

De waarde van data



Bert Dingemans
Augustus 2022

Inhoudsopgave

Samenvatting	3
Wat is data?	3
Definitie van data	3
Aspecten van data	4
Data stromen	5
Datastroom model.	5
Data, informatie en kennis	6
Data, informatie kennis pyramide	7
Data als productiemiddel	8
Data toepassingen	8
Data als productiemiddel	9
Data gedreven raamwerk	10
Data Management	12
Data Management Framework	12
DaMa Body of Knowledge	13
Meer informatie	14
Over de auteur	14

Samenvatting

Veel organisaties onderkennen de waarde van data en willen deze data inzetten om meerwaarde te realiseren. Vaak wordt hiervoor de term data gedreven werken gebruikt. 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.

Wat is data?

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 leeftijd 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 informatie theorie 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.

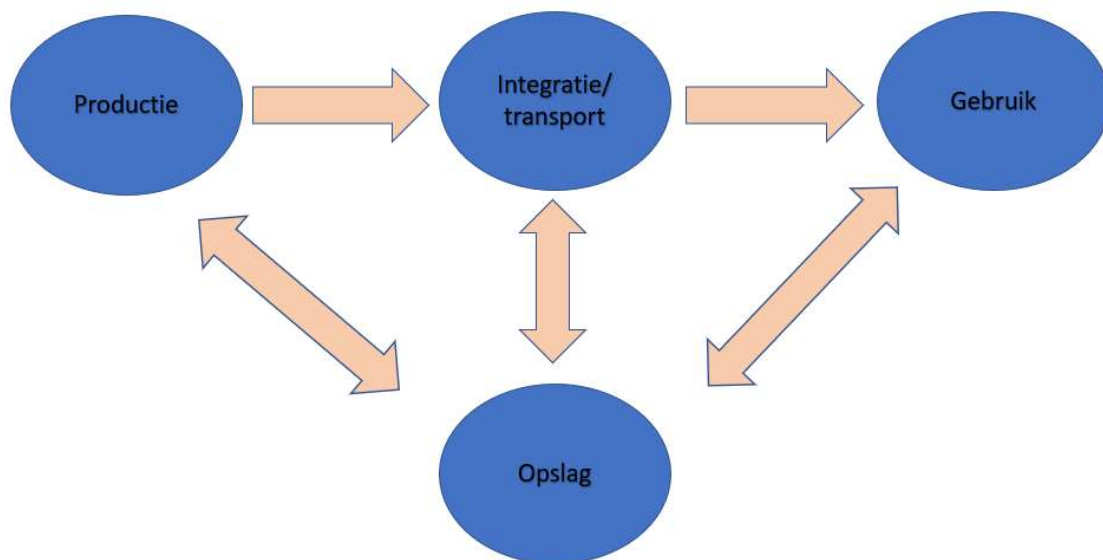
Vanuit de waarde van dat is dit een belangrijk gegeven. Ook tussen de verzender en de ontvanger van data zal bij onduidelijkheid de waarde van data dalen. Er is tenslotte onduidelijkheid en het is maar de vraag of de ontvanger het juiste begrip krijgt of de juiste actie onderneemt.

Data stromen

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.

Datastroom model.

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 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 informatie theorie.

