besturingssystemen, applicatiesoftware, hardware, netwerken, cloud computing.
- Levenscycli van systeemontwikkeling (SDLC's) en benaderingen zoals het 'V'-model en het uniforme proces.
- Benaderingen voor systeemmodellering zoals de UML. Agile ontwikkelingsbenaderingen zoals DSDM en Scrum.
- De relatieve voor- en nadelen van het ontwikkelen van software in plaats van het kopen van kant-en-klare softwareproducten.
- Organisatiestructuren Veel bedrijfsveranderingsprojecten omvatten het tot op zekere hoogte herstructureren van divisies of teams om overdrachten te verwijderen, taken te centraliseren of de klantenservice te verbeteren. Om deze redenen is het belangrijk dat een data-professional een goed begrip heeft van de verschillende organisatiestructuren die zich kunnen voordoen - functioneel, project, matrix, plat, virtueel - en van hun relatieve sterke en zwakke punten.
- Beheer van leveranciers Veel organisaties maken gebruik van externe leveranciers om hun IT-systemen te leveren, hetzij op ad-hocbasis, hetzij via een uitgebreidere outsourcingregeling, die hele bedrijfsprocessen of
- zelfs een hele bedrijfsfunctie kan omvatten. Veel organisaties hebben bijvoorbeeld hun
- salarisadministratieprocessen meerdere jaren uitbesteed, maar sommige hebben dit nu uitgebreid tot een groot deel van het personeelswerk (HR), van werving tot het bijhouden van gegevens.

Selecteren en contracteren van leveranciers valt meestal binnen het domein van de inkoopfunctie. Voor sommige outsourcingcontracten kan de data-professional echter betrokken zijn om ervoor te zorgen dat de bedrijfsprocessen en -systemen efficiënt blijven werken. Dit vereist dat data-professionals een breed begrip hebben van inkoop- en leveranciersbeheerprocessen.

Data-professionals moeten op zijn minst op de hoogte zijn van de verschillende contractuele regelingen die beschikbaar zijn, met name:

- Tijd en materiaal: Wanneer de gecontracteerde partij wordt betaald op basis van de gewerkte tijd en de te leveren prestaties die zijn geleverd; het tijdselement heeft geen betrekking op de verstreken tijd van het project, maar op de hoeveelheid inspanning die is geleverd.
- Levering tegen een vaste prijs: waarbij de gecontracteerde partij de prijs ontvangt die is overeengekomen voor de levering van het werk in overeenstemming met de oorspronkelijke specificatie.
- Risico en beloning: Wanneer de gecontracteerde partij ermee heeft ingestemd om een deel of het volledige risico van het project te dragen. Bijvoorbeeld door middelen te investeren zoals personeelstijd, materialen of kantoorruimte, maar waarbij de potentiële beloningen groter zijn dan bij andere contractuele regelingen.

Data-professionals moeten in staat zijn om met leveranciers in contact te komen om ervoor te zorgen dat ze hun diensten effectief leveren. Dit vereist persoonlijke kwaliteiten zoals communicatie en het opbouwen van relaties, die eerder zijn besproken.

### **Interne leveranciers**

Voor datagedreven werken zijn er naast externe ook interne leveranciers. Deze interne leveranciers leveren bijvoorbeeld datasets aan die gebruikt kunnen worden binnen datagedreven toepassingen. Daarnaast zijn er intern veelal al dataplatformen en tools aanwezig die door interne leveranciers, bijvoorbeeld database beheerders of applicatie beheerders voor datatoepassingen.

Vanuit efficiëntie zijn er meestal kaders om hergebruik van reeds bestaande dataproducten en omgevingen te hergebruiken. Daarom dienen interne leveranciers en data-professionals hierin nauw samen te werken en gezamenlijk zorg te dragen dat deze kaders worden nageleefd.

### **Externe leveranciers**

De data-professional zal te maken krijgen met een breed palet van externe leveranciers. Leveranciers van data gerelateerde consultancy, data-sets, dataplatformen en tooling maar ook leveranciers van cloudgebaseerde dataplatformen zullen data-professionals benaderen met hun producten en diensten.

De data-professional dient zich bewust te zijn van de vaardigheden die externe leveranciers hebben om hun dataproducten onder de aandacht te brengen van de data stakeholders. Externe leveranciers zullen er alles aan doen om deze stakeholders, inclusief de data-professional zelf, om hen een positief beeld te geven van de door hen geleverde producten en diensten. Dit vraagt specifieke vaardigheden van de data-professional om de werkelijke kenmerken van de producten van leveranciers op waarde te schatten voor de stakeholders binnen de eigen organisatie, en hen hiervan te overtuigen.

### **Over data adviseren**

Data is een werkveld gericht op de verandering in organisaties en daarbuiten en het beschrijven van deze veranderingen. Vandaar dat vanuit elke data-professional de stakeholders geadviseerd worden. Met name stakeholders op strategisch niveau en de gebruikers van datagedreven toepassingen zullen door data professionals geadviseerd worden.

Dit stelt eisen aan de adviesvaardigheden van de data-professional.

Daarnaast zullen de verschillende data-professionals in een organisatie om ervoor zorg te dragen dat er een consistent en gelijklopend advies wordt gegeven aan de strategische stakeholders vanuit de verschillende data deelgebieden.

