

Grip op data-architectuur

Grip op data-architectuur

Bert Dingemans

Schrijver: Ir. Ing. Bert Dingemans

Coverontwerp: Rudy Barten

ISBN: 9789465209265

© Ir. Ing. Bert Dingemans

Voorwoord

Heeft een boek specifiek over data-architectuur in de praktijk toegevoegde waarde? Data krijgt op dit moment volop aandacht vanuit allerlei gezichtspunten, zoals technisch, maatschappelijk, management en ethisch. Die vraag is daarom met een volmondig ja beantwoorden.

Of je nu werkt op het vlak van datamanagement, kunstmatige intelligentie, data gedreven werken, enterprise-architectuur of modellering er is altijd een relatie met data. Het structureren van al deze relaties met data en van de gegevens onderling is daarbij het werkveld van de data-architect. Alle activiteiten op het vlak van data-architectuur kunnen daarmee toegevoegde waarde leveren aan een efficiëntere en wendbare gegevensvoorziening.

Dit boek bestaat uit vier delen, in ieder deel wordt vanuit een ander gezichtsveld gekeken naar data en architectuur. Ieder deel bestaat uit een aantal inhoudelijk hoofdstukken en wordt afgesloten met een aantal praktische tips.

In het eerste deel wordt gekeken naar de omgeving van de data-architect. Vanuit de rol van de architect worden stakeholders en concerns behandeld. Verder wordt ingegaan op hoe de omgeving gemodelleerd kan worden om inzicht te krijgen in de onderlinge verbanden binnen de stakeholders en concerns.

In het tweede deel ligt de focus op de wijze waarop de data-architect kaders kan stellen aan de data zoals aanwezig in organisatie. Hierbij wordt uitgegaan van het sturen van verandering binnen projecten. Er wordt uitgebreid stilgestaan bij data-kwaliteiten en hoe deze gebruikt kunnen worden om meetbare kaders te stellen. Daarnaast wordt ingegaan op andere kadervormen zoals patronen, risico's en principes.

Modelleren van data-architectuur is het onderwerp van het derde deel. Op basis van verschillende modelleertechnieken zoals ER-diagrammen, UML en ArchiMate wordt een beeld gegeven van data-architectuurmodellering.

Het vierde deel gaat in op belangrijke producten van de data-architectuur namelijk patronen en bouwblokken. Hierbij is de keuze gemaakt om een beperkt voorbeelden van patronen en bouwblokken in detail te beschrijven.

Het boek is een herziene uitgave van de publicatie “Data-architectuur in de praktijk” verschenen in 2014. In elf jaar is er veel veranderd in het werkveld van de data-architect.

Het boek is beschikbaar in een papieren- en een elektronische versie. Daarnaast is er het ondersteunende website waarop hulpmiddelen, checklists en een register met data-architectuur voorbeelden te vinden zijn.

Zonder hulp van een aantal mensen had dit boek niet de huidige vorm gekregen. Drie personen wil ik met name noemen. Allereerst Huup de Jong voor zijn suggesties over conceptuele data modellering. Ten tweede Yorrick Tillemans voor zijn bijdragen aan de data-architectuur patronen. Als laatste mijn partner Anneke Hubert voor haar steun. Rest mij iedereen veel inzichten over data-architectuur toe te wensen bij het lezen van dit boek.

Culemborg, Augustus 2025.

Inhoud

Voorwoord	5
Inhoud	7
Data-architectuur producten.....	11
Deel 1: Data-architectuur en omgeving	12
Wat is data?.....	13
Definitie van data	13
Aspecten van data	15
Data, informatie en kennis	16
Data, informatie kennis pyramide	16
Data als productiemiddel	17
Data toepassingen	17
Productiemiddel	19
Data Management.....	21
Data Management Raamwerk	21
DaMa Body of Knowledge	22
Context van data-architectuur	25
Inleiding	25
Definities	25
Gerelateerde werkvelden	27
Waarde van de data-architect	30
Inleiding	30
Activiteiten van de traditionele architect	30
Nieuwe rolinvulling voor de architect	33
De rol van de data-architect	42
Inleiding	42
Veranderen	42
Dimensies van data-architectuur	43
Inventariseren, modelleren, begeleiden	47
Beïnvloedingsstrategieën	49
Projecten	50
Roluitvoering	51

Data-architectuur producten.....	54
Inleiding	54
Overzicht data-architectuur Producten	55
Vaardigheden van een data-architect	61
Inleiding	61
Competenties data-architect	62
Stakeholders van de data architect	74
Inleiding	74
ISO 42010 model	74
Stakeholders bij data-architectuur	76
Overzicht stakeholders data-architectuur	77
“Onzichtbare stakeholders”	83
Stakeholder management	84
Concerns	85
Inleiding	85
Inventarisatie	85
Concerns modelleren	87
Concerns bij de data-architectuur	88
Tips over de waarde van data-architect	92
Inleiding	92
Overzicht van tips	92
Tot slot	95
Deel 2 Data-architectuur kaders	96
Data kwaliteiten	97
Inleiding	97
Dimensies	97
Data kwaliteiten en maatregelen	101
Data Kwaliteit maatregelen werkprocessen	103
Data Kwaliteit maatregelen RDBMS	105
Data Kwaliteit maatregelen software	107
Kwaliteit Maatregelen data integratie	109
Data Kwaliteit Maatregelen user interface	110
Datakwaliteiten en Maatregelen per dimensie	112