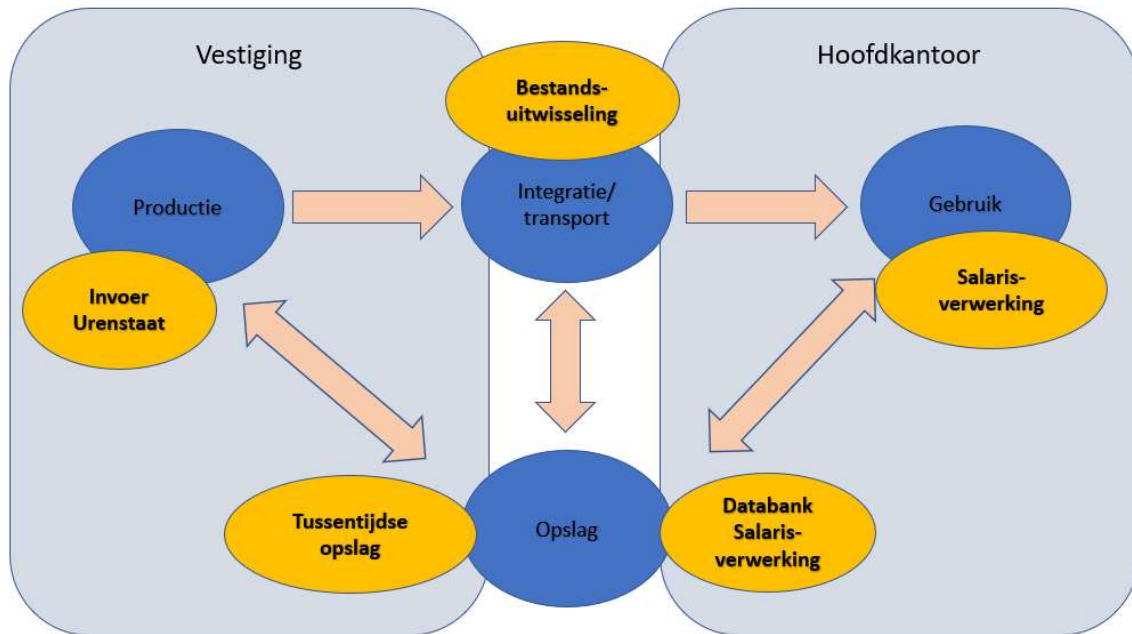
Echter tussen de productie en het gebruik gebeurt iets interessants vanuit het perspectief van de data.

In het voorgaande hoofdstuk gingen we al kort in op datamodellen. Hier komt dat terug. 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 die er heel anders uitziet dan het model voor de reisprijs berekening. 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.



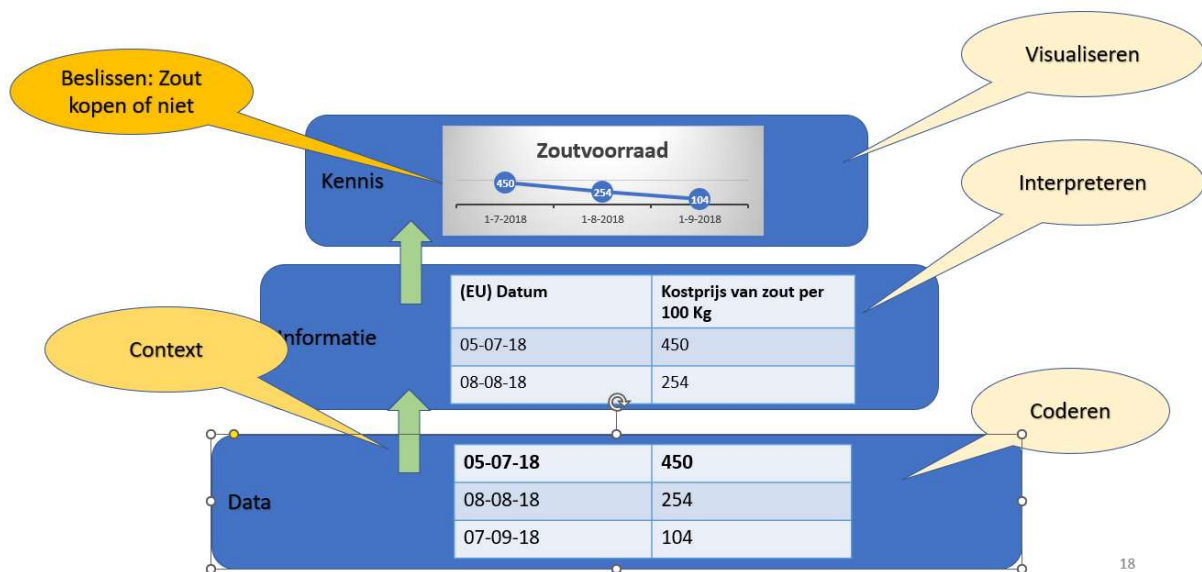
In het voorbeeld van tijdregistratie zal een medewerker uren invoeren in een urenstaat. Dat heeft een (data)model waarmee het eenvoudig is voor de medewerker om de urenstaat in te voeren. Echter wil men van de urenstaten van de verschillende medewerkers salaris- en verlofregistratie doen dan zal hierbij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

Data, informatie 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.

