

DATAMODELLERING TOEPASSEN DATA GOVERNANCE

Inleiding

In dit whitepaper wordt een toepassingsgebied beschreven voor datamodellering. Een toepassing is een werkveld op het vlak van architectuur of modellering waarbij een aantal data modelleervormen met elkaar gecombineerd worden.

Deze specifieke modelleervormen zijn beschreven in een serie whitepapers. In de whitepapers over toepassingsgebieden gaan we in hoe de verschillende modelleervormen met elkaar gecombineerd worden ter ondersteuning van dit werkveld.

Deze combinatie maakt het vervolgens mogelijk om op adequate wijze een model te communiceren voor dit toepassingsgebied. In een aantal gevallen wordt alleen documentatie geproduceerd, in andere situaties kunnen ook andere zaken geproduceerd worden zoals source code of templates etc.

Doel

Data Governance is een onderdeel van Data Management. Data Governance is een werkveld dat binnen elke organisatie waar data wordt verwerkt relevant is. Echter bij de ene organisatie is dit werkveld niet of nauwelijks (impliciet) ontwikkeld terwijl andere organisaties hier een tot in detail gewerkte structuur voor hebben uitgewerkt.

Wordt data geproduceerd, verwerkt, opgeslagen en getransporteerd dan is Data Governance relevant. Binnen deze activiteiten worden er bewerkingen gedaan op data waarbij de structuur van deze data en de veranderingen hierin relevant zijn, er zijn eisen rond privacy en beveiliging, er zijn economische- en of financiële vraagstukken rond de data. Rond al deze aspecten en vragen dient er iemand in de organisatie verantwoordelijk te zijn en precies dat aspect wordt ingeregeld met behulp van Data Governance.

Data modellering en Data Governance lijken in eerste instantie weinig met elkaar gemeen te hebben, echter niets is minder waar. Enerzijds is de waarde van de data gelegen in de structuur van de data en de mogelijkheid om de structuur naar behoefte te veranderen. Anderzijds is data voor steeds meer organisaties een productiemiddel en heeft het daarmee waarde voor zowel de organisatie als haar omgeving.

In dit whitepaper behandelen we de Data Governance vanuit het perspectief van het DaMa International, het consortium achter de DaMa Body of Knowledge (DMBoK), waarbij we regelmatig verwijzen naar achtergrond informatie binnen de DMBoK. De hoofdpagina voor deze site is te vinden via <http://dama.org>

Context

Data Governance introduceren is voor sommige organisaties een complex en langdurig proces, bij andere organisaties kost dit minder effort. Dit hangt veelal af van de volwassenheid van de organisatie rond service processen. Is bijvoorbeeld reeds het eigenaarschap van werkprocessen of applicaties ingeregeld dan is het inregelen van data governance relatief eenvoudig door aan te sluiten bij de reeds aanwezige structuren.

Data Governance is een onderdeel van Data Management. Data Management is een aantal aan elkaar gerelateerde bedrijfsprocessen met focus op de diverse aspecten van data. In het DaMa Body of Knowledge wordt het onderstaande raamwerk voor deze bedrijfsprocessen uitgewerkt



Bron: DMBok

Steeds meer organisaties maken de stap om het data governance vorm te geven. Dat is verklaarbaar. Dataverwerking wordt steeds complexer, de waarde van data in de organisatie wordt steeds meer onderkend. Daarnaast neemt de hoeveelheid data waarmee organisaties werken meer en meer toe en ontstaan er big data toepassingen. Als laatste is de ontwikkeling van data analytics en predictive- en prescriptive analytics een belangrijke reden om data governance naar een hoger niveau te brengen.

DOELEN VAN DATA GOVERNANCE

Doelen van data governance zijn met name gericht op het structureren van de activiteiten rond data en op het beleggen van verantwoordelijkheden rond de data. In dit whitepaper lichten we een aantal doelen toe van data governance indien ze relevant zijn voor data modellering

- **Data strategie**, definiëren van strategieën, policies, architectuur , procedures en richtlijnen rond data binnen de organisatie.

- **Waarde van data**, bepalen van de waarde van data voor verschillende datasets binnen de organisatie en het inregelen van maatregelen en processen in de organisatie die de waarde van data op niveau houden of verhogen.
- **Organisatie rond data issues**, bepalen op welke wijze data issues worden opgelost en hoe de verantwoordelijkheden belegd zijn. Hiervoor worden veelal rollen als data eigenaar en data steward in de organisatie geïntroduceerd.

