

EEN REPOSITORY VOOR META DATA MANAGEMENT

Januari 2022

Bert Dingemans

<http://interactory.nl>



INHOUDSOPGAVE

Een repository voor Meta Data Management.....	1
Inleiding.....	3
Meta Data	3
Data Management Body of Knowledge.....	3
Meta Data raamwerk	5
Stakeholders.....	6
Requirements	7
Opzet meta data domein.....	8
Beschrijvend model	10
Conceptueel datamodel	10
Logisch datamodel.....	16
Fysiek data model	21
Kaderstellend model	26
Data principes	26
Data Kwaliteiten	31
Data security en -privacy	36
Toepassingsmodellen.....	40
Data eigenaar en steward	40
Data Bronnen	45
Data Toegang	50
Meta Meta Model.....	55
Tooling & aanvullende producten.....	55
Evaluatie	56
Over de auteur	56

Inleiding

In dit whitepaper gaan we in op hoe we een repository kunnen opzetten voor Meta Data Management. Meta Data is een belangrijk kennisgebied van data management en daarmee relevant voor data gedreven organisaties.

Ten behoeve van Meta Data zijn er diverse Meta Data tools beschikbaar. Denk hierbij aan Collibra, Axon of Talend. Echter bij organisaties zijn veelal ook generieke modellering en tools beschikbaar voor modelleren van software, enterprise architectuur of requirements. In dit whitepaper gaan we in op een framework voor de generieke situatie.

In de volgende hoofdstukken introduceren we meta data als concept, een eenvoudig raamwerk en geven we voorbeelden van meta data modelleren op basis van het uitgewerkte raamwerk. Hiermee kun je in je eigen organisatie meta data introduceren op een natuurlijke manier door die delen relevant voor de situatie binnen de organisatie stapsgewijs te introduceren.

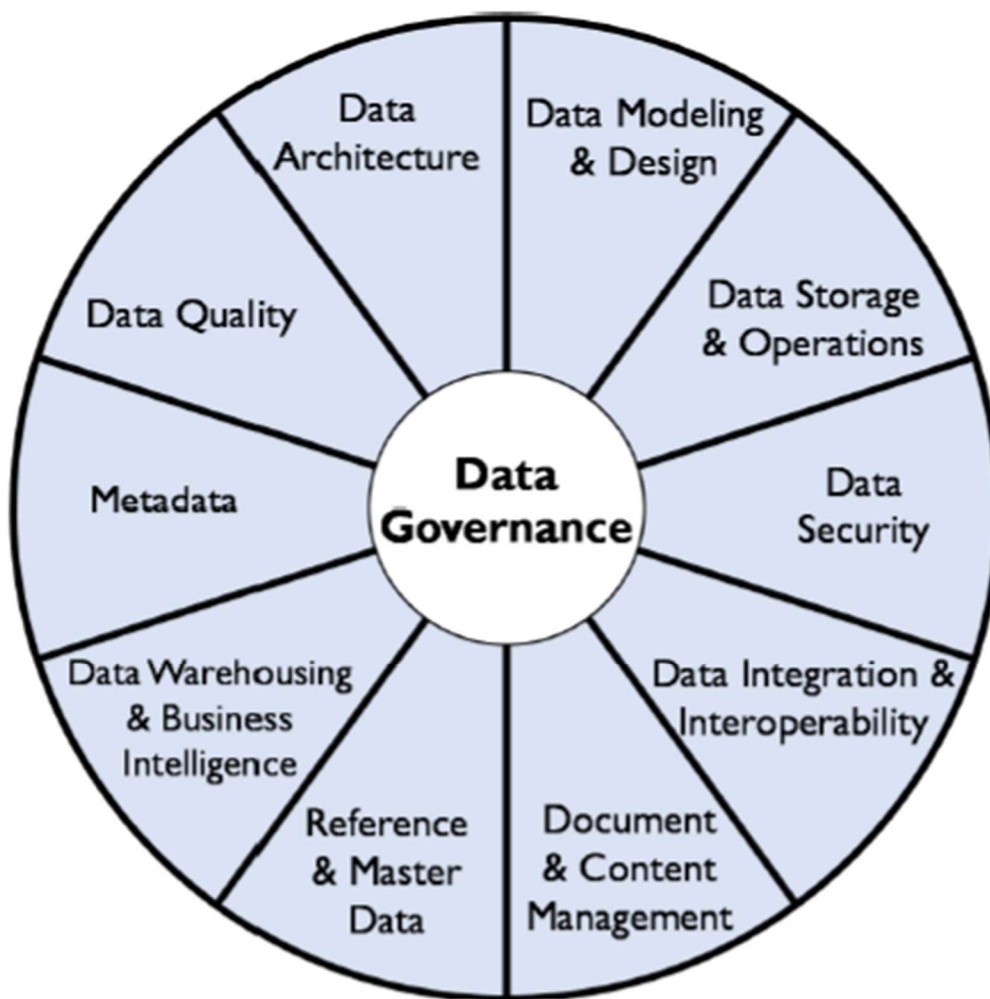
Meta Data

Binnen wikipedia wordt voor meta data de volgende definitie gegeven: *Metadata zijn gegevens die de karakteristieken van bepaalde gegevens beschrijven. Het zijn dus eigenlijk data over data.* Deze definitie is krachtig in haar eenvoud. Vaak wordt van meta data gezegd dat meta data context toevoegt aan data. Het is hierbij van belang welke context dat is. Zo kun je voorstellen dat de meta data van een boek een andere context geeft dan bijvoorbeeld een foto in JPEG formaat of een tabel in een relationele database.

In dit whitepaper is de context van de meta data enterprise data management. Dat betekent dan ook dat we data gaan verzamelen over data entiteiten vanuit het perspectief van data management om context toe te voegen aan deze data entiteiten. Voor data management is een internationaal raamwerk aanwezig name het Data Management Body of Knowledge ontwikkeld door DaMa.org. In dit whitepaper geldt dit raamwerk als startpunt voor ons model. In de volgende paragraaf wordt het DMBOK kort toegelicht.

DATA MANAGEMENT BODY OF KNOWLEDGE

Door de wereldwijde data management community (dama.org) is in de afgelopen jaren een model ontwikkeld in de Data Management Body of Knowledge (DMBoK). Dit is een praktisch uitgewerkt raamwerk met elf kennisgebieden. Hieronder een visuele representatie van het raamwerk en een korte definitie van ieder kennisgebied



Bron: DMBOK

- **Data Governance:** Is het uitvoeren van controle en beheer omtrent het beheer van data assets. Data Governance stuurt alle andere dataprocessen
- **Data architectuur:** Managen, ontwikkelen en beheren van de requirements en principes rond data
- **Data modelleren:** Is het ontdekken, analyseren en beschrijven van data requirements in de vorm van gestandaardiseerde modellen die een data structuur beschrijven
- **Data storage en operations:** 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
- **Reference en Master Data:** Managen van generieke en algemene (herbruikbare) data en referentie data (codelijsten e.d.)

- **Datawarehousing en BI:** 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

In het DMBok is meta data een separaat kennisgebied en is in detail uitgewerkt. Hiermee kunnen we de verschillende data entiteiten binnen een organisatie in de context van de afzonderlijke data management kennisgebieden plaatsen.

META DATA RAAMWERK

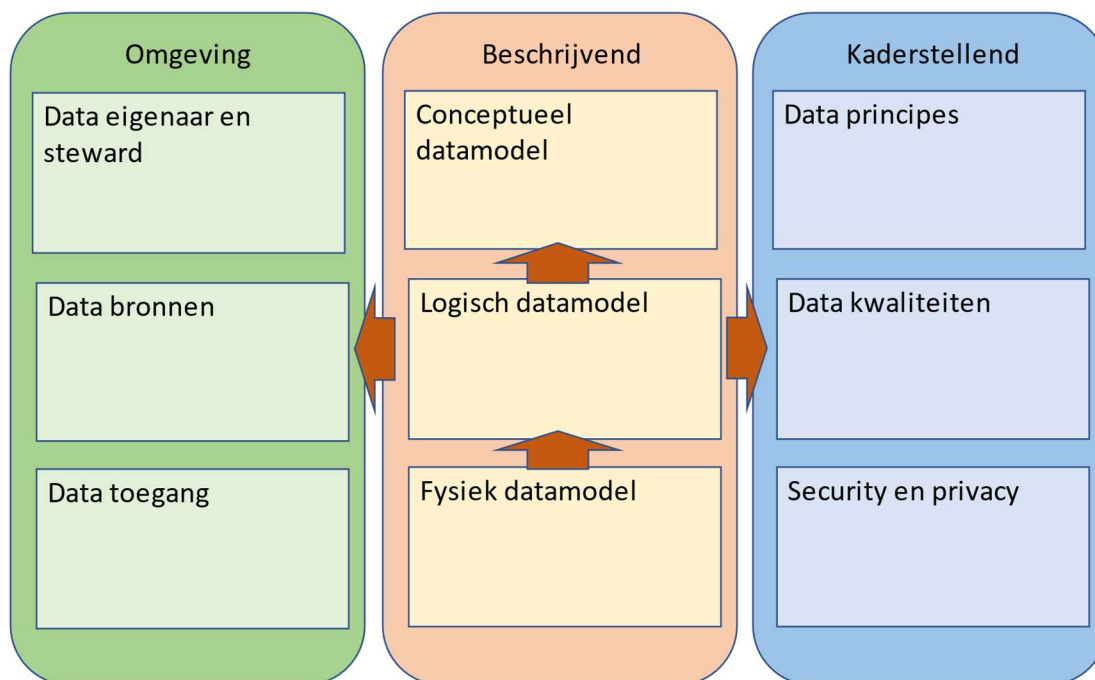
Voor meta data management kan een raamwerk uitgewerkt worden. Met dit raamwerk wordt een registratie gedefinieerd voor de verschillende data management kennisgebieden. Dit raamwerk bestaat veelal uit een groeimodel en is specifiek voor iedere organisatie. Enerzijds vanwege de structuur van de organisatie en anderzijds door de volwassenheid van de organisatie op het vlak van data management.

In onderstaande afbeelding zie je een drie bij drie matrix waarin een beperkt aantal kennisgebieden zijn gedefinieerd. Deze matrix zal als indeling gelden voor de rest van dit whitepaper.

In de afbeelding worden drie categorieën voor kennisdomeinen genoemd namelijk beschrijvend, kaderstellend en omgeving. De onderwerpen binnen deze domeinen kunnen beschouwd worden als een startpunt voor het uitwerken van meta data management vanuit het perspectief van data management. Dit gaan we doen door voor ieder onderwerp een aantal modelleerwijzen te introduceren waarmee de stakeholders zich op eenvoudige wijze een adequaat beeld kunnen vormen van de situatie van dit meta data onderwerp.

In de afbeelding zijn naast de negen kennisdomeinen een viertal pijlen opgenomen. Deze pijlen geven aan dat de kennisdomeinen met elkaar verbonden dienen te zijn. Hierbij neemt het beschrijvende model een centrale plaats in, meer specifiek het conceptueel model. De entiteiten in dit model fungeren feitelijk als centraal koppelpunt voor alle andere aspectgebieden.

In de volgende hoofdstukken zullen we van ieder onderwerp een beschrijving geven van een registratie. Echter deze registratie doen we op basis van een aantal gestandaardiseerde modelleertechnieken zoals ArchiMate, UML Klasse diagrammen en ER Diagrammen. Ook hierbij is het uitgangspunt van een startpunt gebruikt. Bij een verdere uitwerking van het meta data model is het desgewenst eenvoudig mogelijk om extra modelleertechnieken toe te voegen.