In de afbeelding zien we een eenvoudig willekeurig voorbeeld. In de onderste laag zien we data. Echter deze data bestaat 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 datastroom 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 als productiemiddel

Uit het voorgaande hoofdstuk is gebleken dat data waarde kan hebben voor een organisatie. Veel organisaties zijn daarom op zoek naar deze waarde van data en zoeken naar toepassingen voor data als productiemiddel. In dit hoofdstuk kijken we zowel naar toepassingen als naar data als productiemiddel.

Data 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 bij een bank een lening aanvragen dan zal de bank data gebruiken om te analyseren of jouw gedrag en toekomstperspectief een risico is bij het verstrekken van een lening. De besluitvorming binnen de bank over een lening zal dus resulteren in het wel of niet verstrekken van een lening.
- **Verantwoording**, dit moet iedereen op enigerlei wijze afleggen. Een medewerker moet verantwoording afleggen over wat hij of zij gedaan heeft, wat de resultaten van het werk zijn aan de manager. Zie de voorgaande voorbeeld van de urenstaat. 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 maak via de website van de kapper een afspraak om mijn haar te laten knippen. De afspraak wordt daarmee in het afsprakeninformatiesysteem van de kapper gezet. Op basis van de input vanuit dit afsprakenstelsel naar het planningssysteem wordt de planning gemaakt om te bepalen hoeveel kappers er moeten werken op een bepaalde dag en tijd.
- **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 dataproducenten kunnen er nieuwe producten en diensten ontstaan. Bijvoorbeeld de pakketbezorgers verzamelen voor het interne proces al data over waar een pakketje zich bevindt in het interne bezorgproces. Echter zij bieden vervolgens tracking en tracing aan zoals je als ontvanger van een pakketje kunt zien waar je pakket zich bevindt in het proces en wanneer het afgeleverd gaat worden.
- **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.

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 sturing 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 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 allerlei 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.

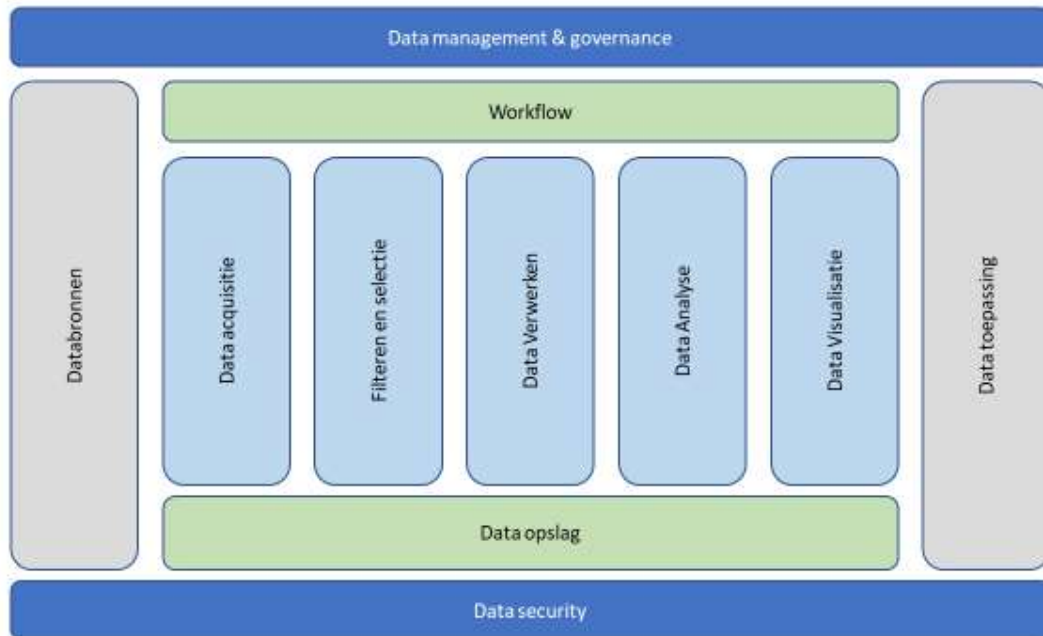
- **Kunstmatige intelligentie en Data Science**, dit zijn activiteiten die voortborduren op het transformeren en analyseren van data met moderne werkwijze zoals algoritmen en machinaal leren en andere moderne technieken. Met deze moderne methoden verkrijgen die met traditionele hulpmiddelen (nog)niet mogelijk zijn. Bijvoorbeeld de zitplaatszoeker die de klant informeert hoe vol de trein zal zitten en in welk treinstel nog wel zitplaatsen zijn.
- **Big Dataoplossingen 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 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 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.