GOVERNANCE EN DATA MODELLERING

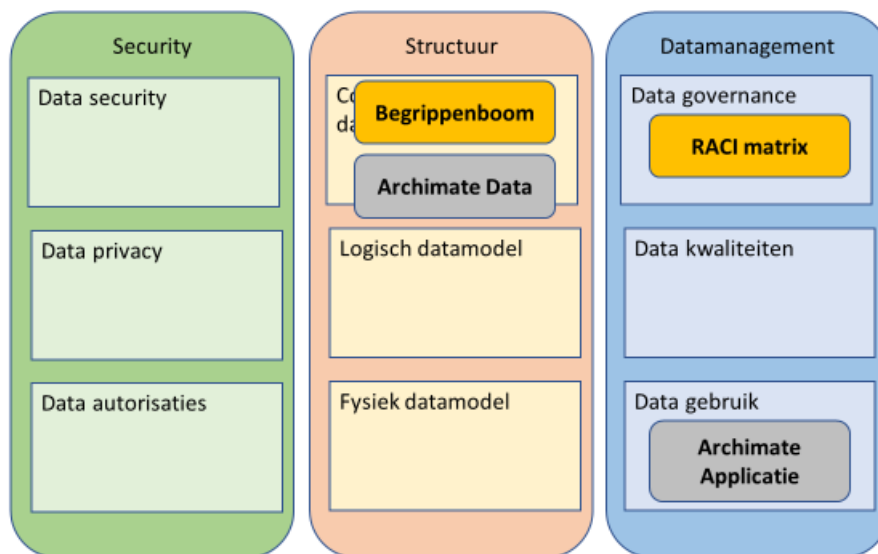
Data governance en met name het toewijzen van data eigenaar en data steward rollen kan op eenvoudige wijze met data modellering worden uitgewerkt. Door dit te combineren met een aantal andere datamodelleerwijzen ontstaat een aantal weergaven cq viewpoints op de data die de invulling van de rol van data steward en data eigenaar eenvoudiger kan maken en verbanden zichtbaar maakt.

Binnen data governance zijn de rollen van data eigenaar en data steward als volgt te definiëren:

- **Data eigenaar**: is eindverantwoordelijk voor de waarde van de data voor die datasets waarvan deze eigenaar expertise/kennis heeft. De data eigenaar is een rol die bij voorkeur binnen de business belegd wordt. Daar is kennis omtrent het gebruik en de waarde van de data aanwezig. De eigenaar rol is voornamelijk strategisch van aard.
- **Data steward**, voert een aantal operationele activiteiten rond de waarde van de data uit. De steward is daarbij gedelegeerd vanuit de data eigenaar. Afhankelijk van de taken die een data steward toebedeeld krijgt ligt deze rol binnen de operatie van ICT of business.

Notatiewijzen

Voor data modellering binnen data governance zijn een aantal notatiewijzen relevant. Een aantal is essentieel, en een aantal is ondersteunend. Onderstaande afbeelding geeft een beeld van de notatiewijzen die vervolgens kort worden toegelicht.