Organisatievolwassenheid	147
Data-principes	148
Inleiding	148
Doelen en stakeholders en principes	148
Principes als kaders bij veranderingen	149
Opbouw principe	150
Voorbeeld principes	152
Principes en projecten	152
Principe register	158
Andere kadervormen	159
Acceptatiecriteria	159
Risico's	160
Architectuurbesluit	161
Tips voor architectuurkaders	162
Inleiding	162
Overzicht van tips	163
Tot slot	165
Deel 3 Beschrijvende architectuur	166
Beschrijvende data-architectuur	167
Inleiding	167
Data-architectuur modelleren raamwerk	167
Modelleertalen	169
Stakeholders	170
Requirements	171
Verschijningsvormen van data-architectuur modellen	173
Opzet data-architectuur modelleerwijzen	178
Beschrijvende architectuur	181
Kaderstellende architectuur	214
Data management	230
Data platform en toepassingen	244
Tooling & aanvullende producten	261
Evaluatie	261
data-architectuur metamodel	262

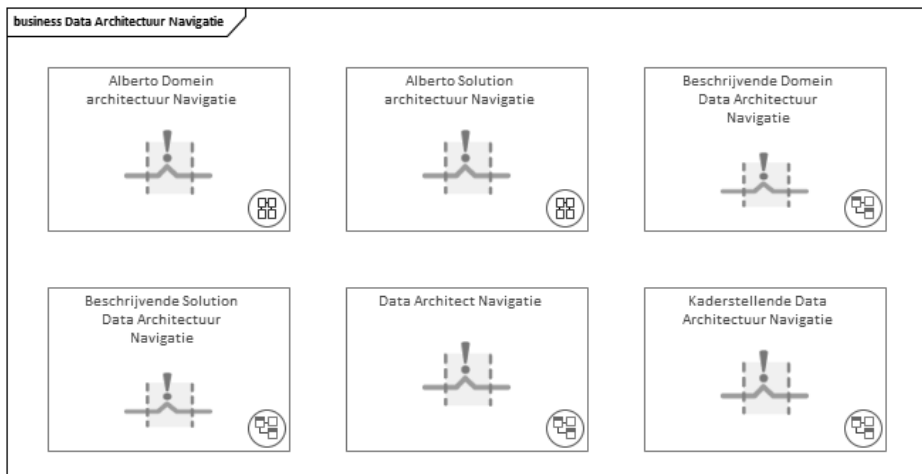
Tips voor data-architectuur modelleren	263
Inleiding	263
Overzicht van modelleertips	263
Tot slot	271
Deel 4 Data-architectuur bouwblokken	272
Data-architectuur patronen.....	273
Wat is de IKEA metafoor	273
Architectuur patronen en -bouwblokken	276
Definitie patronen	276
Definitie bouwblokken	277
Patroon en bouwblok register	278
Voorbeelden van patronen	279
Data gedreven werken	279
Data gedreven grondplaat	279
Data integratie	285
Logische datamodellen	294
Fysieke datamodellen	300
Metadata	301
Master en Referentie Data	306
Business Intelligence, DWH en data analyse	323
Tips voor patronen en bouwblokken.....	333
Inleiding	333
Overzicht van data patronen	333
Referenties	336

Data-architectuur producten

Dit boek “Grip op data-architectuur” is deel van een combinatie van een aantal producten. Hiermee is het voor een data-architect mogelijk om op basis van deze combinatie een gestructureerde en effectieve nieuwe data-architectuur diensten en producten te introduceren. In onderstaande overzicht de producten behorend bij dit boek.

Voor de lezers van het papieren boek zijn de producten uit het combinatiepakket te vinden via:

<https://data-docent.nl/dataarchitectuurlinksNL.aspx>



Voorbeelden van de data-architectuur producten zoals:

- Architectuur principes.
- Data-architectuur modellen.
- Register met een aantal data-architectuur patronen.
- Inrichting van een data-architectuur repository.
- Data-architectuur metamodel.

De online content is uitgewerkt in een repository. De inhoud is via een webapplicatie rechtstreeks te raadplegen. De repository kan worden gedownload als een Sparx Enterprise Architect file die met deze tool kan worden geraadpleegd en desgewenst uitgebreid naar de eigen context. De modellen zijn ook beschikbaar voor download in het XMI uitwisselformaat voor andere modelleer tools.