Stakeholders

Meta data management modellen richten zich op een grote groep van diverse stakeholders die zich voornamelijk binnen de business en ICT bevinden. Het model dient daarom ontsloten te worden voor stakeholders met grote verschillen in kennisniveau, maar ook in interesse van de verschillende onderwerpen in het meta data raamwerk.

Onderstaande opsomming geeft de belangrijkste stakeholders:

- **Gebruikers van data**, feitelijk is dit iedereen in de organisatie die gebruik maakt van data (wie doet dat niet). Zij stellen daarmee bijzondere eisen aan de eenvoud van het raamwerk omdat zij incidenteel gebruik maken van de uitwerkingen in het raamwerk. Anderzijds kun je verwachten dat deze stakeholder niet bereid is om complexe modellen te analyseren.
- **Data eigenaar en -steward**, met name de eigenaar is eindverantwoordelijke de data entiteiten en de meta data die van de verschillende kennisdomeinen wordt vastgelegd. Dit is een relatief kleine groep met een grote mate van betrokkenheid. Je mag dan ook verwachten dat zij bereid zijn om de verschillende modellen te analyseren en te interpreteren. Desgewenst zijn zij bereid om trainingen te volgen op dit vlak.
- **Functioneel (Applicatie) Beheerders**, een groep stakeholders die diepgaande kennis heeft van (standaard) informatiesystemen en het onderliggende datamodel inclusief de

door de leverancier gebruikte objecten en hun definities. Zij zullen in staat zijn om ook meer complexe modellen van de meta data te interpreteren.

- **Data-, Applicatie- of Informatiearchitect**, zal veelal kennis hebben van, de architectuurdoelen, -kaders, het bedrijfs-, applicatie- en datalandschap in de huidige- en de gewenste situatie. Ze hebben daarnaast veelal diepgaande kennis van data patronen, diagnostische technieken of het opstellen van de verschillende modelleertechnieken.
- **Aspect adviseurs**, bijvoorbeeld security en privacy officers, hebben behoefte aan het ontsluiten van specifieke onderdelen van de meta data. Met name bijvoorbeeld security en privacy classificaties en maatregelen om risico's rond deze aspecten te verminderen. Het is een relatief kleine groep stakeholders met specifieke wensen, echter zij zullen veelal bereid zijn om te investeren in het analyseren van deze specifieke aspectmodellen.

Requirements

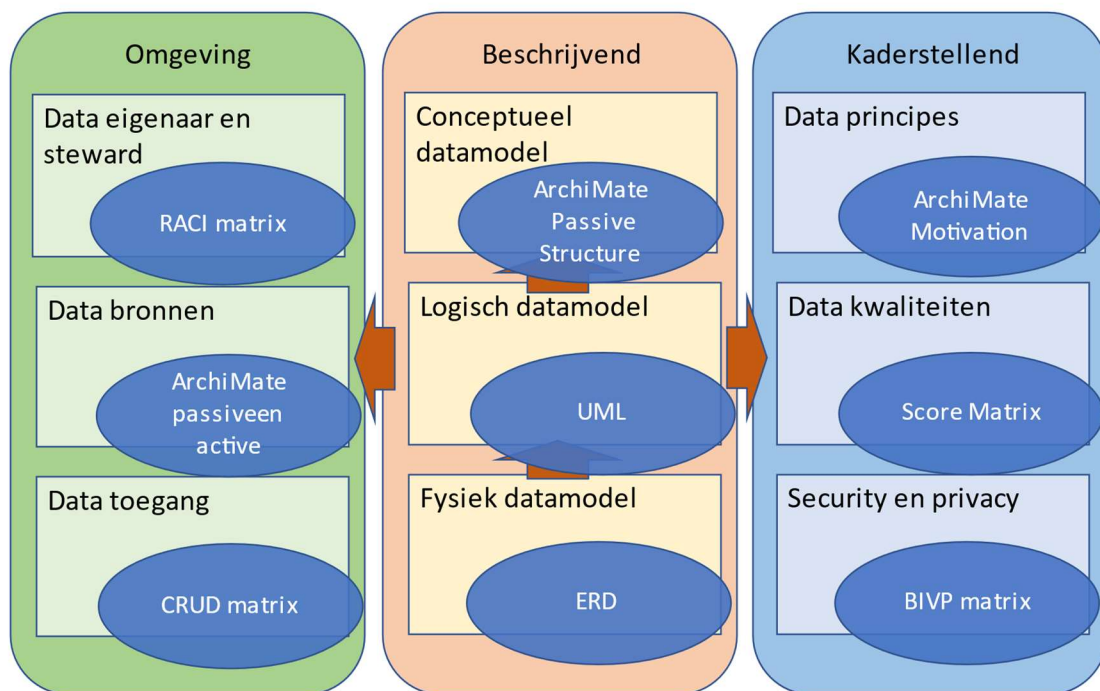
In bovenstaande paragrafen zijn een aantal verschillende kennisdomeinen en stakeholders beschreven. Dit geeft aan dat er een aantal requirements gesteld worden aan het meta data raamwerk en de modelleertechnieken gebruikt als uitwerking voor de meta data onderwerpen in het raamwerk:

- De gebruikte modelleertechnieken dienen analyseerbaar te zijn door verschillende stakeholders
- De modellen zijn bij voorkeur beschreven in een open standaard zoals UML, ER, ArchiMate.
- Er wordt gebruik gemaakt van een beperkt aantal modelleervormen zoals:
 - Lijst
 - Boom
 - Matrix
 - Score matrix
 - Gestereotypeerde matrix
 - Gestereotypeerde graaf
- Modellen vanuit de verschillende meta data domeinen dienen met elkaar verbonden worden
- Er is een verschillende mate van detail mogelijk, globaal voor overzicht en detailanalyse voor ondersteuning van behoeften van de verschillende stakeholders.
- Verschillende vormen van ontsluiting, bijvoorbeeld in PDF documenten of ontsluiting via het web of een wiki.

Opzet meta data domein

Op basis van de drie bij drie matrix en de requirements kan de opzet van het meta data domein gedefinieerd worden. Dit doen we door voor elke cel in de matrix een modelleerconventie te bepalen. Bijvoorbeeld door de keuze van een bestaande modelleervorm of door het selecteren of definiëren van een metamodel voor het uit te werken aspect. Binnen de ArchiMate modelleertaal wordt dit gedefinieerd als een viewpoint.

In onderstaande afbeelding zie je een voorstel voor de uitwerking per aspect. Echter dit is nog een grofmazige indeling van het metamodel. In de volgende hoofdstukken zullen we vervolgens een samenvattend model maken en vervolgens voorbeelden uitwerken van het meta data model voor het desbetreffende aspect.



Vervolgens wordt er op basis van een eenvoudige case een aantal voorbeelden gegeven van het metamodel/viewpoint voor het desbetreffende aspectgebied. Hiervoor maken we gebruik van een raamwerk. Dit raamwerk bestaat uit de volgende onderdelen:

- **Modelleerwijze**, gebruikte modelleerconcepten en -wijze.
- **Modelleertaal**, gebruikte modelleertaal inclusief eventuele viewpoints of andere template inzet.

- **Entiteiten**, gebruikte entiteiten binnen dit aspectgebied inclusief eventuele te gebruiken attributen en andere eigenschappen. Eventueel uitgebreid met een stereotypering en weergave van de entiteit.
- **Associaties**, associaties tussen de bovengenoemde entiteiten, eventueel uitgebreid met een stereotypering en weergave van de associatie.
- **Notatie**, beschrijving van hoe de bovengenoemde entiteiten en associaties gebruikt worden om het aspectgebied te beschrijven.
- **Voorbeeld afbeelding(en)** voorbeeld diagram(men) van het model en de notatie met een beschrijving van onderstaande eigenschappen
 - **Kenmerken**
 - **Toepassingen**
 - **Gerelateerde aspectgebieden en notatiewijzen**

Het voorbeeld wordt uitgewerkt op basis van een model uitgewerkt in Sparx Enterprise Architect. Dit is een eenvoudig voorbeeld van een DatAcademy en geeft een beeld hoe een dergelijke repository opgebouwd kan worden. Voor de geïnteresseerden is een repository bestand beschikbaar. Neem daarvoor contact met me op via email.

In het voorbeeldbestand zijn een aantal extra modellen en modules uitgewerkt als voorbeeld. De drie bij drie matrix is een startpunt voor de inrichting van meta data management, in dit whitepaper noemen we dit het primaire metamodel. In het voorbeeld zijn al andere modellen en modules verder ingericht dat noemen we in dit whitepaper de secundaire modellen. Waar relevant in onderstaande hoofdstukken worden de secundaire modellen ook uitgewerkt.

Beschrijvend model

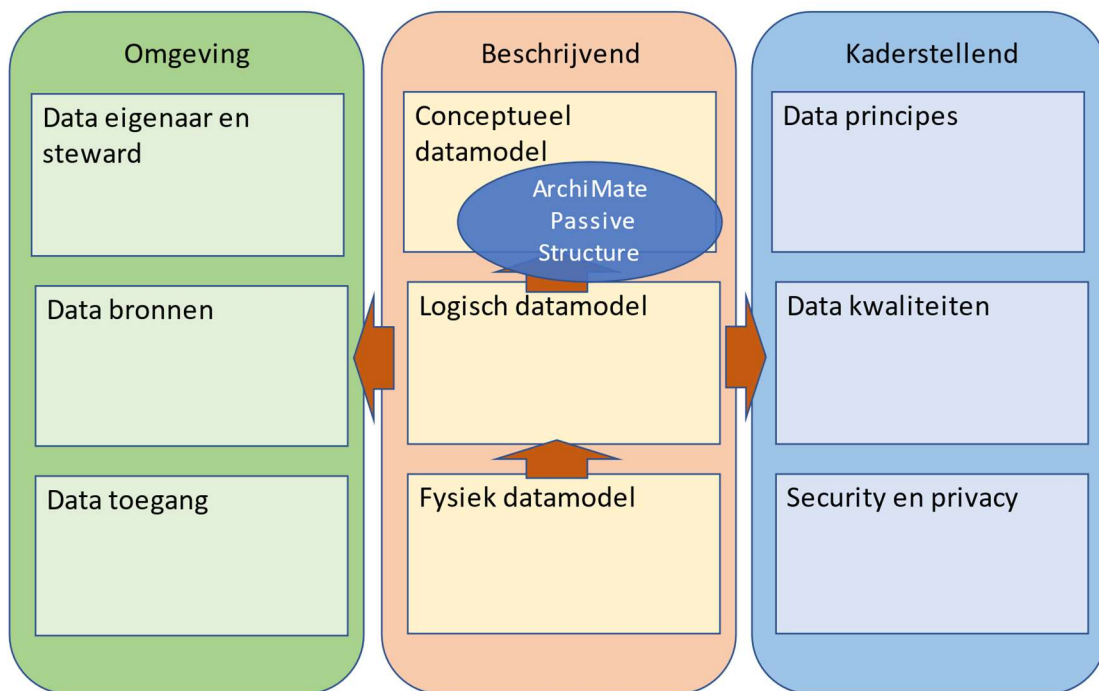
In het beschrijven model worden een aantal data modellen uitgewerkt van abstract (conceptueel model) naar technisch platform specifiek (fysiek model). Dit model is feitelijk het centrale model voor meta data waaraan de andere domeinen gekoppeld zijn.