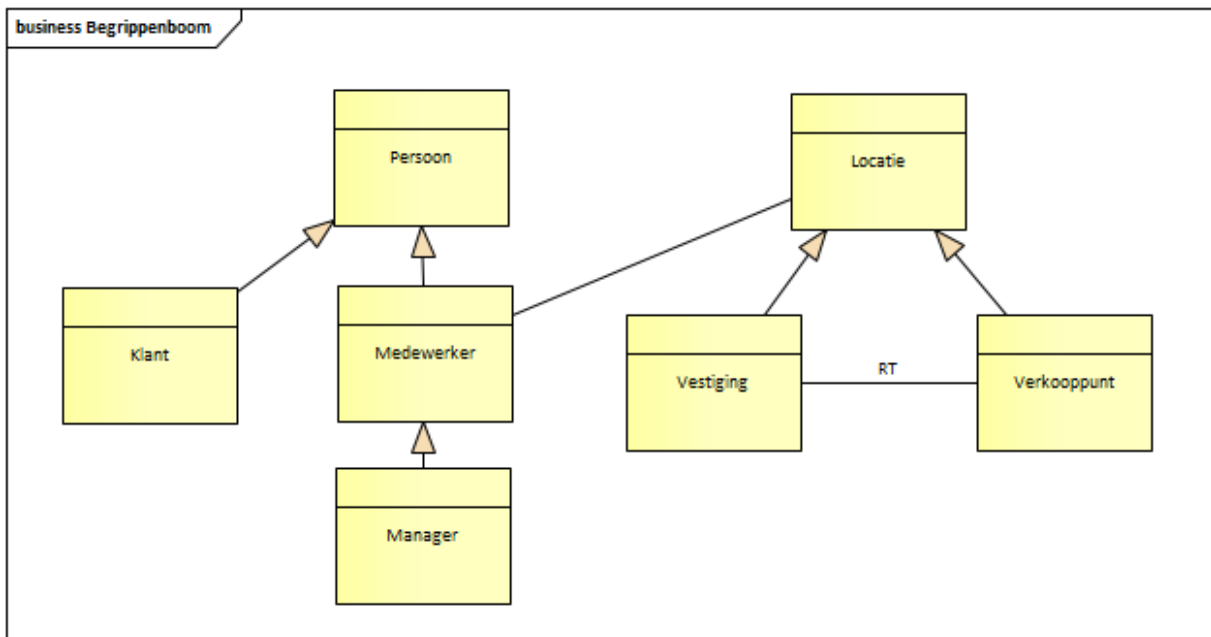
CONCEPTUEEL DATAMODEL

Het conceptueel datamodel is voor data governance een essentieel onderdeel dat zorgt voor de beschrijving welke data entiteiten cq data sets relevant zijn binnen de organisatie. Dit is input voor het definiëren welke rol, zoals onder andere data owner en data steward, een betrokkenheid heeft bij deze data entiteiten. Daarnaast wordt via het conceptueel model de verbinding gelegd naar andere werkvelden zoals security, privacy en quality.

Een belangrijk onderdeel van de rol van data eigenaar en steward is het beschrijven van de definities rond de data concepten in de organisatie. Ook hierbij speelt het conceptuele data model en met name de begrippenboom een belangrijke rol.

Onderstaande afbeelding geeft een beeld van een eenvoudige begrippenboom uitgewerkt binnen de ArchiMate notatie. Meer informatie over de notatiewijze is te vinden via:

<http://assistent.interactory.nl/cmsForm.aspx?formid=50027&webcontentid=249>



Naast het uitwerken van de data entiteiten zal in een aantal organisaties de behoefte bestaan om naast de uitwerking van de begrippen meer detail toe te voegen aan het conceptuele datamodel. Bijvoorbeeld door een onderscheid te maken tussen business objecten en data objecten. Meer informatie over deze notatiewijzen is te vinden via:

<http://assistent.interactory.nl/cmsForm.aspx?formid=50027&webcontentid=248>

DATA GOVERNANCE

Voor het modelleren van de data governance is in eerste instantie van belang om de rol of actor voor het eigenaarschap en het stewardship te definiëren. Dat kan door het opstellen van een lijst maar het opstellen van een ArchiMate business laag model hoort ook tot de mogelijkheden.

Deze bedrijfsrollen of actoren in relatie brengen tot de data entiteiten zoals uitgewerkt in het conceptuele model is de kern van de modellering rond data governance. Hiervoor is de RACI matrix een goed hulpmiddel. Onderstaande afbeelding geeft een voorbeeld van een RACI diagram, detailinformatie over de modelleervorm is te vinden via:

<http://assistent.interactory.nl/cmsForm.aspx?formid=50027&webcontentid=255>

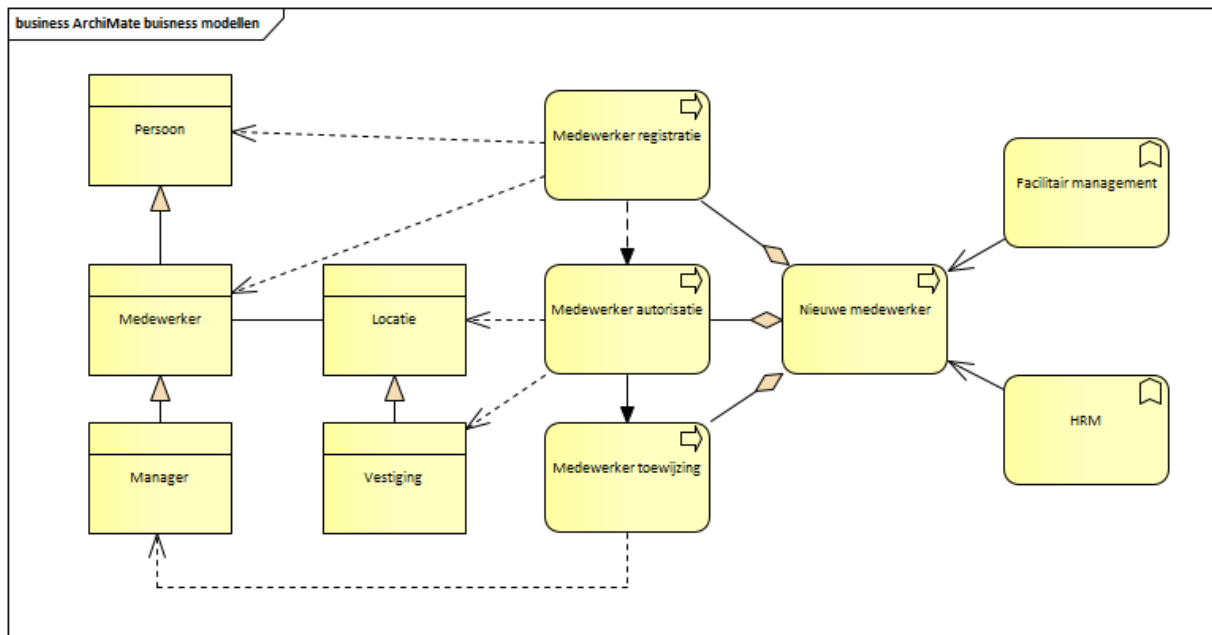
Target \ Source	Manager	Medewerker	Salarisadministratie
Klant	RA	I	CI
Locatie	RA	I	CI
Manager	I		RACI
Medewerker	RA	CI	CI
Persoon	I	I	RACI
Verkooppunt	RA	I	CI
Vestiging	RA	I	CI