Deel 1: Data-architectuur en omgeving

In dit eerste deel wordt ingegaan op de omgeving van de data-architect.

Deze omgeving wordt vanuit meerdere gezichtsvelden beschouwd.

Allereerst wordt ingegaan op de verschillende werkvelden gerelateerd aan data-architectuur. Ook de stakeholders van de data-architect worden beschreven. De stakeholders worden gecategoriseerd en er wordt ingegaan op zichtbare en onzichtbare stakeholders. Verder wordt gekeken naar stakeholdermanagement een bijzondere manier van het categoriseren van stakeholders.

Via stakeholders wordt gekeken naar welke concerns of belangen de verschillende stakeholders hebben. Door concerns in kaart te brengen ontstaan toepasbare eisen voor de gewenste situatie binnen de architectuur.

De rol van de architect in relatie tot de omgeving en de belangen die daar bestaan wordt beschreven in een hoofdstuk in dit deel. De manier waarop een architect zijn rol invult heeft gevolgen voor de successen die gehaald worden. Ook is de rol afhankelijk van de volwassenheid en de cultuur van de organisatie. Er dient dan ook een match te zijn tussen rolinvulling en inrichting en cultuur van de organisatie.

Wat is data?

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.

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.

Is data een waardevol productiemiddel voor organisaties, dan zullen we data willen managen. Willen we data managen binnen de verandering van een organisatie dan zal data-architectuur essentieel zijn.

Definitie van data

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

Een **gegeven**, ook wel **datum** [wikipedia]:

- 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.

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

parallele 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 parallele 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 concepten 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 parallele 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 en vervolgens de juiste actie onderneemt.

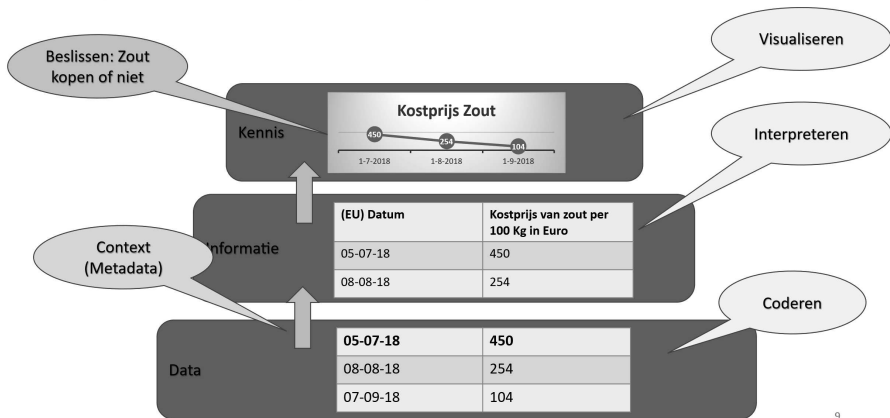
Data, informatie en kennis

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.

Data, informatie, kennis, inzicht en beslissen



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 ontvangerszijde van de datastroom niets. Er mist namelijk iets essentieels 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 metadata 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 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. Zoals met een urenstaat plaatsvindt. Ook organisaties moeten verantwoording afleggen aan allerlei controlerende- en overheidsinstellingen zoals 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 kortere doorlooptijd te realiseren of een betere dienst te leveren aan de uiteindelijke afnemers.
- **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 afsprakensysteem 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 op het spoor 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 organisaties 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.

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 voor 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 geschreven, 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 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 omdat ze me niet meer kunnen bereiken.

Data Management

Hierboven 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 Raamwerk

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 raamwerk in de eigen organisatie heeft een aantal belangrijke voordelen:

- Een raamwerk is ontwikkeld door een consortium van organisaties die het belang onderkennen van de knelpunten en de mogelijke oplossingen binnen data management
- Een raamwerk 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.
- Raamwerken bieden de mogelijkheid om aan cherry picking te doen.

Er zijn een paar veel toegepaste data management raamwerken:

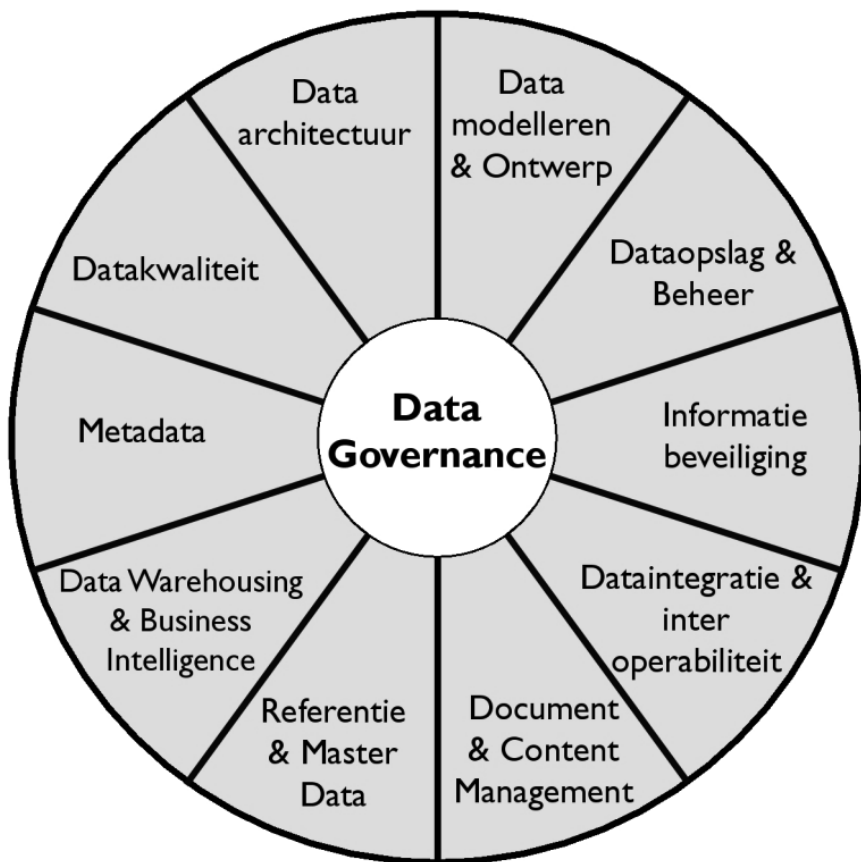
- DCAM- het Data Management Capability Assessment Model - is ontwikkeld door leden van de EDM Council en is het industriestandaardraamwerk voor datamanagement. DCAM definieert de reikwijdte van de capaciteiten die nodig zijn om een volwassen Data Management-discipline op te zetten, mogelijk te maken 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.

Hier gaan we dieper in op DaMa en de DaMa Body of Knowledge (DMBoK). Reden is dat dit in de Nederlandse situatie, op het moment van schrijven van dit boek, het meest toegepaste data management raamwerk 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 (elektronische) publicatie. Daarnaast zijn er trainingen en is er een certificering ontwikkeld.



Copyright © 2017 by DAMA International

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.
- **Metadata:** 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.

Interessant is dat data-architectuur zelf een kennisgebied is binnen het DMBOK raamwerk. Data-architectuur is hierbij gericht op stakeholders binnen data gedreven werken, kaders die daarbij ontstaan en vervolgens het begeleiden van de verandering in een organisatie naar data gedreven werken.

Context van data-architectuur

Data-architectuur is een werkveld dat verweven is met een aantal andere werkvelden zoals data management en enterprise architectuur. Wat zijn definities van data-architectuur en wat is de impact op de werkzaamheden van de architect gerelateerd aan de verschillende werkvelden.

Inleiding

In het vorige hoofdstuk hebben we gezien dat data een waardevol productiemiddel is en dat we data daarom willen managen. Daar is reeds data-architectuur als kennisgebied in data management aan de orde gekomen. Echter data-architectuur is gerelateerd aan meerdere werkvelden binnen een organisatie.

Daarom gaan we in dit hoofdstuk allereerst kijken naar de verschillende definities die gebruikt worden rond data-architectuur. Vervolgens gaan we in op de gerelateerde werkvelden.

Definities

Rond data-architectuur zijn er meerdere beschrijvingen te vinden. Opvallend hierbij is dat vanuit een gezichtspunt een andere definitie van data-architectuur gegeven wordt.

Onderstaande definitie vanuit Togaf is opgesteld vanuit het perspectief van enterprise architectuur. Het beschouwt data-architectuur gericht op een aspect in een enterprise namelijk data. Opvallend is dat in deze definitie er niet ingegaan wordt op het beschrijven van data hoe dat in organisaties gebruikt wordt.

Een beschrijving van de structuur van de belangrijkste typen en bronnen van gegevens, logische gegevensactiva, fysieke gegevensactiva en gegevensbeheerbronnen van de onderneming.
[Togaf]

DaMa benoemt in de definitie de enterprise en geeft daarbij duidelijk aan dat de vereisten rond data in kaart moeten worden gebracht en daarbij ingaan op hoe de veranderingen in de organisatie afgestemd dienen te worden op de data. Er wordt hier niet ingegaan op het beschrijven van de data. Dat is logisch want in de DMBok is hiervoor een separaat kennisgebied namelijk datamodelleren.

“Enterprise data-architectuur” is: een geïntegreerde set specificatie-artefacten die strategische datavereisten definiëren, de integratie van data-activa begeleiden en data-investeringen afstemmen op de bedrijfsstrategie. [DaMa]

In de definitie afkomstig van Wikipedia (Nederlands) zie je dat data-architectuur hier beschouwd wordt als onderdeel van de enterprise architectuur. Mooi om te zien dat hier de aanbod- en vraagzijde van data in organisaties benoemd wordt. Ook goed om te zien hoe hier ingegaan wordt op het beschrijvende- en voorschrijvende karakter van data-architectuur. De verschillende data artefacten worden hier benoemd maar die is wat beperkt van opzet naar mijn mening.