CONCEPTUEEL DATAMODEL

Het conceptueel data model heeft tot doel om op semantisch niveau de data- of bedrijfsconcepten binnen een domein in kaart te brengen. Dit wordt veelal gedaan door het opstellen van een begrippenlijst en/of een begrippenboom. Dit in kaart brengen bestaat allereerst uit het opstellen van een lijst van begrippen met de daarbij behorende definities of beschrijvingen. Met name in omvangrijke domeinen zoals grote organisaties of clusters van verschillende organisaties is het opstellen van een begrippenlijst (in het Engels veelal Glossary genoemd) een complexe activiteit.

Is een begrippenlijst als conceptueel datamodel opgesteld dan kan vervolgens een begrippenboom opgesteld worden waarbij verschillende hiërarchieën en verbanden tussen de conceptuele data entiteiten gemodelleerd kunnen worden.

Het doel van een conceptueel datamodel is om op semantisch vlak overeenkomst te bereiken bij de betrokken stakeholders wat de betekenis of de definitie van een begrip is. Vandaar dat de combinatie van begrip en beschrijving van het begrip een randvoorwaarde is voor een conceptueel datamodel.



Modellerwijze

Modellerwijze bestaat uit drie verschijningsvormen:

- Begrippenlijst
- Begrippenboom
- Gestereotypeerde graaf

Modelleertaal

Voor het conceptuele model zijn meerdere modelleertalen mogelijk. Denk hierbij aan glossary-, ArchiMate- en UML modellen. Het voorbeeldmodel is uitgewerkt in een ArchiMate model.

Entiteiten

Binnen het conceptuele model staan twee zaken centraal, namelijk:

- **Begrip of concept**, de naam zoals dit binnen de domeinpopulatie gebruikt wordt
- **Definitie**, omschrijving van het begrip waarover overeenstemming is binnen de populatie. Met name die overeenstemming kan een uitdaging zijn. In dat geval zie je nog wel eens een definitie die zo algemeen is dat het altijd geldend is. Dat zal in een later stadium discussie of problemen gaan geven bij de implementatie van het gedetailleerde datamodel.
- **Synoniemen of aliases**, begrippen met dezelfde betekenis binnen het domein maar met een ander begrip.

Associaties

Leggen van verbindingen tussen de bovengenoemde entiteiten:

- **Generalisaties of Broader terms**, met name binnen begrippenboom veel toegepast, hiermee wordt hiërarchie aangebracht in de begrippen. Bijvoorbeeld Vervoermiddel - Fiets - Racefiets is een voorbeeld van een hiërarchie waarbij vervoermiddel een generalisatie is van een fiets.
- **Specialisatie of Narrower term**, vergelijkbaar met een generalisatie alleen dan in de richting van de specialisatie. Dus fiets is een specialisatie van vervoermiddel
- **Relatie of related term**, een associatie tussen twee begrippen die aan elkaar gerelateerd zijn. Bijvoorbeeld doordat zij hetzelfde begrip als generalisatie hebben. Bijvoorbeeld Mountainbike is gerelateerd aan Racefiets.

Notatie

De notatie van de verschillende verschijningsvormen is de volgende

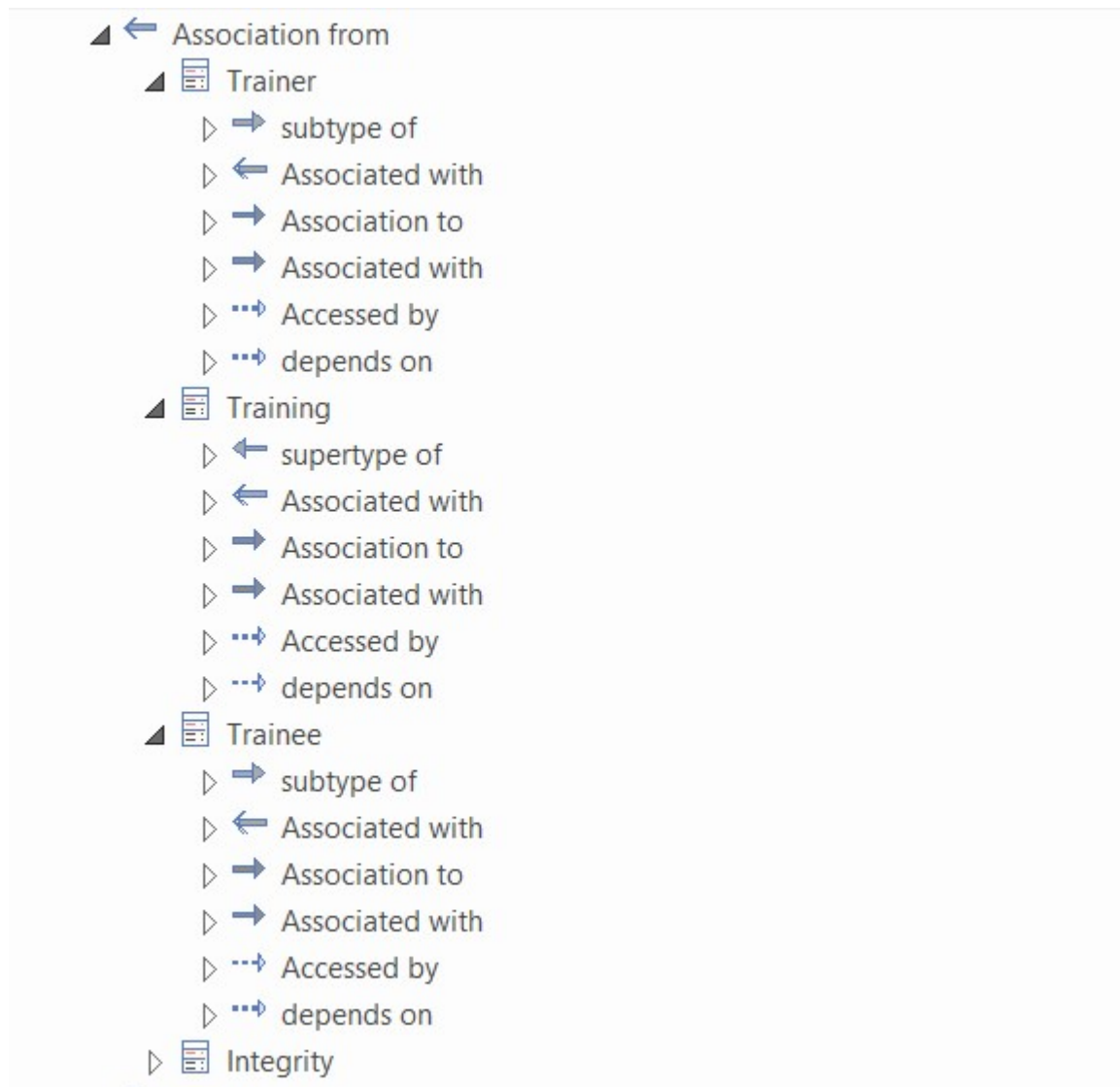
- **Begrippenlijst** bestaande uit een lijst met de naam van het begrip en een definitie
- **Begrippenboom**, een hiërarchische weergave van het conceptueel datamodel in een hiërarchie. Deze is optioneel en met name relevant voor omvangrijke conceptuele data modellen
- **Gestereotypeerde graaf**, weergaven van de hiërarchie van het conceptueel datamodel in een graaf. Eventueel zijn er secundaire modellen in deze weergave mogelijk. Denk hierbij aan het leggen van verbanden van het conceptueel datamodel naar bedrijfsprocessen, applicatielandschap en andere enterprise architectuur modellen. Deze modelleerwijze is optioneel.

AFBEELDING

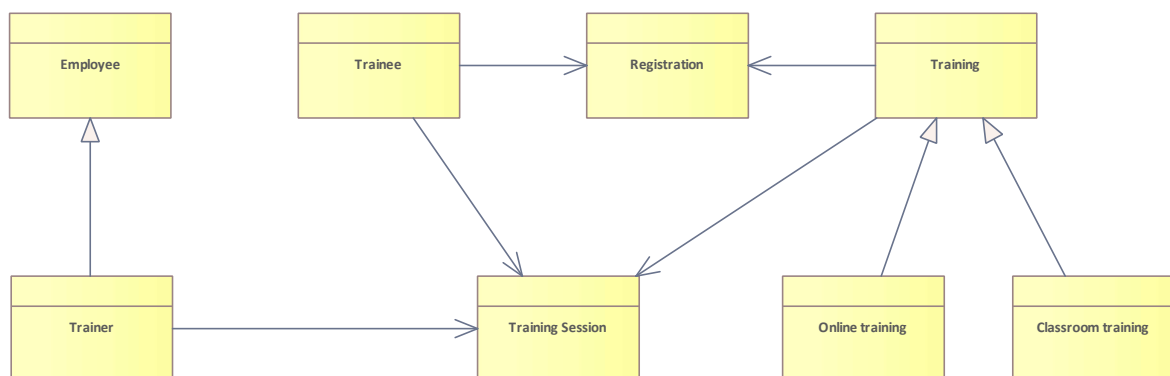
Onderstaande afbeeldingen geven een voorbeeld van de drie hierboven genoemde verschijningsvormen.

Name
Employee An employee is a person that has a function for supporting the trainees in our training company
Training A training is an offer to the market of knowledge transfer about a specific area or subject
Classroom training
Online training
Registration A registration is a record in which is described that a trainee participates in a training
Training Session A training session is part of a training in which the trainer and trainee exchange information and knowledge
Trainer A trainer is an employee who participates in the education activities of our company
Trainee A trainee is a person participating in one or more trainings of our company

Een voorbeeld van een concepten- of begrippenlijst. Een relatief eenvoudige opsomming van het concept en een bijbehorende definitie eventueel uitgebreid met extra kenmerken zoals synoniemen, eigenaar en de status van het begrip.



Een voorbeeld van een hiërarchie van de entiteiten in het conceptuele datamodel. Deze afbeelding is een weergave specifiek voor het gebruikte modelleertool. Echter ook andere modelleertools zullen een dergelijke weergave kennen.



Het voorbeeld is uitgewerkt in ArchiMate, deze modelleertaal bestaat uit een gestereotypeerde graaf waarin op eenvoudige wijze te zien is welke verbanden er zijn tussen de concepten op

basis van verschillende soorten associaties, zoals specialisatie, aggregatie en een associatie. De concepten zijn in dit voorbeeld uitgewerkt als ArchiMate business objecten

Kenmerken

Het conceptueel model (begrippenlijst en begrippenboom) heeft de volgende kenmerken:

- Krachtige notatiewijze waarin begrippen op eenvoudige wijze aan elkaar gerelateerd kunnen worden
- Eenvoudig toepasbaar bij gebruik voor stakeholders zonder modelleerervaring
- Kan goed gebruikt worden in interactieve workshops
- Toepasbaar op hoge abstractie niveaus, voornamelijk conceptueel
- Goed eerste startpunt bij een top down benadering van een objectmodel
- Goed model om discussie op gang te brengen tussen domeinexperts
- Belangrijk hulpmiddel bij het opstellen van datamodellen bij ketenintegraties.
- Hiërarchieën kunnen complex worden als er veel generalisaties worden gebruikt

Toepassingen

Conceptuele datamodellen worden vooral toegepast op een hoog abstractieniveau van datamodellering. Het biedt een goed startpunt voor het in kaart brengen van het gegevensdomein. In complexe domeinen is de begrippenboom een goed startpunt om te komen tot een gezamenlijk domeinmodel waarbij de begrippen de hoogste hiërarchie omvatten.