Data gedreven raamwerk

Rond data gedreven werken is een algemeen raamwerk te gebruiken gebaseerd op een big data patroon de data pipe. Dit raamwerk is afkomstig van Arcitura en is een detaillering van de datalevensloop en wordt veel toegepast in (big) data integratie projecten. Het is feitelijk een raamwerk waarin je verschillende projectactiviteiten, deliverables en modelleervormen kunt afbeelden. Dit helpt om de complexiteit op eenvoudige wijze in kaart te brengen. Onderstaande afbeelding toont de data pipe waarmee je een data gedreven toepassing kunt realiseren.



Bron: Arcitura

De datapipe wordt gelezen van links naar rechts en toont de stappen die genomen moeten worden om de productie van data in de bronnen in te kunnen zetten in een data toepassing (datagebruik). Daarnaast zijn er een aantal extra dimensie toegevoegd namelijk data management en security aspecten. In onderstaande opsomming een korte toelichting op de onderdelen:

- **Databronnen**, gegevensverzamelingen die gebruikt worden als grondstof voor de data toepassing
- **Data acquisitie**, activiteiten die de verkrijging van de relevante databronnen bewerkstelligen
- **Filteren en selectie**, bewerken van de gegevens vanuit de databronnen tot die datasets die relevant zijn voor de toepassing
- **Verwerken**, transformatie, manipulatie en verrijking om het datamodel geschikt te maken voor een adequate data analyse, visualisatie en toepassing
- **Analyse**, activiteiten waarbij de getransformeerde data gebruikt wordt voor het zoeken naar verbanden, patronen of statistische verhoudingen
- **Visualisatie**, zichtbaar maken van de analyse resultaten ter ondersteuning van de analisten of voor presentatie aan andere stakeholders
- **Toepassen**, inzet van data in verschillende vormen van besluitvorming. Hier vindt de uiteindelijke creatie van waarde uit data plaats.
- **Dataopslag**, opslag van de data en tussenproducten voor later gebruik in vervolgstappen of andere datapipes
- **Security**, beveiligings- en privacy aspecten van data
- **Data management**, data is een asset en daarom dient er management op plaats te vinden om de waarde van data te verhogen of behouden
- **Workflow**, automatiseren en standaardiseren van bewerkingsstappen op de data

Zoals reeds genoemd is dit een detaillering van de datalevensloop en gaat dit met name in op de integratie aspecten voor het creëren van waarde uit data bij de transformatie van data uit de bronnen tot de realisatie van waarde bij de uiteindelijke toepassing.

Data Management

In het vorige hoofdstuk hebben we gezien dat data beschouwd kan worden als een productiemiddel. Dat houdt dan vervolgens in dat we data moeten managen zoals we ook Assets - , Human Resources -en Financial management implementeren. Data Management implementeren kan het beste gedaan worden via een gestructureerde aanpak.

Nu is het mogelijk om dit voor de eigen organisatie uit te werken. Waarschijnlijk is het beter en efficiënter om gebruik te maken van een bestaand Data Management Framework. In dit hoofdstuk gaan we hier kort op in.