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 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. [Wikipedia (NL)]

In de definitie van Wikipedia (Engels) zie je eveneens de combinatie van een voorschrijvende en beschrijvende architectuur. En gaat daarbij in op hoe data verwerkt, gebruikt en opgeslagen worden.

Data-architectuur bestaat uit modellen, beleid, regels en standaarden die bepalen welke data wordt verzameld en hoe deze wordt opgeslagen, gerangschikt, geïntegreerd en gebruikt in datasystemen en organisaties. Data is doorgaans een van de verschillende architectuurdomeinen die de pijlers vormen van een enterprise-architectuur of oplossingsarchitectuur. [Wikipedia (EN)]

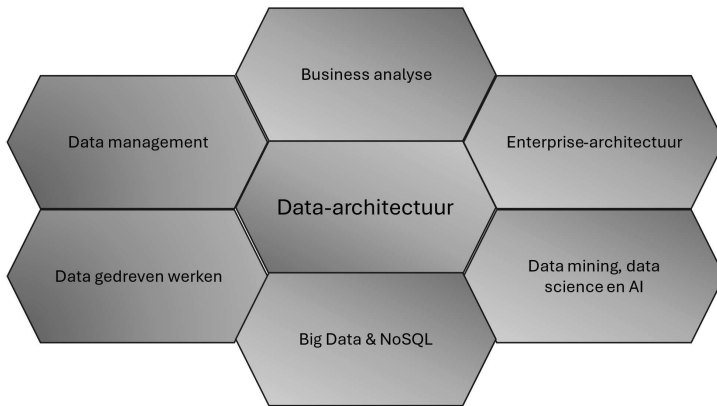
Op basis van bovenstaande definities is kenmerkend dat er meerdere gezichtspunten zijn naar data-architectuur. Overeenkomst is dat data een productiemiddel is met waarde en op die manier behandeld dient te worden. In dit boek gebruiken we verder de volgende samenvattende definitie van data-architectuur.

Data gedreven organisaties zien data als een waardevol productiemiddel. Daarom wordt bij veranderingen in organisaties het belang van data als aspect van de veranderingen beschouwd. Data wordt daarom uitgewerkt in een beschrijvende en voorschrijvende architectuur. Hiermee wordt een beschrijvend beeld geschetst hoe data in de organisatie wordt ingezet. Met de voorschrijvende architectuur worden kaders gesteld aan de verandering zodat deze verandering bijdraagt aan de realisatie van een effectieve data gedreven organisatie.

Gerelateerde werkvelden

In het vorige hoofdstuk hebben we de definities behandeld. Daar bleek reeds dat er vanuit verschillende werkvelden verschillende definities bestaan over data-architectuur. Dat wordt veroorzaakt door het feit dat verschillende werkvelden eigen gezichtspunten hebben naar data en architectuur.

Hieronder een afbeelding van de werkvelden die een directe relatie hebben met data-architectuur. Hierbij geen uitputtende lijst waarbij data-architectuur en de genoemde werkvelden van invloed zijn op elkaar. We lichten de werkvelden kort toe en geven aan wat de relatie is naar data-architectuur.



- **Enterprise architectuur:** Dit omvat de activiteiten in een organisatie of enterprise waarbij de veranderingen in een organisatie gecontroleerd kunnen plaatsvinden. Binnen enterprise architectuur zijn er meerdere disciplines zoals applicatie- bedrijfs-, informatie- en technische architectuur. Data-architectuur is een soortgelijke discipline. Wederzijds is er een nauwe relatie. Het maken van beleid en kaders in het ene domein zal het andere domein beïnvloeden. Veranderingen in het applicatielandschap hebben invloed op het datalandschap en vice versa.
- **Business Analyse:** Veranderingen in de organisatie kijkt naar de strategie, missie en visie van organisaties en vertaalt deze visies naar hoe de organisatie zal worden ingericht om invulling te geven aan de missie en visie. Vanuit alle perspectieven zoals bedrijfsprocessen, personeel, cultuur en ook de informatiesystemen wordt vanuit business analyse beschreven hoe de business verandering eruit ziet. De relatie met data is hierbij dat de veranderingen in de organisatie ook van invloed is op het toepassen, gebruiken en bewaken van de data. Daar is dan ook een nauwe relatie tussen data-architectuur en business analyse.
- **Data management:** Dit werkveld wordt bij steeds meer organisatie als een werkveld geïntroduceerd. Data is een waardevol productiemiddel en dient daarom beheerd te worden. Data management begint bij een data strategie en bij data governance. Daar ontstaat een nauwe relatie met data-architectuur. Want vanuit de strategie en de governance zal de data-architect kaders stellen aan de verandering en in een beschrijvende data-architectuur een

beeld geven van hoe data management geïmplementeerd wordt in de organisatie.