Houdt er rekening mee dat ondanks de eenvoud van de notatiewijze het opstellen van een conceptueel model een complex traject kan zijn, zeker bij een complex domein of binnen een organisatiecontext waar rond datamodellering weinig volwassenheid is.

Gerelateerde domeinen

Gezien het hoge abstractieniveau van het conceptueel model is het een centraal model voor meta data. In onze drie bij drie matrix zijn alle andere domeinen op enigerlei wijze gerelateerd aan het conceptueel datamodel

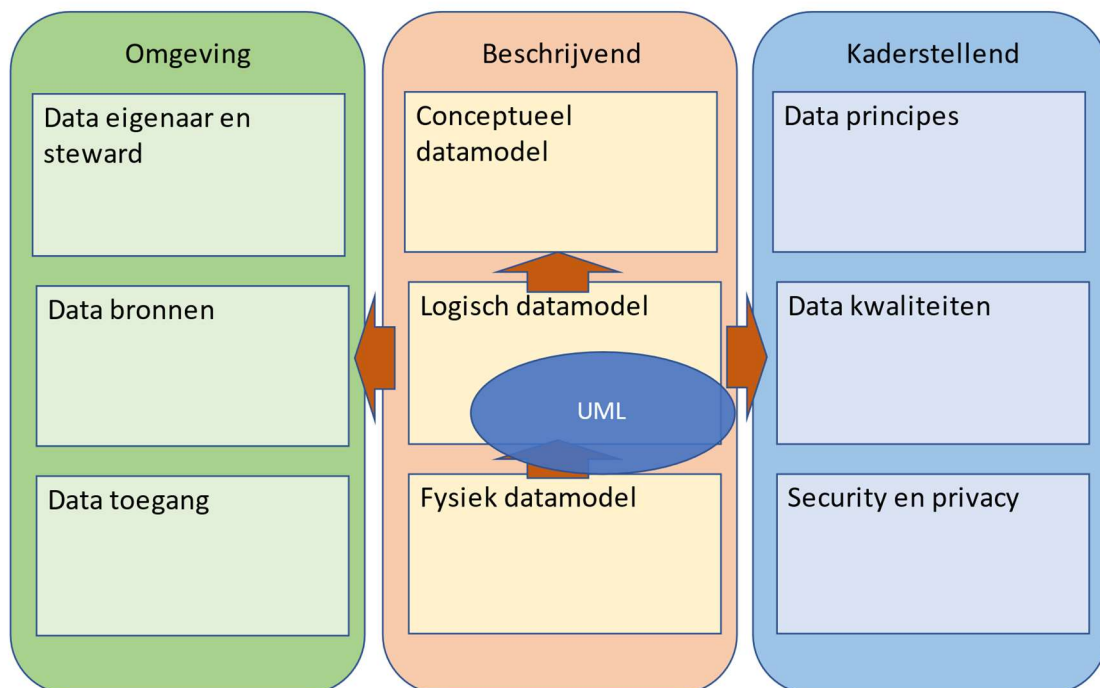
Binnen ArchiMate is in het voorbeeld met behulp van business objects een begrippenboom opgesteld. Deze objecten worden vervolgens als koppelpunt naar bijvoorbeeld bedrijfsprocessen of -functies ingezet.

Ook wordt de begrippenboom veelvuldig gecombineerd met een Logisch Data Model. Hiermee ontstaat een hybride datamodel waarbij de begrippenboom als startpunt dient op een hoog abstractieniveau en logisch datamodel de detaillering van deze begrippen uitwerken in de klassen, attributen en associaties.

LOGISCH DATAMODEL

In deze paragraaf wordt de datamodelleervorm basis UML klassemodel beschreven als modelleerwijze voor logische datamodellering. Deze modelleervorm staat in verhouding tot een aantal andere modelleervormen.

Voor het modelleren van informatie of data is het logisch datamodel. Hierbij is het van belang dat het uitgangspunt is dat het de structuur van gegevens beschrijft. Het logisch datamodel vereist een aantal eigenschappen die ervoor zorgen dat de modellen relatief eenvoudig kunnen blijven (zeker bij basis modellen) maar toch veel zeggingskracht hebben. Dat maakt dat ze geliefd zijn in veel situaties in de informatievoorziening.



Modelleerwijze

In het logisch datamodel worden de concepten vanuit het conceptueel model in detail uitgewerkt met entiteiten, attributen, associaties en een aantal extra uitbreidingen zoals beperkingen, domeinen en cardinaliteiten. Omdat er meerdere associaties mogelijk zijn tussen twee of meer entiteiten is de modelleerwijze een gestereotypeerde graaf.

Modelleertaal

Voor het logisch model zijn er meerdere modelleertalen beschikbaar. De twee meest gebruikte talen zijn Entity Relationship Model en de Unified Modeling Language (UML). In dit whitepaper gebruiken we voor het logisch datamodel de UML (klasse) modelleertaal.

UML is een modelleertaal die voornamelijk in de softwareontwikkeling gebruikt wordt. UML kent veel verschillende diagramtechnieken die in meer of mindere mate met elkaar in verband staan. Een aantal veel toegepaste diagrammen uit UML zijn:

- **Klassediagram**
- **Objectdiagram**
- Sequence diagram
- Component diagram
- Use case diagram

Voor het modelleren van het logisch datamodel zijn eigenlijk alleen het klassediagram en het objectdiagram van belang. Het UML klassediagram heeft tot doel om de structuur van gegevens te beschrijven. Hierbij is het van belang dat het uitgangspunt is dat het de structuur van gegevens beschrijft binnen de organisatie rekening houdend met de behoeften van de verschillende soorten stakeholders.

Entiteiten

- **Klasse:** is het hoofdconcept. Een klasse kenmerkt zich in dat het is opgebouwd uit attributen en dat het een associatie kan hebben met andere klassen. Klassen zijn in natuurlijke taal meestal een zelfstandig naamwoord. Voorbeelden zijn Persoon, Medewerker of Activiteit
- **Attribuut of eigenschap,** zijn de beschrijvende onderdelen van een klasse en voegen hiermee de detaillering toe aan deze klassen. Voor attributen worden een aantal extra kenmerken beschreven, vaak zijn dat het datatype van het attribuut en of het attribuut optioneel is of niet. Een attribuut is in natuurlijke taal veelal een zelfstandig naamwoord, echter kenmerkend is dat er een overkoepelende entiteit is. Voorbeelden van attributen van een persoon zijn voornaam, achternaam of geboortedatum.

Op basis van deze basis concepten zijn reeds zeer krachtige modellen op te stellen. Naast het klasse diagram in de UML basis notatie is het vermelden van het object model relevant. Het object model wordt veelal gebruikt om het klasse diagram concreet te maken met voorbeelden vanuit de organisatie. Bij de notatie paragraaf wordt dit toegelicht.

Associaties

Leggen van verbindingen tussen de bovengenoemde entiteiten:

- **Associatie of relatie,** is een verbinding tussen twee klassen en geeft verbanden tussen logische entiteiten weer. Van een associatie worden net als bij eigenschappen extra kenmerken beschreven. Bijvoorbeeld de cardinaliteit en de rol die een klasse binnen de associatie vervuld.
- **Specialisatie** maakt het mogelijk om een “is een” relatie te leggen. Dus Medewerker is een Persoon. Hierbij zijn er een aantal modelleerconventies opgenomen die de opgestelde modellen vereenvoudigen.
- **Aggregatie,** voor het modelleren van een “heeft” associatie, dus een modelleerwijze om een composiet of samengestelde klasse te modelleren.

- **Compositie**, voor het modelleren van een bestaat uit associatie. Is daarmee sterker dan een aggregatie.

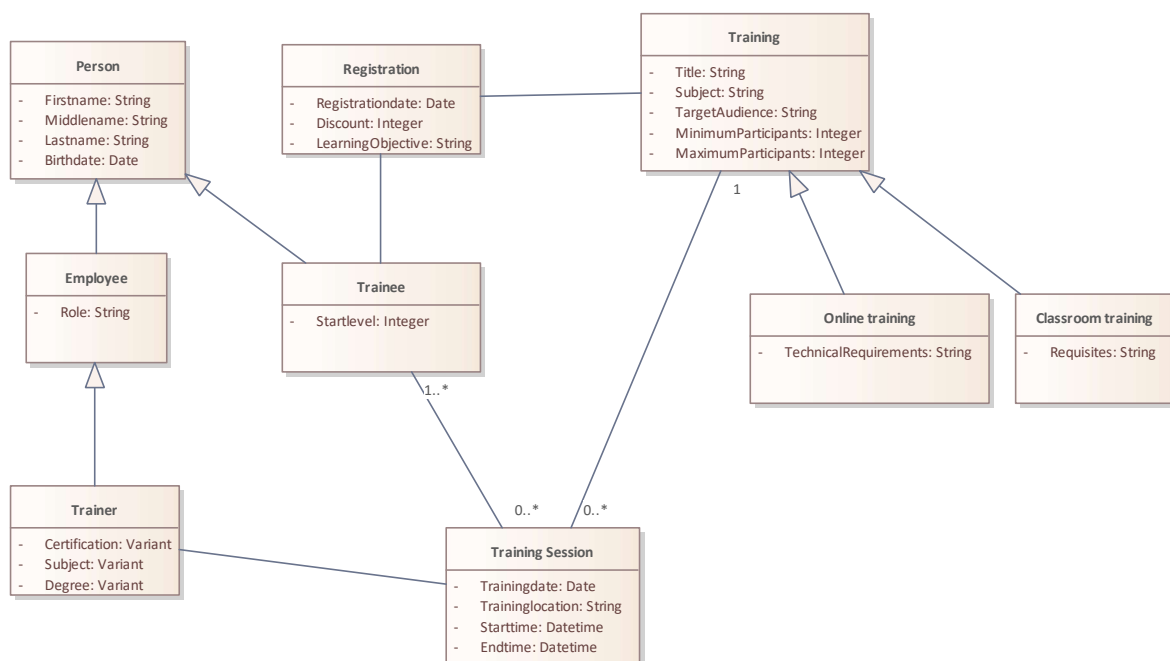
Notatie

Het UML klassediagram is een verrijkte graaf waarbij met name de associaties een aantal verrijkingen kent die de taal de zeggingskracht. Hierbij zijn het klasse diagram waarin een weergave wordt gegeven van de details van de entiteiten waaronder minimaal de associaties (inclusief de genoemde) specialisaties en de eigenschappen van de entiteiten.

Het objectmodel is een vereenvoudigde weergave van het bovengenoemde klasse diagram met name voor de stakeholders met minder ervaring in het lezen van UML klassediagrammen.

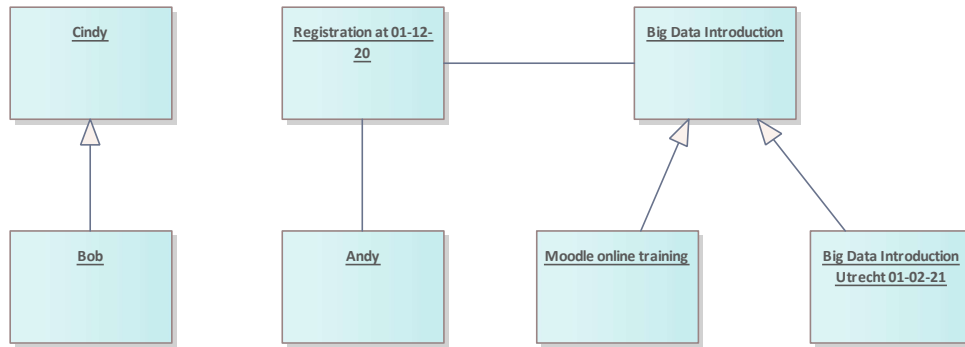
AFBEELDING

Hieronder twee voorbeelden voor logische data modellering



In het voorbeeld is te zien hoe de entiteiten of klassen gemodelleerd worden inclusief de detaillering van de attributen. Van de attributen zijn in dit voorbeeld ook de datatypen gemodelleerd. Desgewenst is dit uit te breiden met extra kenmerken zoals bijvoorbeeld optionaliteit.