### **Governance adviseren**

De governance rollen en met name de rollen gericht op data-governance zijn verantwoordelijk voor het beleid rond data in de organisatie. Het zijn rollen die sterk in opkomst zijn, met name door data gedreven werken en het introduceren van data management.

Daarmee zijn deze data-governance rollen belangrijke stakeholders voor de data-professionals. Data-governance en data-professionals dienen nauw met elkaar samen te werken om ervoor te zorgen dat de data-strategie wordt geïntroduceerd in de organisatie en vertaald wordt naar beleid, oplossingen en kaders om sturing te geven aan de verandering in de organisatie naar data gedreven werken.

### **Data werkveld adviseren**

Rond het werkveld kent de data-professional veel stakeholders. Niet alleen op strategisch niveau ook op tactisch en operationeel niveau. Daarbinnen is een nieuw werkveld ontstaan, waar data-engineers, data beheerders en data-analisten en -scientist actief zijn.

Deze rollen in het datawerkveld hebben specifieke requirements maar brengen ook specifieke kennis in rond het data werkveld.

Daarom dient de data-professional de verschillende rollen in het data werkveld te adviseren over hoe data ingezet kan worden in hun werkveld maar daarbij wel rekening houden met de kaders en requirements die vanuit stakeholders uit andere domeinen een rol spelen.

### **Management ondersteunen**

Data-professionals adviseren meerdere soorten stakeholders binnen organisaties maar doen dat ook vanuit verschillende gezichtspunten.

Bij organisaties die data-gedreven willen werken of data-management willen introduceren is het management van de organisatie een relevante stakeholder.

Bij het uitwerken van de organisatie strategie zullen aspecten, requirements en eisen vanuit data-gedreven perspectief een rol spelen die de data-professional gebruikt voor het adviseren van het management.

### **Ontwerp en documentatie managen**

Data-professionals ontwikkelen meerdere ontwerpen en documenten. Zie de uitwerking met welke soorten producten ontwikkeld worden door data-professionals. Door de verschillende vormen van producten maar ook de verschillende publicatieritmes van ontwerpen en documentatie is het managen van deze producten een belangrijke capability.

Binnen de projecten en programma's die veranderingen introduceren in de organisatie worden er door de data-professionaal solution architecturen en ontwerpen gemaakt die kaders stellen aan de veranderingen in de organisatie. Echter is de behoefte aan organisatieveranderingen hoog dan is kan ontwikkelen van deze ontwerpen onder druk komen te staan door tijdgebrek.

Daarnaast worden steeds meer agile technieken toegepast binnen de organisatie, ook dat kan druk veroorzaken op de werkzaamheden en de agenda van data-professionals.

### **Beschrijvende ontwerpen beheren**

De beschrijvende ontwerpen beschrijven het huidige en gewenste, en eventueel tussenliggende, datalandschap voor een organisatie.

Voor de beschrijvende ontwerp worden veelal gestandaardiseerde architectuurmodelleertalen gebruikt zoals ArchiMate, BPMN, UML en ER. Deze talen worden uitgewerkt op basis van een hybride modelleerwijze waarbij de architect een combinatie van modelleertalen combineert om op basis daarvan een beschrijvende data documentatie te laten.

Aangezien het ontwikkelen van een beschrijvende documentatie een continu proces is dient het managen van de modellen en het bewaken van de kwaliteit van de modellen bewaakt te worden.

### **Metadata register managen**

Bij het uitwerken van een datatoepassing en een dataplatform ontstaan al snel allerlei ontwerpen, documenten en modellen die nauw met elkaar samenhangen. Hierdoor ontstaat een situatie die dusdanig complex is dat de inhoud van de toepassingen en het platform gemanaged moet worden.

Het managen van de documentatie bestaande uit de modellen en producten kan gedaan worden door inzet van een metadata register of repository te introduceren. Dit is een informatiesysteem dat zorgdraagt voor het managen van de documentatie inhoud en het bewaken van de kwaliteit.

Veelal wordt een architectuur register gebruikt door alle architecten binnen een organisatie. Zij dienen daarom nauw samen te werken om zorg te dragen dat het architectuur register een

gezamenlijk ingericht en gemanaged informatiesysteem wordt voor de gezamenlijke architectuur.

### **Kaderstellende architecturen managen**

Voor het sturen van veranderingen in de organisatie in projecten en programma's gericht op data gerelateerde werkvelden is de kaderstellende architectuur een belangrijk product.

Welke kadervormen ingezet worden en de inhoud van deze kaders dienen daarom gemanaged te worden. Bij het uitwerken van de kaders is het uitgangspunt dat deze weinig aan verandering onderhevig zijn. Echter organisaties veranderen continue en dat houdt in dat de kaderstellende architectuur gemanaged dient te worden zodat de kaderstellende architectuur de juiste kaders stelt aan de veranderingen in de organisatie.

### **Modelleren**

Het werkveld van data is complex en deze complex kan inzichtelijk gemaakt worden door modellen van het werkveld op te stellen.

De data-professional zal daartoe meerdere modelleertechnieken toepassen om hiermee de verschillende stakeholders met modellen te ondersteunen op basis waarvan deze stakeholders inzicht krijgen van deze complexiteit en op basis daarvan een gefundeerde beslissing kunnen nemen.