- **Data gedreven werken:** Organisaties willen vanuit data, informatie en kennis inzicht krijgen uit de data die binnen een organisatie ingezet wordt. Hiervoor worden veelsoortige initiatieven uitgevoerd om data gedreven werken te introduceren in de organisatie. Vanuit data-architectuur zal aan deze veranderingen in de organisatie voorschrijvende- en beschrijvende architectuur worden gesteld.
- **Big Data & NoSQL:** Organisaties verzamelen steeds meer vormen van data. Niet alleen grote hoeveelheden data maar ook veelvormige data worden verzameld ter ondersteuning van een data gedreven organisatie. Daardoor wordt het beschikbaar stellen, ontsluiten en opslaan van de data complex. Reden om nieuwe technieken, paradigma's en dataformaten in te zetten. Aan deze ontwikkelingen zal vanuit data-architectuur een kaderstellende architectuur worden opgesteld. Echter rekening houden met de kansen en perspectieven van deze nieuwe technieken is hierbij vanuit architectuur essentieel.
- **Data mining, data science en AI:** Op deze werkvelden zijn de ontwikkelingen ronduit stormachtig te noemen. Dat biedt ongekende mogelijkheden maar brengt ook een aantal risico's voor de stakeholders binnen en buiten de organisatie met zich mee. Data-architectuur dient daarom deze ontwikkelingen op adequate wijze te ondersteunen, beschrijvende architecturen op te stellen, kaders te stellen en de verschillende concerns van de stakeholders rond deze werkvelden voldoende te borgen. Dit wordt gedaan door afwegingen te maken rond de verschillende concerns en op basis daarvan gewogen adviezen te geven aan deze ontwikkelingen.

Waarde van de data-architect

Data heeft waarde voor de organisatie, vanuit data-architectuur kunnen de veranderingen in organisatie om die waarde te creëren ondersteunen. Op welke wijze kan een data-architect waarde leveren aan de begeleiding van die data gerelateerde verandering in een organisatie.

Inleiding

Steeds meer organisaties onderzoeken de mogelijkheden (en onmogelijkheden) van data gedreven werken, data management, Big Data en Data Analytics toepassingen. Maar hoe zit het met de toegevoegde waarde van de architect in deze data centrische soort trajecten. Bij een opdrachtgever waar ik een big data project heb ondersteund werd geopperd dat het prima zonder architect kon. Dat is misschien wat kort door de bocht. Echter de architect zal zichzelf wel heel duidelijk moeten afvragen: Wat is mijn toegevoegde waarde? Duidelijk is in ieder geval dat de “traditionele architect” zichzelf opnieuw zal moeten uitvinden.

Activiteiten van de traditionele architect

Mogelijk komt de term “traditionele architect” wat denigrerend over. Dat is niet de insteek hier. De term wordt gebruikt om aan te geven op welke aspecten de veranderingen in de data-architectuur plaatsvinden. Beschouw de “traditionele architect” als een vertrekpunt voor het schetsen van de rol van een data architect in een data gedreven context.

In veel gevallen bestaat het werk van de architect, uit twee soorten werkzaamheden. Dit zijn voorschrijvende- en beschrijvende werkzaamheden. Met voorschrijven worden kaders en beperkingen gesteld aan veranderingen op basis van concerns van stakeholders. Beschrijven geeft een beeld van de toekomstige situatie.

In de paragrafen hieronder gaan we in op een aantal activiteiten en producten die de “traditionele architect” hoogstwaarschijnlijk onder de loep moet gaan nemen.

Architectuur principes

Voorschrijvende architectuur is vaak gebaseerd op een set van principes. Deze principes zijn daarmee een algemeen kader gericht op het sturen van de verandering, veelal in projecten. Principes zullen vaak gericht zijn op gestructureerde data die opgeslagen wordt in relationele databases. Bij

data gedreven toepassingen zullen er ook datasets zijn die minder gestructureerd zijn.

Gevolg is dat er drie mogelijkheden zijn:

- Herschrijf de principes zodat ze ook inzetbaar zijn voor minder gestructureerde data. Dit heeft als voordeel dat de set van principes relatief beperkt blijft, nadeel is dat de principes mogelijk nietszeggend of minder toepasbaar blijken.
- Maak een aantal principes gericht op laag gestructureerde datasets binnen big data toepassingen. Daarmee wordt er een specifieke set van principes geïntroduceerd. Nadeel is dat de grens tussen laag en hoog gestructureerd moeilijk te bepalen kan zijn. Voordeel is dat je een beperkte specifieke set van principes introduceert voor je big data initiatieven.
- Introduceer een getrappt model van data principes voor data. Door de principes te stapelen kun je data indelen in categorieën waarbij per categorie bepaald wordt welke data principes binnen die categorie geldend zijn.