Ook voor de associaties zijn een aantal extra kenmerken opgenomen zoals de cardinaliteit of meervoudigheid van de klassen van de associatie. Ook de specialisatie (driehoek op de associatielijn) zie je terug in dit voorbeeld.



Voorbeeld van een objectmodel wat een voorbeeld representatie is met concrete objecten van het klasse model.

Kenmerken

Het logisch datamodel op basis van het UML klassenotatie heeft de volgende kenmerken:

- Een internationale en breed geaccepteerde standaard
- Rijke notatie voor logische datamodellering
- Interactie met leveranciers wordt eenvoudiger
- Het genereren van diverse andere output zoals XSD's
- Eenvoudig om te zetten naar documentatie zoals berichtenboeken
- Notatie is rijk en modellen worden snel complex
- Kans op te grote mate van detaillering
- Het uitwerken van modelleer- en naamgevingsconventies is noodzakelijk
- Het modelleren van relaties met services en processen is niet mogelijk in modellering
- Moeilijk te begrijpen voor Niet ICT'ers

Toepassingen

Het UML klassediagram wordt vooral toegepast binnen de logische data modellering en in een aantal gevallen binnen de fysieke modellering. Binnen dit werkveld is het een zeer krachtige en veel gebruikte modelleerwijze. Het heeft de mogelijkheid om voldoende detail toe te voegen (met name bij de attributen en de associaties).

Een extra gebruikstoepassing van het UML klassediagram is de inzet van deze notatie binnen software modellering en -ontwikkeling. Hierbij wordt het UML klassediagram veelal verrijkt met meer details en het UML klassediagram kan dan vervolgens omgezet worden tot werkende software programmatuur door middel van codegeneratie.

Houdt er rekening mee dat ondanks de eenvoud van de notatiewijze het opstellen van een UMLK basis klassediagram een complex traject kan zijn, zeker bij een complex domein of binnen een organisatiecontext waar rond de informatievoorziening en logische modellering weinig volwassenheid is.

Gerelateerde domeinen

Deze UML klassennotatie heeft met veel notatiewijzen een relatie. In onderstaande opsomming een overzicht:

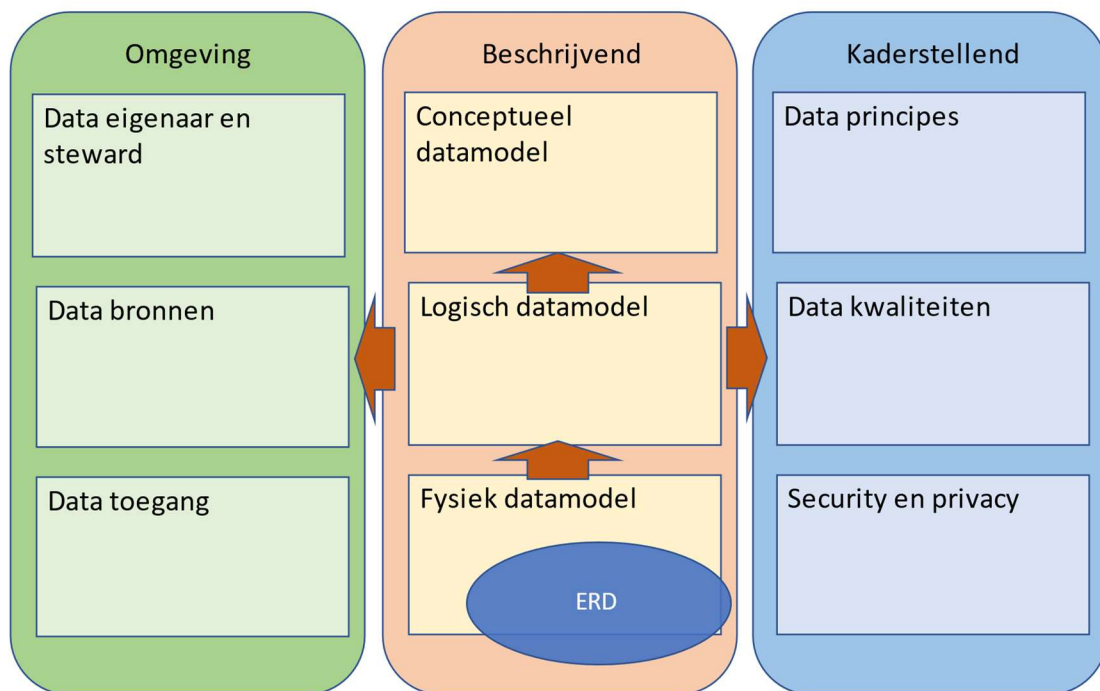
- Conceptueel Datamodel of begrippenlijst en -boom, hiermee worden de verbanden gelegd vanuit het conceptuele datamodel naar het logische datamodel.
- Fysieke datamodellering met de ER notatie waarmee een koppeling gelegd wordt van het logische model in UML naar de fysieke opslagstructuur in een relationele database.
- Fysieke datamodellering met de XSD schema definities waarmee een verband wordt gelegd van het logische model in UML naar een fysieke inrichting voor gestandaardiseerde berichtuitwisseling op basis van XML of JSON.

FYSIEK DATA MODEL

Fysieke datamodellen hebben tot doel om tot in detail de fysieke structuur van een relationele database te modelleren. Veelal wordt daarbij al rekening gehouden met de eisen die vanuit een specifiek databaseplatform gesteld worden aan de inrichting van een database. Hierdoor is het mogelijk om deze ER modellen rechtstreeks om te zetten naar een fysieke inrichting in een database. Dit staat bekend als de Data Definition Language (DDL).

Daarnaast is het mogelijk om extra fysieke modellen op te stellen voor specifieke platformen of implementaties. Denk hierbij aan berichtuitwisseling met XML of JSON berichten of datawarehouses die een specifiek fysiek datamodel kennen.

ER staat voor Entity-Relationship en ER diagrammen geven een gedetailleerde notatiwijze voor het modelleren van deze twee concepten. Daarnaast kun je met ER diagrammen de kolommen in een tabel modelleren en eventueel een aantal belangrijke constraints in relationele databases zoals primaire en verwijzende sleutels.



Modellerwijze

In het fysieke datamodel worden de klassen uit het logisch model in detail uitgewerkt met technische entiteiten zoals tabellen, kolommen in tabellen en primaire en verwijzende sleutels. Omdat er meerdere associaties mogelijk zijn tussen twee of meer entiteiten is de modellerwijze een gestereotypeerde graaf.

Modelleertaal

Voor fysieke datamodellen zijn meerdere platform specifieke talen aanwezig. De ER diagrammen worden met name gebruikt voor database modellering. De XML Schema Definitions worden gebruikt voor het modelleren van (semi) gestructureerde berichten in data integratie.

Entiteiten

Binnen ER diagrammen worden voornamelijk fysieke datamodellen weergegeven in een verrijkte graaf. ER diagrammen zijn opgebouwd uit de volgende concepten:

- **Tabel:** is het hoofdconcept. Een tabel kenmerkt zich in dat het is opgebouwd uit kolommen en dat het een relatie kan hebben met andere tabellen. Tabellen zijn het hoofdconcept voor het opslaan bij elkaar behorende gegevens in een relationele database. Het is de fysieke implementatie van het logische concept klasse.
- **Kolom,** zijn de onderdelen van een tabel die ervoor zorgdragen dat de eigenschappen uit een klasse (in het logische model) binnen een tabel van elkaar onderscheiden kunnen worden. Kolommen worden bij de implementatie gebruikt bij het ophalen en wegschrijven van gegevens uit een database. Bijvoorbeeld binnen Select en Insert/Update/Delete commando's. Voor een kolom kunnen in een ER diagram extra gegevens worden bijgehouden. Bijvoorbeeld het datatype van de inhoud maar ook of de kolom null (leeg) waarden mag bevatten.

Associaties

Leggen van verbindingen tussen de bovengenoemde entiteiten:

- **Relatie,** is een verbinding tussen twee tabellen en geeft verbanden tussen fysieke entiteiten weer. Het draagt er zorg voor dat bij de implementatie in een relationele database deze afdwingt dat de relaties aan de eisen voldoen van deze database. Hiermee wordt ervoor zorggedragen dat verwijzingen van de ene tabel naar de andere altijd correct zijn.
- **Primary keys of primaire sleutel,** dragen zorg voor het uniek identificeren van rijen in een tabel. Op basis hiervan zijn tabellen op correct wijze aan elkaar te verbinden. Is een belangrijk aspect van tabellen in een relationele database.
- **Foreign keys of verwijzende sleutels,** bij bepaalde relaties in een relationele database wordt gebruik gemaakt van een ouder - kind relatie. Verwijzende sleutels dragen zorg dat in de kind tabel op correcte wijze een verwijzing is opgenomen naar de sleutel in de ouder-tabel. De database bewaakt te allen tijde de correctheid van deze verwijzingen.

Notatie

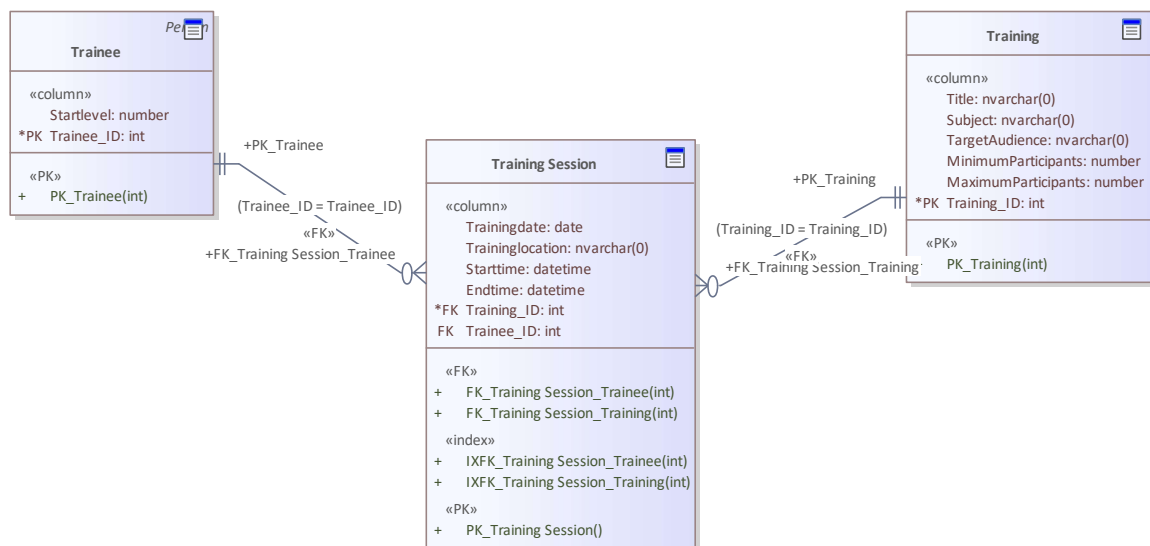
Het ER diagram is een verrijkte graaf waarbij met name de relaties een aantal verrijkingen kent die de taal de zeggingskracht geeft. In onderstaande afbeelding een voorbeeld van een ER diagram.