#### **Data modelleren**

Het analyseren van de gegevens die binnen een bedrijfssysteem zijn opgeslagen en gebruikt, biedt waardevolle inzichten in hoe dat systeem werkt. Een datamodel weerspiegelt de informatiebehoefte van de organisatie en de bedrijfsregels die inherent zijn aan de structuur van de data. Er zijn verschillende technieken die worden gebruikt om gegevens te modelleren.

#### **Data toepassing modelleren**

De datatoepassing modelleren is een kernactiviteit van de data-professional. De modellen zijn een vereenvoudiging van de complexe inrichting van data binnen organisaties. Door deze vereenvoudiging op basis van gestructureerde (architectuur) modelleertechnieken krijgt de data-professional inzicht in de situatie en kan hij de situatie analyseren. Op basis hiervan kan hij oplossingen beschrijven die bijdragen aan een beheerbaar en inzetbaar datalandschap voor de organisatie en de omgeving.

#### **Business case visualiseren**

Naast de eerder beschreven persoonlijke kwaliteit 'Faciliteren', zijn er technieken die data-professionals kunnen gebruiken om discussie te stimuleren, ideeën te genereren en resultaten weer te geven tijdens een vergadering, workshop of focusgroep. Deze technieken omvatten: mindmapping, open space-technologie, brainstorming, cartoons of houtkoolschetsen.

Bij het faciliteren wordt vaak gebruik gemaakt van visualisatietechnieken. Ze zijn snel te begrijpen en eenvoudig uit te leggen, en helpen om deelnemers aan de workshop te betrekken, of het nu gaat om het verkennen van zakelijke problemen, strategische keuzes of vereisten. Visuele representaties van informatie kunnen low-fidelity tekeningen zijn of kunnen worden geproduceerd met behulp van geautomatiseerde tools, waarbij het mogelijk is om verschillende scenario's te modelleren zonder uitgebreid opnieuw te tekenen.

### **Data toepassing ontwikkelen**

Ontwikkelen, testen en uitrollen van softwaremodules, scripts en algoritmen in datagedreven toepassingen

#### **Data algoritmen toepassen en ontwikkelen**

Toepassen van algoritmen voor data analyse, data-science inclusief het ontwikkelen van scripts voor het verhogen van de kwaliteit van de datasets toegepast in de toepassing. Rond toepassen wordt ook verstaan de vaardigheid om de juiste algoritmen te selecteren op basis van de business case en de datasets

#### **Data software ontwikkelen**

Ontwikkelen van de software in de vorm van scripts en executables, inclusief het gebruik van libraries, scripting talen en compiler voor het ontwikkelen van de data gedreven toepassingen. De toepassing dient data te ontsluiten, transformeren maar ook het ontwikkelen van data visualisaties bijvoorbeeld in dashboards

#### **Data toepassingen testen**

Testen van de gehele datagedreven toepassing inclusief de ontsluiting van databronnen via APIs en connectoren, de transformatie en integratiestappen. Testen dienen daarbij herhaalbaar en schaalbaar zijn en rekening houden met de inzet van toepassingen in product situaties tov ontwikkel/testsituaties.

#### **Data toepassingen uitrollen**

Uitrollen van de toepassing in de organisatie en beschikbaar stellen aan de (eind)gebruikers.

Uitrollen omvat ook het zorgdragen van releases en het integreren van de datagedreven toepassing in de product situatie van de organisatie.