Data Management Framework

Steeds meer organisaties zien het belang van data management en willen dit implementeren in de eigen organisatie. Veelal wordt gekozen voor een bestaand raamwerk om dit te doen. Introduceren van een bestaand data management framework in de eigen organisatie heeft een aantal belangrijke voordelen:

- Een framework is ontwikkeld door een consortium van organisaties die het belang onderkennen van de knelpunten en de mogelijke oplossingen binnen data management
- Een framework heeft een gestructureerde opzet en geeft in een aantal gevallen scenario's gestandaardiseerde werkwijzen.
- Daarnaast hebben veel data management frameworks een assessment om een nulpunt rond de volwassenheid van data management in de organisatie te bepalen.
- Frameworks bieden de mogelijkheid om aan cherry picking te doen.

Er zijn een paar veel toegepaste data management raamwerken:

- [Mike 2.0](#): Method for an Integrated Knowledge Environment (MIKE2.0) is een open source delivery framework voor Enterprise Information Management. Het biedt een uitgebreide methodologie die kan worden toegepast op een aantal verschillende projecten binnen de informatiebeheerruimte. Hoewel het aanvankelijk gericht was op gestructureerde gegevens, is het doel van MIKE2.0 om een uitgebreide methodologie te bieden voor elk type informatieontwikkeling.
- [DCAM– het Data Management Capability Assessment Model](#) – is ontwikkeld door leden van de EDM Council en is het industriestandaardraamwerk voor datamanagement. DCAM definieert de reikwijdte van de capaciteiten die nodig zijn om een volwassen Data Management-discipline op te zetten, mogelijk te maken en in stand te houden. Het behandelt de strategieën, organisatiestructuren, technologie en operationele best practices die nodig zijn om gegevensbeheer in uw hele organisatie met succes aan te sturen, en zorgt ervoor dat uw gegevens digitale transformatie, geavanceerde analyses zoals AI en ML en gegevensethiek kunnen ondersteunen.
- [DAMA](#). The Data Management Association International is een internationale not-for-profit organisatie van data management professionals. DAMA International heeft wereldwijd verschillende afdelingen en belangenbehartigers. Data Management is ontwikkeling en executie van architecturen, regelgeving, praktijk en procedures om op een goede manier de volledige data management lifecycle van een organisatie te ondersteunen. Stichting DAMA NL

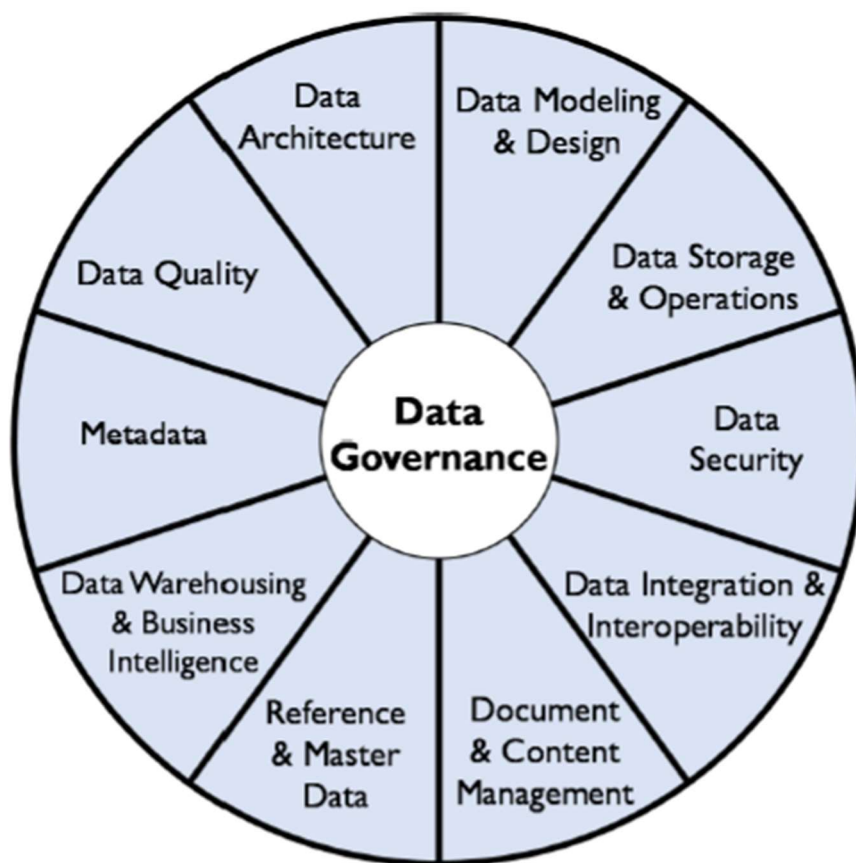
is een Nederlandse afdeling die voor de Nederlandse markt haar activiteiten ontplooit om deze doelstellingen te ondersteunen.