In de notatie bestaat een tabel bestaat uit een naam en dat een tabel kolommen bevat. Voor de kolommen geldt dat het data type getoond worden en de cardinaliteit [0..1] als het optioneel is of verplicht [1..1].

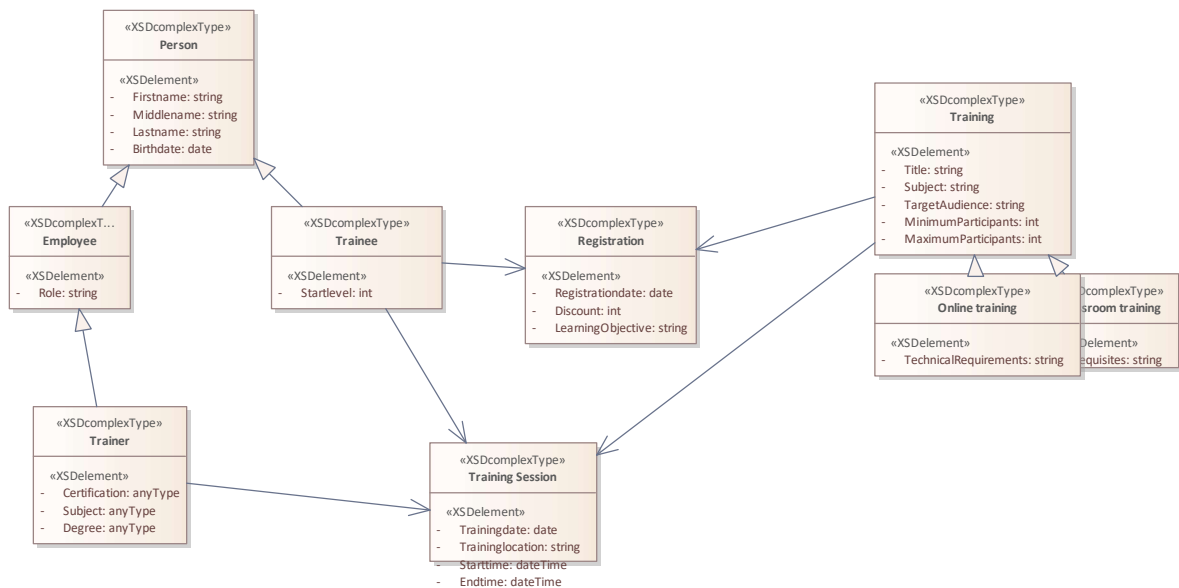
Voor de relaties zijn er meerdere extra kenmerken te zien. Bijvoorbeeld welke rol de tabel vervuld in de relatie (veelal op basis van primary keys en foreign keys). Daarnaast wordt voor de relatie aangegeven wat de cardinaliteit is. Hiermee kan een database zorgdragen voor het correct houden van de tabelinhoud met de relaties tussen tabellen binnen de database.

Als laatste zijn de indexen te noemen. Dit zijn beperkingen in tabellen die gemodelleerd kunnen worden in een ER diagram. In bovenstaande afbeelding zijn deze constraints opgenomen onder de kolommen.

AFBEELDING



In deze afbeelding zie je de tabeldefinitie, bovenin staat de naam van de tabel, er is een sectie voor de specificatie van de kolommen. Verder de extra elementen zoals sleutels en indexen in een eigen sectie.



Deze afbeelding is een voorbeeld voor een XML Schema Definitie, een fysiek model voor berichtuitwisseling.

Kenmerken

Het fysieke datamodel heeft de volgende kenmerken:

- Een internationale en breed geaccepteerde standaard
- Rijke notatie met name voor fysieke datamodellering
- Interactie met leveranciers wordt eenvoudiger vanwege standaardisatie
- Het genereren van diverse andere output zoals DDL scripts voor databases of schema's voor berichtuitwisseling.
- Notatie is rijk en modellen kunnen complex worden
- Grote mate van detaillering
- Mogelijkheid om bestaande relationele databases geautomatiseerd om te zetten naar een ER model of het genereren van berichtdefinities op basis van XSD
- Het modelleren van relaties met niet relationele entiteiten is niet mogelijk in modellering
- Moeilijk te begrijpen voor Niet ICT'ers

Toepassingen

Het ER diagram wordt vooral toegepast binnen de fysieke data modellering en in een aantal gevallen binnen de logische modellering. Binnen dit werkveld is het een zeer krachtige en veel gebruikte modelleerwijze. Het heeft de mogelijkheid om voldoende database specifieke details toe te voegen (met name bij de relaties, de kolommen en de sleutels).

Een extra gebruikstoepassing van het ER diagram is de inzet van deze notatie binnen database modellering en -ontwikkeling. Hierbij wordt het ER diagram omgezet tot een daadwerkelijke database implementatie of omgekeerd een database implementatie wordt omgezet tot een ER diagram.

Houdt er rekening mee dat ondanks de eenvoud van de notatiewijze het opstellen van een ER diagram een complex traject kan zijn, zeker bij een complex domein of binnen een organisatiecontext waar rond de informatievoorziening en logische modellering weinig volwassenheid is.

Gerelateerde domeinen

Deze ER notatie heeft met veel notatiewijzen, met name in de fysieke en logische modellering een relatie. In onderstaande opsomming een overzicht:

- **Conceptueel datamodel op basis van ArchiMate Datamodellering** of begrippenlijst en -boom, hiermee worden de verbanden gelegd vanuit het conceptuele datamodel naar het logische datamodel en desgewenst het fysieke datamodel.
- **UML klassediagram** waarmee een koppeling gelegd wordt naar het logische model in UML klassediagram vanuit een fysieke opslagstructuur in een relationele database.
- **XSD schema** waarmee een verband wordt gelegd van het fysieke datamodel (veelal via een logisch model in UML) naar een fysieke inrichting voor gestandaardiseerde berichtuitwisseling.
- **CRUD matrix**, voor het modelleren van welke persoon, rol of database user of -group toegang krijgt tot de gegevens in een tabel.

Kaderstellend model

Vanuit diverse werkvelden kunnen kaders gesteld worden aan de productie en het gebruik van data. Denk bijvoorbeeld aan data architectuur, security en privacy maar ook aan het stellen van kwaliteitseisen aan data.

Kenmerkend hierbij is dat de kaders richting geven aan het gebruik van de data, maar ook aan andere aspecten zoals de opslag, het gebruik en incidenteel ook aan de wijze waarop de data geproduceerd wordt.

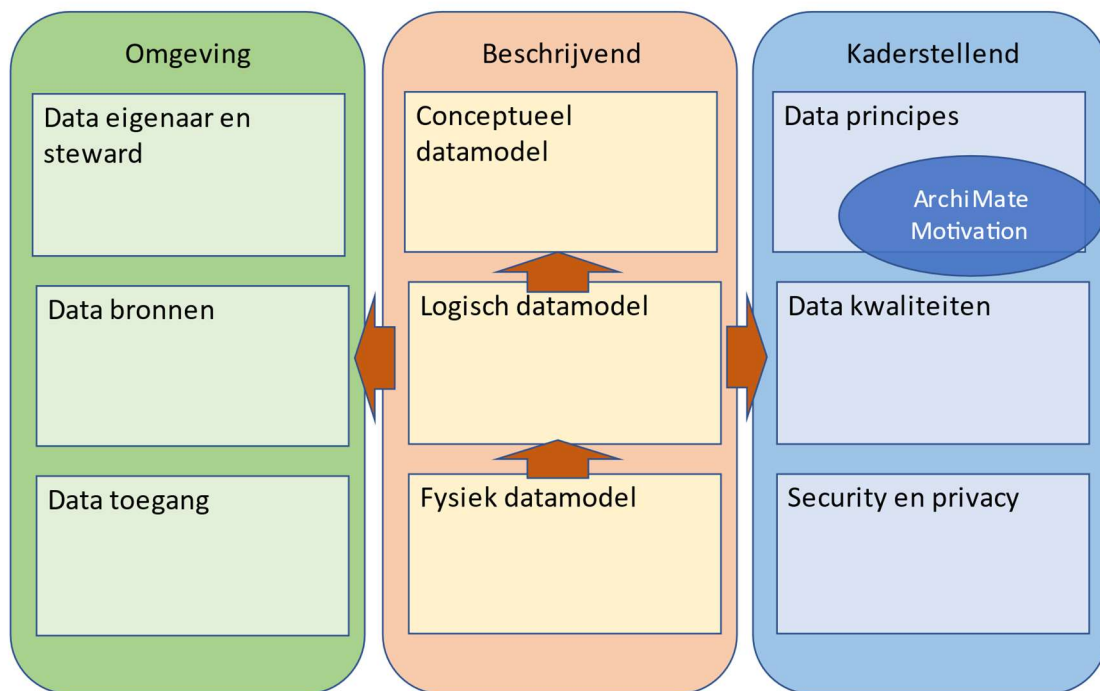
Kaders kunnen op meerdere wijzen beschreven worden bijvoorbeeld als risico, eis of beperking. In dit whitepaper sluiten we aan op de modelleerwijze van ArchiMate en werken we de kaders uit op basis van requirements en principes. Waarbij we principes beschouwen als generalisaties van de requirements.

Naast de kaders zul je binnen een metadata management model ook een uitwerking zien van de maatregelen die genomen kunnen worden om de gestelde kaders te realiseren. We zullen hierbij een aantal eenvoudige voorbeelden uitwerken in de volgende paragrafen.

DATA PRINCIPES

Data principles modellering heeft tot doel om de relatie van bedrijfsentiteiten, data entiteiten en artefacten met stakeholders en hun concerns te beschrijven. Hierbij wordt een duidelijk onderscheid gemaakt tussen de entiteiten in de architectuurlagen en worden er meestal per laag afzonderlijke modellen opgesteld. Hierbij wordt met behulp van principes een verbinding gelegd van de data entiteiten, bijvoorbeeld in het conceptueel datamodel naar de behoeften en doelen van de stakeholders binnen en buiten de organisatie.

Het doel van deze modelleerwijze is om deze verbanden tussen stakeholders en hun concerns in relatie te brengen met de data entiteiten op verschillende lagen. Daarnaast wordt met de principes veelal ook de definitie uitgewerkt op basis van de voor deze concepten geldende standaarden. Voor de data entiteiten is reeds een beschrijving opgesteld binnen de primaire modellering.



Modelleerwijze

Voor het modelleren van principes kunnen lijsten, bomen en gestereotypeerde grafen gebruikt worden. Zeker in de situatie waarin ook de doelen en stakeholders gemodelleerd worden zal veelal de gestereotypeerde graaf gebruikt worden.

Modelleertaal

Voor de data principes is er binnen ArchiMate een extensie beschikbaar namelijk de motivation extensie. Hierin zitten onder andere de voor ons relevante elementen zoals principe en stakeholder.

Entiteiten

Binnen de ArchiMate motivation modellering wordt gebruik gemaakt van grafen bestaande uit entiteiten en associaties. Deze concepten worden afzonderlijk beschreven:

Basisentiteiten

- **Conceptuele datamodel met Businessobject**, dit zijn de entiteiten op het hoogste abstractieniveau in deze modellering. Deze entiteiten worden gerelateerd aan de andere entiteiten in de motivation extensie zoals principes, requirements en stakeholders.
- **Data objecten**, beschrijven de data entiteiten waaruit een business object is opgebouwd. Een business object kan namelijk zijn opgebouwd uit meerdere data

objecten die vervolgens door verschillende logische en fysieke applicaties benaderd worden. Dit is een secundair modelleer element en wordt alleen toegepast in omvangrijke of complexe metadata modellen.