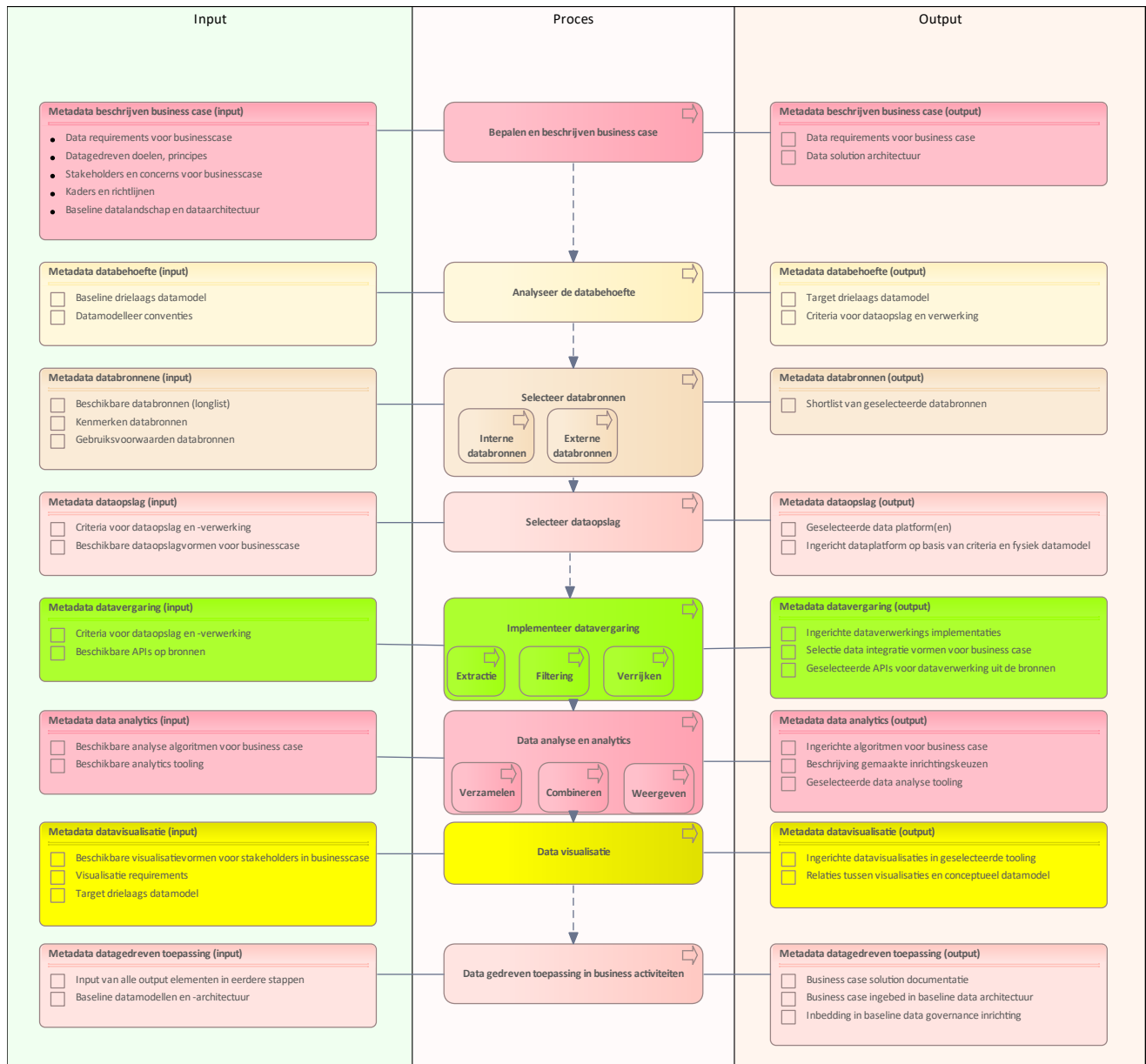
#### **ETL/ELT scripts ontwikkelen**

Ontwikkelen van script voor het ontsluiten van data uit de verschillende databronnen, transformeren naar een analyseerbare dataset bijvoorbeeld in een datawarehouse omgeving of een omgeving voor dataanalyse en -science bijvoorbeeld in datalakes.

# Metadata bij data gedreven werken

In dit hoofdstuk gaan we kijken welke metadata relevant is bij iedere stap in een datagedreven project. Er is hierbij een indeling in de metadata die benodigd is om deze projectstappen uit te voeren en welke metadata geproduceerd wordt in iedere processtap. Dit wordt deels gemodelleerd in ArchiMate voor de bedrijfsprocesstappen. Daarnaast met een aantal checklist items voor de input en output.

## Metadata Data gedreven project proces



Dit diagram toont de stappen die ingezet kunnen worden om een data gedreven project in projectstappen in te delen. Daarnaast voegen we hier aan de processtappen een aantal checklists toe met metadata onderwerpen die bij de beschreven projectactiviteit relevant zijn. In dit diagram erkennen we input en output metadata voor de projectstappen inclusief de daarbij horende sub projectactiviteiten. In dit diagram zie je metadata die via een input beschrijving (linker kolom) geconsumeerd worden door de procesactiviteit. Daarnaast zie je welke metadata binnen een procesactiviteit geproduceerd wordt als output (rechterkolom). Veel van de metadata die in een proces activiteit geproduceerd wordt zal in een latere proces activiteit ook weer geconsumeerd worden. Dat is in dit diagram model niet apart aangegeven. We gaan uit van een metadataregister zodat alle geproduceerde metadata beschikbaar is binnen dit project en aanpalende metadata gerelateerde activiteiten.

Input	Processtap	Output
<p><b>Datagedreven doelen, principes</b>  Bij een organisatie waar data gedreven werken en data management een bepaalde mate van volwassenheid heeft zal er al een data strategie zijn uitgewerkt in data gedreven doelen.  Zijn de data gedreven doelen duidelijk dan zijn er in veel gevallen ook al data architectuur principes uitgewerkt op basis van de data gedreven doelen. Deze data architectuur principes zijn dan kaderstellend aan de business case die hier beschreven wordt</p> <p><b>Stakeholders en concerns voor businesscase</b>  Bij een business case dient bepaald te worden wat de stakeholders zijn en welke concerns zij hebben. Vaak zijn deze stakeholders en concerns in een referentie architectuur al uitgewerkt als een lijst van mogelijke stakeholders en concerns. Deze kunnen dan als checklist cq hulpmiddel worden gebruikt om te bepalen welke concerns en stakeholders voor deze business case. Dit helpt ons om adequaat te bepalen welke requirements gelden voor de uit te werken business case</p> <p><b>Data requirements voor businesscase</b>  Lijst van requirements van de diverse stakeholders voor deze business case. Bij een omvangrijk datagedreven traject zullen de stakeholders en de bijbehorende concerns uitgewerkt worden in een model waarin de relaties tussen de stakeholders en de requirements en de relatie onderling uitgewerkt worden in een requirements model.</p> <p><b>Kaders en richtlijnen</b>  Naast de data gedreven doelen, concerns en</p>	<p><b>Bepalen en beschrijven business case</b>  Deze stap beschrijft de business case. Dit is de eerste stap in het proces en heeft een belangrijke relatie met de data gedreven toepassing die ontwikkeld gaat worden.  Meestal wordt de business case opgesteld door de data governance rollen in nauwe samenwerking met de data-professional.</p>	<p><b>Data requirements voor business case</b>  Bij het uitwerken van deze specifieke business case worden een aantal requirements uitgewerkt. Daarmee zijn deze requirements in een latere processtap bepalend voor de inrichting van het data gedreven project. Bij organisaties waar data architectuur principes is het ook mogelijk dat de implicaties van de data architectuur principe voor deze business case uitgewerkt worden.</p> <p><b>Data solution architectuur</b>  Bij het uitwerken van de business case voor dit data gedreven project zal een beschrijving gemaakt worden in de vorm van een solution architectuur. Hierin zijn enerzijds alle relevante stakeholders, concerns, principes en requirements uitgewerkt en in een aantal gevallen wordt hier ook al een (gedetailleerde) beschrijving gegeven van de inrichting. Meestal is een agile aanpak kenmerkend voor data gedreven business cases. In dat geval zal de uitwerking.</p>