DATA GEBRUIK

Voor de data eigenaar en de data steward is het handig als er inzage is waar de datasets in de organisatie worden ingezet. Dat is niet noodzakelijk maar zeker bij een complex landschap of een hoge volwassenheid van de organisatie rond data management zal hieraan steeds meer behoefte ontstaan.

Data gebruik kan op meerdere niveaus gemodelleerd worden, bijvoorbeeld door in kaart te brengen in welke bedrijfsprocessen een data entiteit gebruikt wordt of door te modelleren in welke applicaties een data entiteit gebruikt cq gemuteerd wordt. Onderstaande afbeelding geeft een voorbeeld van een koppeling tussen de bedrijfsprocessen en de entiteiten. Meer informatie over beide notatiewijzen is te vinden via:

<http://assistent.interactory.nl/cmsForm.aspx?formid=50027&webcontentid=244> en
<http://assistent.interactory.nl/cmsForm.aspx?formid=50027&webcontentid=246>



Kenmerken

Data governance komt bij steeds meer organisaties hoger op de prioriteitenlijst te staan. Problemen rond rapportages en compliance kunnen daaraan ten grondslag liggen. Maar ook het creëren van waarde uit data is een reden om een aantal verantwoordelijkheden rond data beter te beleggen

Data governance biedt vanuit data modelleringsperspectief een aantal interessante modelleerbehoefte, met name de combinatie van het conceptuele model met rollen in de organisatie en daarnaast het modelleren van het gebruik van data entiteiten in de organisatie is de kern in een data governance datamodel. Bij de introductie van data modellering van een data governance model zijn de volgende stappen relevant:

- Begin klein en start met een eenvoudig conceptueel datamodel zoals de begrippenboom
- Zorg voor de definities van de begrippen en leg verbanden met behulp van associaties
- Stel een lijst van bedrijfsrollen en -actoren op die de rol van data eigenaar en steward gaan vervullen
- Leg een RACI matrix aan om het eigenaarschap en stewardship te modelleren
- Werk eventueel het data gebruik uit in ArchiMate (core) modellen
- Verfijn het model en breidt het verder iteratief uit.

Producten

De producten voor een data governance vanuit data modelleringsperspectief zijn samengevat:

- Conceptueel datamodel
- RACI matrix
- Model van bedrijfsrollen en -actoren
- Modellen rond het gebruik van datasets in de organisatie

Tooling

Zoals reeds genoemd zijn er rond data governance meerdere producten te vinden, veelal als onderdeel van een data management suite. Gartner heeft hiervoor een aantal documenten opgesteld, een andere aardige ingang is: <https://blog.panoply.io/12-data-management-tools>

Ook generieke tooling kan ingezet worden. Inrichten op basis van Wiki's is een goede mogelijkheid, maak bij de inrichting van een dergelijke omgeving direct rekening met het beheer en de governance van de entiteiten. Een dergelijke omgeving dient namelijk in sync te blijven lopen met de ontwikkelingen binnen de data governance inrichting en dat is geen eenvoudige opgave vanuit beheersperspectief.

Als laatste is het inzetten van generieke (enterprise) architectuurtooling te noemen. Een aantal architectuur tools hebben de mogelijkheid om meerdere modelleertalen met elkaar te combineren waardoor de (data) modelleerbehoefte voor data governance grotendeels kan worden afgedekt.

Evaluatie

Data governance wordt bij steeds meer organisaties een belangrijk werkveld. Inzetten van data governance kan veel redenen hebben, echter vrijwel altijd dient er een antwoord gevonden te worden op problemen rond het eigenaarschap van data.

Binnen data governance speelt data modellering een belangrijke zo niet centrale rol. Met name het leggen van verbanden tussen de data entiteiten en de eigenaren en stewards is essentieel. In een vroeg stadium nadenken welke modelleervormen relevant zijn, hoe deze aan elkaar verbonden worden en hoe de stakeholders daarbij betrokken zijn ondersteunt de introductie van data governance.

In dit whitepaper hebben we een combinatie van modelleervormen beschreven die een (minimale) set is van notatiewijzen op basis waarvan data governance in organisaties gemodelleerd kunnen worden.

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