- **Artefacten**, zijn de fysieke implementaties van de data objecten en zijn daarmee nog meer technisch gericht en beschrijven hoe de data objecten zijn opgenomen in bijvoorbeeld tabellen en berichten en in welke systeemsoftware. Ook dit is een secundair modelleerelement.

Motivation-entiteiten

- **Stakeholder**, dit zijn de entiteiten die groepen van bedrijfsrollen die een aantal concerns hebben vertegenwoordigen. Deze entiteiten worden gerelateerd aan de data entiteiten in de drie lagen.
- **Principe**, is een algemene stelling die veelal meerdere abstracte requirements en concerns aggregereert tot één stelling. Heeft een gestandaardiseerde opbouw. Veelal zijn principes getrapt ingedeeld en hebben daarmee een overeenkomst met een begrippenboom.
- **Requirement**, is een beschrijving van de concerns van de stakeholders uitgedrukt in generieke behoeften. Ook dit is een secundair modelleerelement
- **Constraint**, komt meestal voort uit een of meerdere requirements en is een eis of een beperking die gesteld wordt aan een of meerdere data entiteiten. Ook dit is een secundair modelleerelement

Associaties leggen verbindingen tussen de bovengenoemde entiteiten:

- **Associatie**, tussen al deze entiteiten kan de meest generieke verbinding gelegd worden. Deze kan gebruikt worden om een associatie op basis van een rol te leggen
- **Aggregatie en compositie**, hiermee wordt het mogelijk om samenvoegingen te maken van entiteiten die andere entiteiten omvatten. Bijvoorbeeld een Order is een aggregatie van Orderregels
- **Specialisatie**, waarmee boomstructuren opgebouwd kunnen worden. Hiermee kunnen hiërarchische structuren opgebouwd worden zoals de begrippenboom.
- **Realisatie**, tussen data- en bedrijfsobjecten kan een realisatie associatie gelegd worden waarmee een data object een bedrijfsobject creëert.
- **Access** waar mee een data object toegang heeft tot een artefact in de technische laag.
- **Influence**, geeft aan hoe een motivation concept de data concepten beïnvloed, dit kan zowel positief als negatief zijn.

Notatie

In de notatie van de ArchiMate motivation datamodellering wordt er een model opgesteld dat de data entiteiten combineert met stakeholders en de bijbehorende concern concepten zoals requirements en principes. Alle modellen zijn gebaseerd op grafen. Scorematrix is een bijzondere vorm van representatie waarin veelal motivation concepten worden gelinkt naar data entiteiten en er een kwantitatieve score wordt gegeven. Deze wordt als secundair beschouwd.

Gebruiken we een weergave van samengestelde entiteiten dan wordt er veelal een combinatie gemaakt van de relevante data entiteiten en de motivation concepten. Zie de afbeelding in de volgende paragraaf.

AFBEELDING

A central design authority governs the data sources and the data integration layers

A central design authority governs the data sources and the data integration layers and design and technology decisions on these layers are with this design authority only

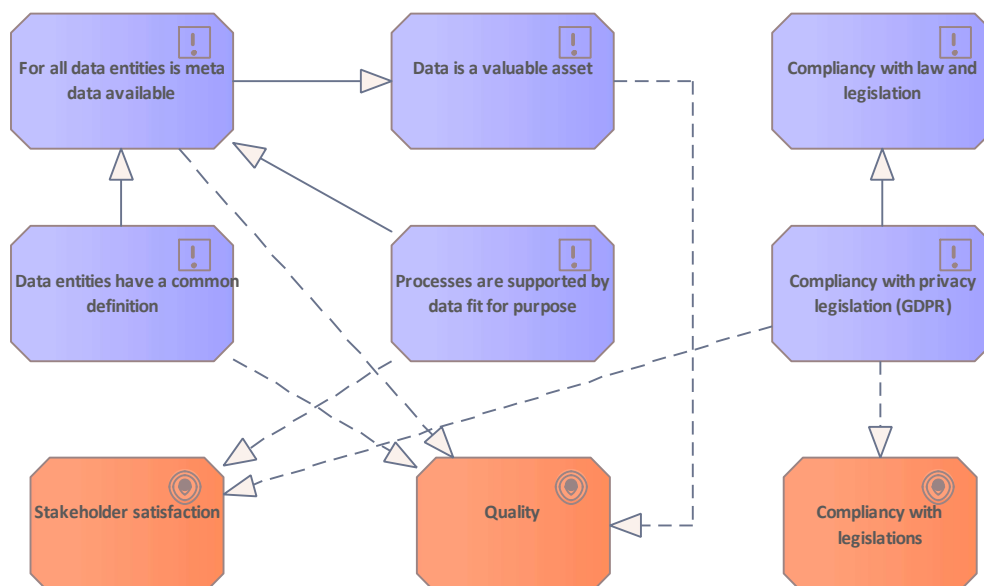
Ability to abstract on the data sources via implementation of specific connectors

Connection and integration of data sources of different types and ability to abstract on the data sources via implementation specific connectors, e.g. for smart meters, wind craft, etc.

Cloud and on-premise deployment options

Cloud and on-premise deployment options with ability to migrate from cloud to on-premise (cloud is preferred for short term temporary solution)

Afbeelding met een eenvoudige lijst van principes. Vaak wordt er een gestandaardiseerd format voor principes gebruikt. Dit model bestaat dat uit een statement, rationale en implicaties. In dit model is hier niet voor gekozen.



In deze afbeelding zie je enerzijds een hiërarchie van principes gemodelleerd met de specialisatie associatie. Anderzijds zie je hoe een aantal principes een doel van de organisatie realiseren.

Kenmerken

De ArchiMate principe datamodellering is een abstracte maar krachtige notatiewijze met de volgende kenmerken:

- Krachtige notatiewijze waarin stakeholders, doelen principes en data entiteiten op eenvoudige wijze aan elkaar gerelateerd kunnen worden
- Eenvoudig toepasbaar bij gebruik voor stakeholders zonder modelleerervaring
- Kan goed gebruikt worden in interactieve workshops
- Toepasbaar op hoge abstractie niveaus, voornamelijk conceptueel in combinatie met andere viewpoints
- Mogelijkheid om verbanden tussen de behoeften in de organisatie en het data model
- Goed model om discussie op gang te brengen tussen domeinexperts
- Hiërarchieën kunnen complex worden als er veel generalisaties worden gebruikt, bijvoorbeeld bij principe modellen.

Toepassingen

Principe modellering wordt vooral toegepast op een hoog abstractieniveau van datamodellering. Het biedt een goed startpunt voor het in kaart brengen van het bedrijfsobjecten en dataentiteiten en stakeholders en concerns. In complexe domeinen is met name een conceptueel data model in combinatie met principes een goed startpunt om te komen tot een gezamenlijk domeinmodel waarbij de principes de hoogste hiërarchie omvatten.

Houdt er rekening mee dat ondanks de eenvoud van de notatiewijze het opstellen van een principe model een complex traject kan zijn, zeker bij een complex domein of binnen een organisatiecontext waar rond principe modellering weinig volwassenheid is. Denk bijvoorbeeld aan een explosie van principes of requirements die onvoldoende uitgekristalliseerd zijn.

Gerelateerde domeinen

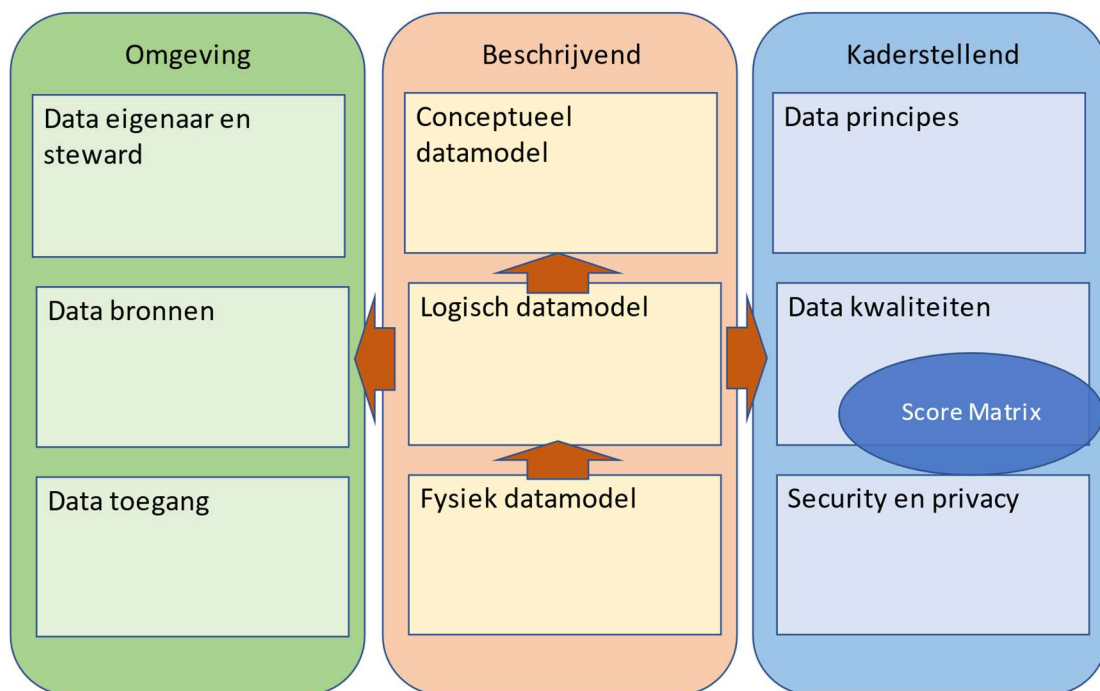
Deze principe datamodellering heeft met veel notatiewijzen een relatie. In onderstaande opsomming een overzicht:

- **Conceptueel datamodel**, hiermee worden de verbanden gelegd vanuit de principes naar de concepten of begrippen
- **Datagebruik modellen**, Link vanuit de principes naar de modellen rond het gebruik en de opslag van de data entiteiten. Bijvoorbeeld principes relevant in bedrijfsprocessen rond het gebruik van data of principes gericht op het applicatielandschap van de data bronnen.

DATA KWALITEITEN

Modelleren van data kwaliteiten kan met een score matrix. Dit is een bijzondere vorm van data modelleren omdat het zowel voor de vraag- als aanbodkant van datasets gebruikt kan worden. Het doel is om kwaliteiten (gemodelleerd als requirements) in relatie te brengen met data entiteiten en vervolgens een numerieke- of ordinale waarden worden toe te kennen.

Score matrices zijn voor verschillende doeleinden te gebruiken, waarbij opvallend is dat dit zowel in de ontwikkelfase als in de beheerfase hulp biedt, hierbij komt wederom zowel de vraag- als aanbodzijde aan bod.



Modelleerwijze

Primaire modelleerwijze is een matrix waarbij over de ene as de data kwaliteiten en over de andere as de conceptuele data entiteiten worden afgebeeld. In de cel wordt vervolgens een waarde of score gegeven van de huidige of de gewenste situatie.

Secundair kan ook een model worden opgesteld van kwaliteit verhogende maatregelen. Dit kan met een gestereotypeerde graaf op basis van ArchiMate.

Modelleertaal

Voor de score matrix is het mogelijk om een willekeurige modelleertaal te kiezen. Omdat andere requirements en principes worden gemodelleerd in ArchiMate doen we dat voor de score matrix ook. De matrix is daarmee een weergave van twee ArchiMate concepten over beide assen.