<p>requirements kunnen er aanvullende kaders en richtlijnen zijn die gelden. Denk bijvoorbeeld aan de webrichtlijnen of de lijst van open standaarden die beiden bij de Nederlandse overheid als kaders en richtlijnen gelden.</p> <p><b>Baseline datalandschap en data architectuur</b>  Vaak zal een data gedreven project gestart worden in een organisatie die al onderweg is met data gedreven werken. Dan is er geen green field situatie meer voor deze business case. De keuzes en geselecteerde tooling zal dan als startpunt gelden voor dit data gedreven project.</p>		
<p><b>Metadata databehoeft (input)</b>  Bij het uitwerken van de databehoeft zullen de requirements en principes zoals beschreven bij voorgaande proces activiteit als input dienen. Specifieke input in deze stap zijn:</p> <p><b>Baseline drielaags datamodel</b>  Voor de databehoeft zal een datamodel uitgewerkt worden, Veelal zal dit gedaan worden in de vorm van een drielaags datamodel. Is data gedreven werken al ingebed in de organisatie dat zijn er waarschijnlijk al datamodellen uitgewerkt voor andere business cases. Deze modellen kunnen als input dienen voor deze proces activiteit. Want de reeds aanwezige modellen zijn een mooi startpunt voor het drielaags datamodel dat voor deze business case uitgewerkt wordt.</p> <p><b>Datamodelleerconventies</b>  Soms zijn bij organisaties al modelleer- en naamegevingsconventies aanwezig voor datamodelleren. Deze dienen als kaders voor de uit te werken drielaags datamodellen.</p>	<p><b>Analyseer de databehoeft</b>  Op basis van de business case kan er een inschatting gemaakt worden van de data behoeft. Dat wordt veelal gedaan door een eenvoudig datamodel op te stellen van de databehoeft voor de business case. Bijvoorbeeld een combinatie van een conceptueel en logisch datamodel van de databehoeft.</p>	<p><b>Target drielaags datamodel</b>  Output van deze procesactiviteit is het drielaagsmodel dat de databehoeft voor deze business case beschrijft. Dit model dient opgenomen te worden in het metadata register zodat ook dit target datamodel voor een latere business case als baseline datamodel geldt.</p> <p><b>Criteria voor dataopslag en verwerking</b>  Op basis van de uitwerking van het drielaagsdatamodel voor deze business case kunnen aanvullende requirements en criteria gelden voor dataopslag en -verwerking. Denk bijvoorbeeld aan bijzondere dataopslag vormen zoals NoSql databases, gedistribueerde dataopslag en verwerking en criteria voor ETL of API inrichting.</p> <p><b>Metadata databronnen (output)</b>  <b>Shortlist van geselecteerde databronnen</b>  Op basis van de inventarisatie van bronnen zal er een selectie gemaakt worden van een aantal databronnen die ontsloten gaan worden in de uitwerking van deze business case. Deze shortlist gaat</p>