In dit whitepaper gaan we dieper in op DaMa en de DaMa Body of Knowledge (DMBoK). Reden is dat dit in de Nederlandse situatie het meest bekende data management framework is.

DaMa Body of Knowledge

Dit Data Management Framework heeft een gestructureerde uitwerking en is gebaseerd op een aantal kennisgebieden. Binnen deze kennisgebieden gebruikt DaMa steeds een zelfde indeling van de relevante onderwerpen in een kennisgebied. Hieronder zie je een afbeelding van de verschillende kennisgebieden binnen DMBoK.

Het DMBoK is verschenen in een (electronische) publicatie. Daarnaast zijn er trainingen en is er een certificering ontwikkeld.



Bron:DMBoK

Hieronder van ieder kennisgebied een korte definitie:

- **Data Governance:** Is het uitvoeren van controle en beheer omtrent het beheer van data assets. Data Governance stuurt alle andere dataprocessen.
- **Data modelleren en – ontwerp:** Is het ontdekken, analyseren en beschrijven van data requirements in de vorm van gestandaardiseerde modellen die een data structuur beschrijven.
- **Data opslag en beheer:** Ontwerp en implementatie van data opslag en -persistentie.

- **Data security:** Activiteiten rond de bescherming van informatie en data door autorisatie, authenticatie, toegang, auditing.
- **Data integratie en interoperabiliteit:** Managen van het transport en consolidatie van data tussen informatiesystemen en organisaties.
- **Document en content management:** Managen en (levensloop)beheer van alle soorten data inclusief documenten en content.
- **Referentie en master data:** Managen van generieke en algemene (herbruikbare) data en referentie data (codelijsten e.d.).
- **Data Warehousing en Business Intelligence:** Planning, ontwikkeling en beheer van activiteiten voor het samenstellen van data ter ondersteuning van besluitvorming en kenniswerkers.
- **Meta Data:** Managen, ontwikkelen en beheren van metadata.
- **Data Kwaliteit:** Activiteiten voor kwaliteitsmanagement van data assets zodat het geschikt is voor gebruik en voldoet aan de wensen van de data consumenten.
- **Data Architectuur:** Managen, ontwikkelen en beheren van de requirements en principes rond data.

Wil je als organisatie data gedreven gaan werken en data als een productiemiddel gaan inzetten dan is het managen van de data een randvoorwaarde. In dit hoofdstuk heb ik kort aangegeven welke frameworks er zijn en binnen DMBOK welke kennisgebieden er zijn in het werkveld data management.

Meer informatie

Wil je meer informatie over de waarde van data. Kijk dan eens naar meer informatie op:

- <http://wpp.interactory.nl/frmSearchList.aspx?module=webcontent&mode=direct&searchfor=dat a%20management>
- <http://wpp.interactory.nl/frmSearchList.aspx?module=webcontent&mode=direct&searchfor=me ta%20data>
- <http://medama.eexpertise.nl/Home.aspx>
- <https://www.dama-nl.org/>

Over de auteur



Bert Dingemans is trainer op het vlak van data architectuur, data management en Big Data. Hij heeft een passie voor modelleren, modelleertools en het effectief inzetten van geautomatiseerde hulpmiddelen om modellen effectief in te zetten in de praktijk. Bert is te bereiken via bert@interactory.nl