Veel traditionele dataprincipes zijn gericht op kwaliteit van data in de bron. Deze kwaliteitsaspecten blijven onverminderd van kracht. Echter bij data gedreven initiatieven met andersoortige datastructuren wordt datakwaliteit een uitdaging.

Veelal is de afstand tot de bron groot, heb je weinig invloed op het eigenaarschap en is terug melden van laag kwalitatieve data knelpunten niet relevant. Gevolg is dan ook dat de kwaliteit verhogende maatregelen in de transformatie van de brondata tot aan het gebruik van de afgeleide data plaats zullen vinden. Denk hierbij niet alleen aan het opschonen van data maar ook aan het inzitten van controlestromen of het combineren van stromen om data met voldoende kwaliteit voor gebruik te kunnen produceren. Gevolg is dan ook dat de datakwaliteiten extra aandacht dienen te krijgen bij het uitwerken van data gedreven oplossingen met veel andersoortige data.

Is het lastig om de principes toepasbaar te maken. Dan kun je overwegen om niet met principes te werken maar te kiezen voor een uitwerking gebaseerd op requirements. Hierbij kun je requirements verzamelen van de stakeholders rond je data gedreven toepassing en op basis daarvan een kader stellend beeld schetsen van de te introduceren oplossing. Je kunt de data kwaliteiten in dat geval als kadervorm gebruiken en beschrijven welke maatregelen het gewenste kwaliteitsniveau bereiken.

Solution Architectuur (PSA)

In principe is het uitwerken van een kaderstellend architectuurdokument bij de start van een verandering binnen een project een goede aanpak. Project Start Architecturen (PSA) en solution architectures zijn kenmerkende voorbeelden van dergelijke architectuurdokumentten. Beide soorten van architectuurdokumentten worden veelal voor de start van een project geschreven.

Bij data gedreven trajecten, zeker bij de innovatieve projecten, zal de traditionele werkwijze met een uitgebreid en gedetailleerd architectuurdokument niet goed werken. Aan het begin van het project zijn er nog teveel onzekerheden, is er nog geen duidelijk beeld over de toekomstige situatie en weet niemand wat de data en ICT inrichting wordt van de uiteindelijke oplossing.

Bijvoorbeeld het werk van een data scientist en een beeld van de relevante algoritmen voor de oplossing zal iteratief ontstaan gedurende de analyse activiteiten binnen het project. Dit heeft echter wel invloed op de uiteindelijke configuratie van de oplossing, helder zal zijn dat dit niet in een start architectuur uitgewerkt kan worden bij aanvang van het project. Betekent dit nu dat een start architectuur overbodig wordt? Nee een start architectuur document is nog altijd nodig maar dat zal wel heel anders van opbouw zal zijn. Het zal een veel minder gedetailleerde beschrijving van de oplossing bevatten. De kaders die gesteld worden zijn algemener van aard en er wordt aangegeven hoe ze later in het project concreet gemaakt worden. Ook is het aan te bevelen om aan te geven in welke situaties contact gezocht zal moeten worden met de architect gedurende het project bij welke mijlpalen en welke onduidelijkheden er interactie noodzakelijk is.

Ondersteunen van datateams

In de vorige paragraaf zijn we ingegaan op de up front werkzaamheden en de veranderingen die optreden rond de architectuurdokumentten. Deze verandering heeft eveneens gevolgen voor de werkzaamheden tijdens het project en de ondersteuning van de ontwikkel- en beheerwerkzaamheden. Omdat er een beperkte architectuur wordt opgesteld bij aanvang van een verandertraject zullen de meeste vragen en uitdagingen tijdens de uitvoering van het traject gaan plaatsvinden. Gevolg is dan ook dat een architect tijdens de uitvoering onderdeel dient uit te maken van het ontwikkelteam. Niet om ontwikkelwerkzaamheden uit te voeren maar om de architecturele vragen te beantwoorden met architecturele oplossingen. Dat beantwoorden zal enerzijds bestaan uit de rechtstreekse ondersteuning van het team, anderzijds zullen er detailbeschrijvingen van delen van de

architectuur gemaakt worden. Deze detailuitwerkingen zullen met het ontwikkelteam maar ook met andere stakeholders worden gevalideerd. Gevolg is dat er een Just In Time architectuur ontstaat. Dit is een significant andere aanpak voor de data-architect.

Knelpunten in de traditionele werkwijze

In de bovenstaande paragraaf zijn een aantal aspecten van de dienstverlening van de architect beschreven die problemen op zullen gaan leveren binnen big data projecten. Blijf je als architect op een “traditionele wijze” acteren in de data context dan zal je toegevoegde waarde afnemen. Reden om na te denken wat de knelpunten zijn rond deze werkwijze en op welke wijze je wel toegevoegde waarde kunt leveren aan data gedreven projecten. In de volgende paragrafen gaan we een aantal mogelijke diensten beschrijven waarmee je toegevoegde waarde kunt leveren aan data gedreven projecten.