		<p>onderdeel uitmaken van de solution architectuur. Daarnaast kan deze shortlist opgenomen worden in het metadata register voor toekomstige data gedreven business cases.</p>
<p><b>Beschikbare databronnen (longlist)</b>  Op basis van het target drielaags datamodel voor de business case kan gezocht worden naar de databronnen waar de inhoud van de business case opgehaald kan worden.  Zeker in organisaties waar al eerder datagedreven business cases zijn uitgewerkt zullen al inventarisaties van databronnen gedaan zijn. Deze inventarisaties zijn eengoede bron van inspiratie voor het selecteren van databronnen voor een business case. Bijkomend voordeel is dat volop hergebruik van databronnen mogelijk wordt. Ook hierbij geldt weer dat een metadata register een belangrijk hulpmiddel is om deze projectactiviteit te faciliteren en daarmee minder complex te maken.</p> <p><b>Kenmerken databronnen</b>  Iedere databron heeft een aantal kenmerken die bij latere projectactiviteiten relevant zijn. Denk bijvoorbeeld aan de structuur van de data (laag tot hoog gestructureerd). API's en ontsluitingsvormen, datakwaliteitsaspecten van de data uit de verschillende databronnen.</p> <p><b>Gebruiksvoorwaarden databronnen</b>  De bronhouder van een databron kan aanvullende eisen stellen aan het gebruik van databronnen. Denk aan security, privacy, financiële en juridische aspecten van het gebruik van de data uit een bron.</p>	<p><b>Selecteer databronnen</b>  Databronnen maken feitelijk geen onderdeel uit van een Data gedreven toepassing. Databronnen zijn de grondstof waaruit de structurering tot informatie plaatsvindt. Daarom is het van belang om goed na te denken wat de juiste bron is. Dit wordt gebaseerd op de analyse van de databehoeftes voor deze business case.</p> <p><b>Interne databronnen</b>  Voor data gedreven toepassingen zal een deel van de data afkomstig zijn vanuit de eigen organisatie. Bijvoorbeeld over de activiteiten van de organisatie en hoe de organisatie de waarde levert binnen het vervaardigen van hun producten en diensten.</p> <p><b>Externe databronnen</b>  Als er intern geen databronnen zijn of als de interne datasets onvoldoende kwaliteit hebben voor de databehoeftes kan gekozen worden voor externe (betaalde) databronnen.</p>	<p><b>Shortlist van geselecteerde databronnen</b>  Op basis van de inventarisatie van bronnen zal er een selectie gemaakt worden van een aantal databronnen die ontsloten gaan worden in de uitwerking van deze business case. Deze shortlist gaat onderdeel uitmaken van de solution architectuur. Daarnaast kan deze shortlist opgenomen worden in het metadata register voor toekomstige data gedreven business cases.</p>

<p><b>Criteria voor dataopslag</b> Op basis van de geselecteerde databronnen, de requirements gelden criteria voor de opslag van data binnen de data gedreven toepassing. Denk hierbij aan de opslag van halffabrikaten van de data maar ook de opslag van de data ten behoeve van de uiteindelijke datagedreven toepassing binnen de laatste activiteit van dit proces.</p> <p><b>Beschikbare dataopslagvormen voor businesscase</b> De halffabrikaten en de data van de datagedreven toepassing zullen persistent gemaakt moeten worden in enigerlei vorm. Denk bijvoorbeeld aan een DWH, een datalake of een cloudbaseerd platform maar wellicht ook data queues behoren tot de mogelijkheden. Ook hierbij zijn er meerdere opties van opslagvormen mogelijk. Veelal is dit reeds uitgewerkt in de vorm van een dataplatform of een data architectuur waarin de verschillende dataopslagvormen in welke context toegepast dienen te worden.</p>	<p><b>Selecteer dataopslag</b> Kies voor een opslag medium, wijze (zoals rdbms, nosql, newsql of polyglot etc).</p>	<p><b>Geselecteerde data platform(en)</b> Beschrijving van de voor deze business case geselecteerde dataplatformen met eventueel extra beschrijvingen. Ook dit wordt opgenomen in het metadata register.</p> <p><b>Ingericht dataplatform op basis van criteria en fysiek datamodel</b> Beschrijving van de inrichting van de geselecteerde dataplatformen desgewenst met een beschrijving van de criteria die bepalend zijn voor de inrichting. Op basis hiervan zal veelal het fysieke datamodel dat reeds is uitgewerkt in het drielaags datamodel verfijnd worden naar de fysieke inrichting van het dataplatform. Uiteindelijk is het eindresultaat een ingericht dataplatform. De beschrijving hiervan kan desgewenst opgenomen worden in de solution architectuur.</p>
<p><b>Criteria voor dataverwerking</b> Op basis van de geselecteerde databronnen, de requirements gelden criteria voor de verwerking van data binnen de data gedreven toepassing. Want de databronnen zullen een datamodel hebben dat niet overeenkomen met het fysieke datamodel van de datagedreven toepassing. Daarom dient dataverwerking en transformatie geïmplementeerd te worden. Veelal zijn hiervoor al een aantal criteria en checklist opgesteld inclusief geselecteerde tooling voor data transformatie</p> <p><b>Beschikbare APIs op bronnen</b> De databronnen bieden data veelal aal via APIs. Dat zijn</p>	<p><b>Implementeer dataverwerking</b> Als bronnen zijn geselecteerd en gefilterd dient de verwerking plaats te vinden. Verwerking is het omzetten van de ruwe data naar een structuur waarop data analyse mogelijk is. Vanwege het volume en de velocity is bij big data gedistribueerde verwerking veelal noodzakelijk Er is een nauwe relatie met data opslag en data integratie.</p> <p><b>Extractie</b> Extraheren van patronen in de data om inzichten te krijgen die wenselijk zijn om de toepassing succesvol te realiseren.</p> <p><b>Filtering</b> Filteren van die data</p>	<p><b>Metadata dataverwerking (output)</b> Ingerichte dataverwerkings implementaties In deze project activiteit vinden veel technische activiteiten plaats voor de inrichting van de dataverwerking. Deze inrichting is een fysieke output van deze proces activiteit</p> <p><b>Selectie data integratie vormen voor business case</b> Voor het ontsluiten van bronnen en het transformeren van de data en het bieden van APIs voor de uiteindelijke datagedreven toepassing zal er een veelheid van integratievormen aanwezig zijn. Vaak is hiervoor ook een data integratie architectuur</p>

<p>bijvoorbeeld APIs voor berichtenverkeer zoals binnen XML en JSON. Maar ook APIs voor toegang tot databases via connectoren met eigen protocollen behoren tot de mogelijkheden.</p>	<p>elementen die in de databronnen aanwezig zijn maar niet nodig zijn voor deze datagedreven toepassing.</p> <p><b>Verrijken</b> Verrijken van de data door het uitbreiden van de data entiteiten zodat meer zeggingskracht ontstaat in de data.</p>	<p>aanwezig. Er dient echter een beschrijving te komen welke integratievorm gekozen is en op basis van welke requirements. De beschrijving van de gekozen integratievorm is relevant voor in de solution architectuur en voor in het metadata register.</p> <p><b>Geselecteerde APIs voor dataverwerking uit de bronnen</b> Voor het ontsluiten van bronnen via APIs en het bieden van APIs voor de uiteindelijke datagedreven toepassing zal er een veelheid van APIs mogelijk zijn. Vaak is hiervoor ook een data integratie architectuur aanwezig. Er dient echter een beschrijving te komen welke APIs gekozen zijn en op basis van welke requirements. Deze lijst van APIs en criteria zijn relevant voor in de solution architectuur en voor in het metadata register.</p>
<p><b>Beschikbare analyse algoritmen voor business case</b> Voor de uitwerking van de business case dienen een aantal beschikbare analyse algoritmen geselecteerd te worden. Hierbij rekeninghoudend met het soort analyse, de datakwaliteiten, de hoeveelheid data en de betrouwbaarheid van de uitkomsten. Ook hierbij kan de inzet van een metadata register behulpzaam zijn waarin een inventarisatie van analyse methoden voor alle data gedreven business cases een collectie kan vormen van toe te passen data-analyse algoritmen</p> <p><b>Beschikbare analytics tooling</b> Lijst van mogelijke tools en platformen om data analyse uit te voeren ten behoeve van de business case. Op basis van andere business cases zal hergebruik ontstaan van analyse tooling en kan</p>	<p><b>Data-analyse en analytics</b> Data-analyse is het gericht zoeken naar (statistische) verbanden in gegevensverzamelingen met als doel profielen op te stellen voor wetenschappelijk, journalistiek of commercieel gebruik. Zo'n verzameling gegevens kan gevormd worden door gebeurtenissen in een praktijksituatie te registreren (aankoopgedrag van consumenten, symptomen bij patiënten, et cetera) of door de resultaten van eerder uitgevoerde wetenschappelijke onderzoeken met elkaar te vergelijken en te herinterpreteren.</p> <p><b>Combineren</b> Combineren of joinen van verschillende datasets om analyse mogelijk te maken.</p> <p><b>Verzamelen</b> Verzamelen van verschillende datasets die aanwezig zijn en invulling kunnen geven aan de</p>	<p><b>Ingerichte algoritmen voor business case</b> Implementatie van de algoritmen voor deze business case is een technische inrichting in het platform</p> <p><b>Geselecteerde data-analyse tooling</b> Data analyse tooling is geselecteerd en inrichting voor gebruik voor deze, en desgewenst andere, business cases.</p> <p><b>Beschrijving gemaakte inrichtingskeuzen</b> Uitwerking van de gemaakte keuzes rond de data-analyse algoritmen op basis van de requirements van de betrokken stakeholders. Deze beschrijving zal opgenomen worden in de solution architectuur voor deze business case en worden opgenomen in het metadata register.</p>

<p>desgewenst de analyse activiteit ingebed worden in het data platform.</p>	<p>data analyse.</p> <p><b>Weergeven</b>  Visualisaties van resultaten, tussenresultaten en verkennende analyse visualisatie in relatie tot de datakwaliteiten. Weergeven is gericht op de doelgroep van data analisten.</p>	
<p><b>Beschikbare visualisatievormen voor stakeholders in businesscase</b>  Visualisatievormen zijn er in allerlei soorten en maten. En sommige visualisatievormen zijn binnen een business case meer geschikt dan de ander. Dit is mede afhankelijk van de doelgroep van de visualisaties. Een overzicht van mogelijke visualisatievormen per doelgroep en context is hierbij relevant voor de te selecteren visualisatievormen voor deze business case.</p> <p><b>Visualisatie requirements</b>  Dit is een subset van de requirements die bij de beschrijving van de business case zijn verzameld en worden hier desgewenst specifiek gemaakt voor de te kiezen visualisatievormen. Desgewenst kunnen eerder bepaalde visualisatie requirements van eerdere business cases hier hergebruikt worden.</p>	<p><b>Data visualisatie</b>  Visualiseren van de data analyse bijvoorbeeld in de vorm van rapportages, dashboards met grafieken en combinaties daarvan. Als laatste kunnen bij de visualisaties ook user interfaces horen voor de gebruikers van de data gedreven toepassing die in het project ontwikkeld wordt. Bijvoorbeeld in apps en websites ter ondersteuning van eindgebruikers.</p>	<p><b>Ingerichte datavisualisaties in geselecteerde tooling</b>  Data visualisaties zijn geïmplementeerd in de geselecteerde tooling binnen het aanwezige data platform voor de data gedreven organisatie of deze business case. De gemaakte keuzes van de visualisatievormen worden opgenomen in de solution architectuur.</p> <p><b>Relaties tussen visualisaties en conceptueel datamodel</b>  Dit is wellicht een verrassende output. De visualisaties worden uiteindelijk aangeboden aan de gebruikers van de visualisaties. Hierbij kan onduidelijkheid ontstaan over de definities van de concepten in de visualisaties. De gebruikers dienen daarom inzage te hebben in het concept datamodel vanuit de data visualisaties (zoals bijvoorbeeld KPI dashboards).</p>
<p><b>Input van alle output elementen in eerdere stappen</b>  Zijn alle voorgaande stappen succesvol doorlopen dan wordt de data gedreven toepassing geïntroduceerd in de organisatie. Deze inbedding betekent ook vaak dat de datagedreven toepassing wordt overgedragen naar een beheerorganisatie. Alle output die in voorgaande stappen zijn uitgewerkt dienen hierbij als input om de data gedreven toepassing te documenteren zodat deze in gebruik genomen kan worden.</p>	<p><b>Data gedreven toepassing in business activiteiten</b>  Introduceren van een oplossing waarmee de data gedreven toepassing ingezet kan worden binnen de organisatie. Het introduceren en begeleiding van het gebruik van de toepassing van de business case binnen de organisatie.</p>	<p><b>Business case solution documentatie</b>  Valideren en archiveren van de business case solution documentatie bijvoorbeeld door opname in een metadata register. Eventueel vaststellen van deze documentatie waarmee deze onderdeel gaat uitmaken van de baseline van de data architectuur</p> <p><b>Business case ingebed in baseline data architectuur</b>  Door het in gebruik nemen van de business case in de organisatie is deze business case onderdeel geworden van</p>