Nieuwe rolinvulling voor de architect

In het voorgaande hoofdstuk zijn we ingegaan op data initiatieven en de rol van de data-architect. We hebben geconstateerd dat de “traditionele rolinvulling” niet meer volstaat. Echter wat wordt de nieuwe rolinvulling van een architect. In onderstaande paragrafen gaan we in op verschillende producten en diensten waarmee de architect big data initiatieven op adequate wijze kan ondersteunen.

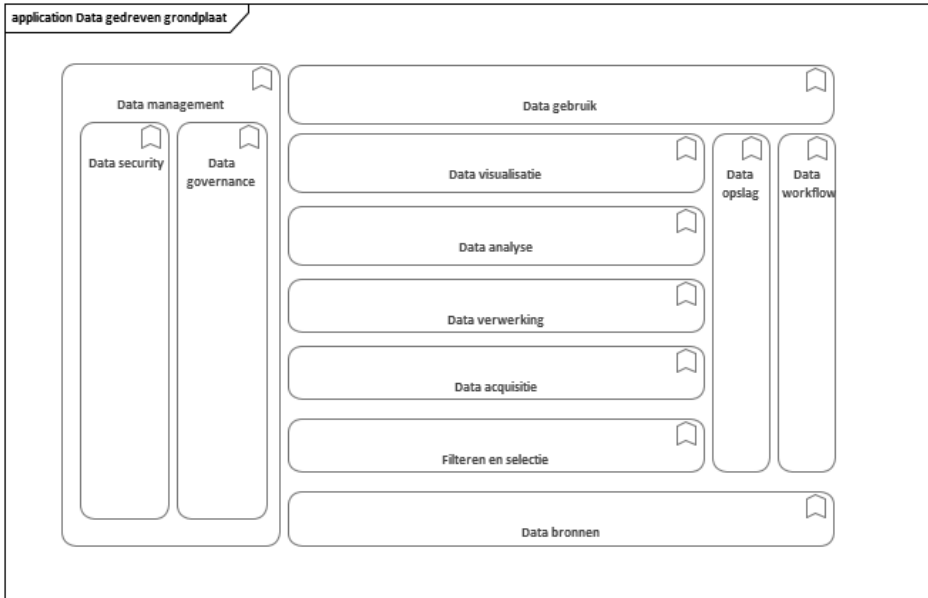
Just in Time Architectuur

Belangrijk aspect van data gedreven initiatieven is dat up front een architectuur opstellen niet goed mogelijk is. Want de activiteiten hebben een innovatief karakter waarbij het onderzoeken van verschillende mogelijkheden centraal staat. Vooraf is daarom ook niet bekend welke onderzoekspaden resultaat gaan opleveren. Het vooraf opstellen van een architectuur is daarmee niet mogelijk.

Betekent dit dat de architect helemaal niets kan voorbereiden? Nee dat is zeker niet het geval! Veel producten en diensten kunnen opgedeeld worden in een “up front” gedeelte en een deel in “Just in time” activiteiten. Idee hierbij is dat het up front als een soort fundament of platform gaat fungeren waarboven op de just in time activiteiten op een efficiënte wijze getest en geïmplementeerd kunnen worden.

Zo zijn er, zonder naar de concrete implementaties te kijken, al veel zaken bekend rond een data gedreven initiatief. Bijvoorbeeld welke bewerkingen er op de data gedaan dienen te worden om oplossingen te implementeren.

Het in kaart brengen van de requirements en constraints van de verschillende stakeholders kan in een vroeg stadium gedaan worden en omgezet worden naar bijvoorbeeld checklists die tijdens de verschillende onderzoeken gevalideerd kunnen worden.



In bovenstaande afbeelding zie je een raamwerk van functionaliteiten die bij big data initiatieven ingezet kunnen worden. Dit model kan voorafgaand aan projecten en onderzoeken opgesteld worden en er kan beschreven worden wat de kenmerken zijn van de verschillende onderdelen. Op dit raamwerk kun je vervolgens principes, requirements en constraints plotten. Maar ook is dit een goed raamwerk om de implementaties binnen projecten op te plotten. Hiermee bied je een eenvoudig oplossing om just in time architectuur te gaan bedrijven waarbij je dergelijke raamwerken aan de start van een project beschikbaar hebt en snel kunt invullen voor een architectuur van het project. Een voorbeeld van een just in time aanpak waar de architect toegevoegde waarde mee levert. Interessant is verder dat dit raamwerk tevens ondersteuning biedt bij data management en de opbouw van de noodzakelijke registers van ontsloten databronnen en van ontwikkelde informatieproducten cq toepassingen. Een activiteit die normaal tijdens de pioniersfase vaak vergeten wordt.