**Baseline datamodellen en -architectuur**

De reeds aanwezige datamodellen en baseline data architectuur dient als input om de uitwerking van de business case samen te voegen tot een baseline architectuur waar deze business case deel van uitmaakt.

de baseline data architectuur. Tenzij deze datagedreven toepassing niet aan de data architectuur voldoet dan wordt zij niet opgenomen in de baseline architectuur maar wordt dit geregistreerd als een bouwen zonder architectuur toepassing met een tijdelijke gebruiksperiode.

**Inbedding in baseline data governance inrichting**

De datagedreven toepassing wordt in gebruik genomen en gaat deel uitmaken van de organisatie. Dan dient ook de data governance zoals eigenaarschap, stewardship en de relatie met de datastrategie van de data gedreven toepassing uitgewerkt te zijn. Met name registratie van de data governance aspecten in het metadata register is in deze fase van groot belang.





## Over de auteurs



Bert Dingemans is trainer en coach op het vlak van data architectuur, data management en data modelleren. Hij heeft een passie voor modelleren, modelleertools en het effectief inzetten van geautomatiseerde hulpmiddelen om modellen effectief in te zetten in de praktijk. Zie ook <https://data-docent.nl>, Bert is te via mail bereiken via [bert@interactory.nl](mailto:bert@interactory.nl).



Yorrick Jan Tillemans werkt sinds 2007 met data en ontwerpt sinds 2017 als data architect grote dataplatformen. Momenteel is hij werkzaam als Data Architect van het dataplatform van NS waarop veel data binnenkomt en wordt verwerkt voor diverse toepassingen. Buiten NS deelt hij zijn kennis en enthousiasme door het geven van masterclasses, trainingen, advies en als spreker op diverse Data & Analytics-events. Zie ook: [www.yorrick.consulting](http://www.yorrick.consulting)

Bert en Yorrick zijn beiden betrokken bij de Data architecten community en de werkgroep metadata van DaMa-NL.

## Bronnen

Arcitura, *Big Data Foundation*, Arcitura, z.j.

Best, Bart de, *Acceptatiecriteria*, Academic service, 2011.

Dingemans, Bert, *Grip op data leerpakket*, Brave new books, 2026.

Erl, Thomas, *SOA design patterns*. Prentice Hall, 2003.

Erl, Thomas & Roger Stoffers, *Field guide to digital transformation, the*, Pearson, z.j.

Hay, David, *Data model patterns, convention of thoughts*, Pearson Education, 1996.

Hohpe, George et al (2008). *Enterprise Integration Patterns : Designing, Building, and Deploying Messaging Solutions*, Addison Wesley, 2004.

Microsoft corporation, *Data Patterns (Patterns & Practices)*, z.j., Microsoft Press.

Mosley, Mark et al, *DAMA guide to the data management body of knowledge 2.0*. Bradley beach: Technics Publications LLC, 2018.

Rijn, Ria. van, *Fasen van volwassenheid in enterprise architectuur en hun invloed op data-architectuur*, [www.architectuurassistent.nl](http://www.architectuurassistent.nl), 2013.

Silverton, Len, *Data model resource book, vol 1,2,3*, the, Wiley, 2001.

Snoeck, Monique et al, *Object oriented enterprise modelling with Merode*, Leuven: Coronet Books, 1999.

The Open Group, *ArchiMate 3.1*, Van Haren Publishing, 2019.

The Open Group, *The Togaf standard, tenth edition*, Van Haren Publishing, 2022.

Wikimedia, *Wikipedia.nl, de vrije encyclopedie*, Wikimedia, z.j..