Entiteiten

Score bestaat uit het mappen van twee concepten en het geven van een score op het kruispunt van deze concepten. In de cellen waar de twee concepten elkaar snijden wordt vervolgens een score gegeven. Dat kan een getal zijn bijvoorbeeld tussen 0 en 10 waarmee aangegeven wordt in welke mate een entiteit aan een eis of kwaliteit voldoet. Andere opties zijn om hier ordinale waarden te gebruiken zoals Hoog - Midden - Laag (of Brons - Zilver - Goud).

Belangrijk is hierbij wel dat deze score vanuit twee perspectieven gegeven kan worden. Allereerst vanuit de aanbod- en vraagzijde. Ten tweede als huidige- en gewenste situatie of baseline en target.

Naast de scores concepten wordt er in een matrix een koppeling gelegd met een tweede dimensie en dit zijn verschillende soorten kwaliteiten, eisen of requirements bijvoorbeeld:

- **Data kwaliteiten**, kwantificeerbare requirements bijvoorbeeld vanuit het DMBOK raamwerk. Gemodelleerd als (ArchiMate) requirements of constraints.
- **Conceptueel Data Model object** in conceptuele modellen.

Secundaire modellering is mogelijk door modellen op te stellen van de verschillende kwaliteit verhogende maatregelen. Hierbij kunnen vanuit de ArchiMate modelleertaal de volgende concepten gekozen worden

- **Deliverable**, een oplevering vanuit een project dat er voor zorgdraagt dat een kwaliteit verhogende maatregel geïntroduceerd wordt
- **Core model entiteit**, elementen vanuit een architectuur laag (bedrijf-, applicatie-, infrastructuurlaag) dat de implementatie van de deliverable is.

Desgewenst is er nog een extra dimensie toe te voegen en dat is de dimensie tijd, de score acties kunnen namelijk per data entiteit en per gebruikersentiteit in de tijd veranderen, met name relevant bij een baseline en target aanpak.

Associaties

De volgende associaties kunnen toegepast voor de koppeling tussen de conceptuele data entiteiten en de kwaliteiten

- **Realisation**, Conceptuele data entiteit realiseert de data kwaliteit (positief)
- **Influence**, Conceptuele data entiteit beïnvloedt de data kwaliteit, meestal in negatieve zin.
- **Association**, verder niet gestereotypeerde associatie tussen kwaliteit en data entiteit

Notatie

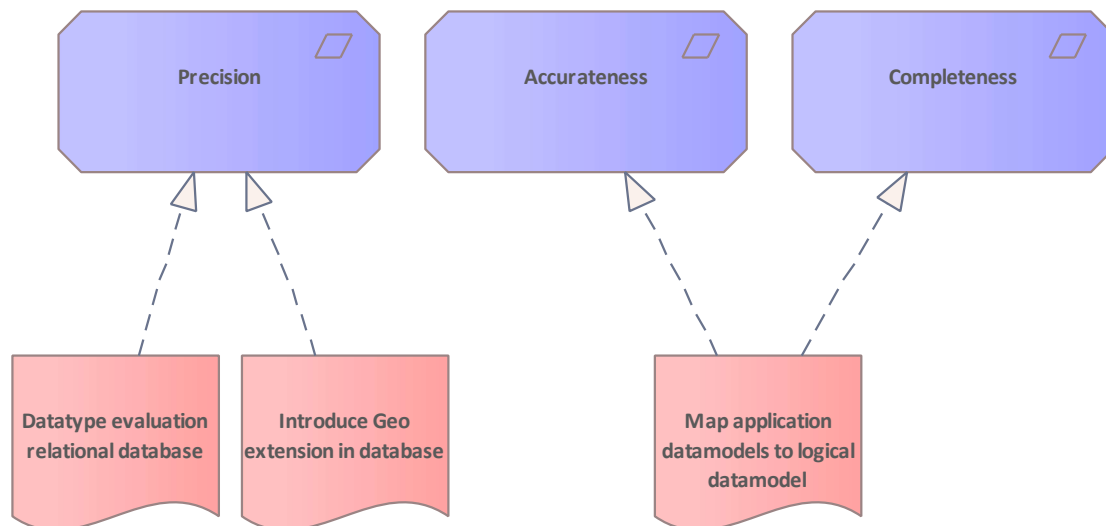
Notatie is relatief simpel maar daardoor eveneens krachtig. Notatie bestaat uit een matrix waarbij in de cellen een score getal of ordinale waarde komt te staan. Op de twee assen staat op de ene as de data entiteit zoals beschreven in de paragraaf concepten. Op de andere as wordt de kwaliteit getoond. Er wordt geen onderscheid gemaakt over welke as welk entiteitstype wordt weergegeven.

Afbeelding in de volgende paragraaf geeft een weergave van een score matrix. In de matrix zie je een voorbeeld op de business laag waarbij wordt weergegeven per bedrijfsobject hoe deze scoort op de datakwaliteiten zoals benoemd in DMBok. Let op dat een cel ook leeg kan zijn of een NaN waarde kan hebben

AFBEELDING

Target +											
+ Source	Accurateness	Actuality	Completeness	Consistency	Precision	Privacy	Reasonableness	Referential integrity	Timeliness	Uniqueness	Validity
Classroom training	3					1		3			1
Employee	3					3		3			1
Online training	3					1		3			1
Person	3					3		3			1
Registration	3	3	3	3	3	1	3	3	3	3	3
Trainee	3	3	3	3	3	3	3	3	3	3	3
Trainer	3	3				3		3			1
Training	3					1		3			1
Training Session	3	3				1		3			3

Voorbeeld van de primaire modellering met een score matrix. Score is in dit geval een numerieke waarde maar kan ook hoog-midden-laag zijn.



Secundair model met daarin kwaliteit verhogende maatregelen die een data kwaliteit realiseren.

Kenmerken

De score matrix is al lang aanwezig in andere werkvelden maar in data managementland wordt deze notatie relatief weinig toegepast. Haar kracht is bewezen in meerdere verschijnings- en toepassingsvormen. Ik heb goede ervaring opgedaan in inventarisaties maar ook in interactieve workshops. De belangrijkste (positieve en negatieve) kenmerken:

- Krachtige notatiewijze waarin drie concepten in één weergave met elkaar worden gecombineerd, namelijk scores (cel) data entiteit (As 1) en eis/requirement of kwaliteit (As 2).
- Eenvoudig toepasbaar bij gebruik voor stakeholders zonder modelleerervaring.
- Kan gebruikt worden in interactieve workshops.
- Toepasbaar op meerdere abstractie niveaus, van logisch tot conceptueel.
- Geschikt voor notatie binnen data management en data kwaliteit context.
- Goed toepasbaar binnen data security en - privacy context.
- Modellen kunnen bij veel entiteiten en kwaliteiten lastiger leesbaar worden.
- Tijdsdimensie kan ingezet worden voor het modelleren van een baseline en target inventarisatie.

Toepassingen

Score matrix wordt in meerdere situaties toegepast maar de meest voorkomende zijn

- Data kwaliteit matrix (hoe scoort een data entiteit op een kwaliteit)
- Data Security en privacy, in welke mate wordt aan eisen vanuit deze werkvelden voldaan.

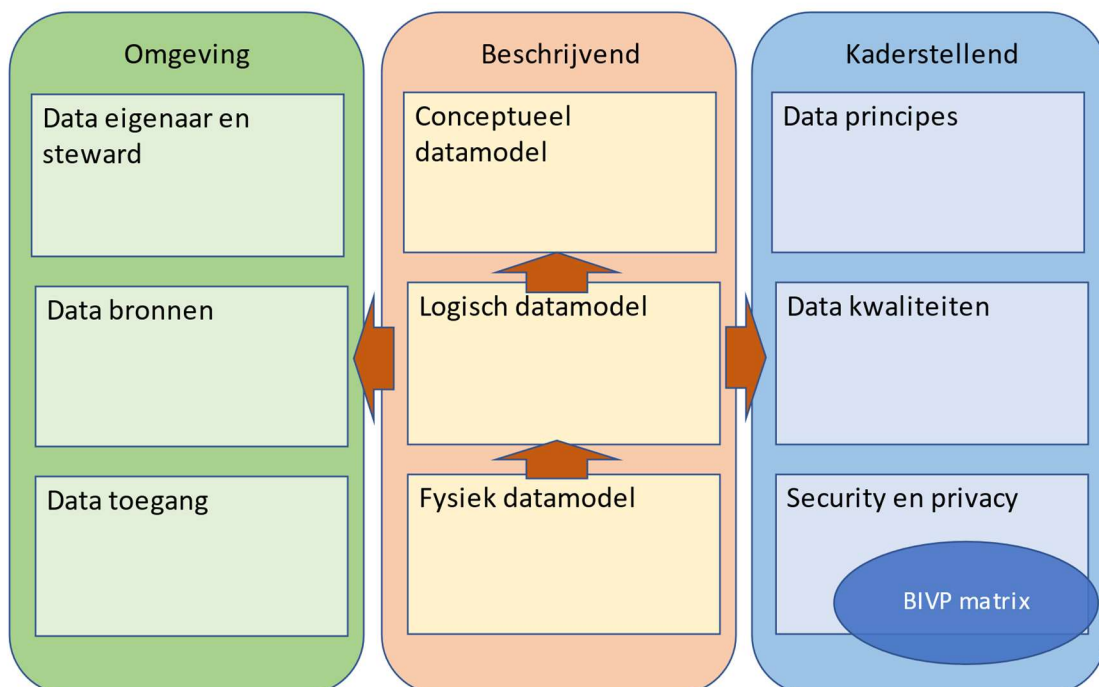
Gerelateerde domeinen

- **Conceptueel data model** waarin de data entiteiten gemodelleerd zijn waarop de kwaliteiten een score krijgen.
- **Security en privacy model**, zijn feitelijk een verbijzondering van de hier genoemde datakwaliteiten en kunnen daarom met dezelfde concepten gemodelleerd worden.
- **Maatregelen model**, secundair model voor kwaliteit modellering en bijbehorende maatregelen.

DATA SECURITY EN -PRIVACY

Modelleren van data security en privacy is gebaseerd op een hiërarchie van maatregelen. Wat het niveau is van deze maatregelen kan gemodelleerd worden met een score matrix. Dit is een bijzondere vorm van data modelleren omdat het zowel voor de vraag- als aanbodkant van datasets gebruikt kan worden. Het doel is om de beveiligingsmaatregelen (gemodelleerd als requirements) in relatie te brengen met data entiteiten en vervolgens een numerieke- of ordinale waarden worden toegekend.

Voor security wordt vaak een BIV classificatie gebruikt BIV staat voor beschikbaarheid, Integriteit en Vertrouwelijkheid. Dit kan uitgebreid worden met een privacy classificatie waarmee een BIVP classificatie ontstaat. Dit zijn feitelijk bijzondere data kwaliteiten en kunnen op soortgelijke wijze gemodelleerd worden.



Modelleerwijze

Primaire modelleerwijze is een matrix waarbij over de ene as de BIVP categorieën en over de andere as de conceptuele data entiteiten worden afgebeeld. In de cel wordt vervolgens een waarde of score gegeven van de huidige of de gewenste situatie.

Secundair kan ook een model worden opgesteld van kwaliteit verhogende maatregelen. Dit kan met een gestereotypeerde graaf op basis van ArchiMate.

Modelleertaal

Voor de BIVP matrix is het mogelijk om een willekeurige modelleertaal te kiezen. Omdat andere requirements en principes worden gemodelleerd in ArchiMate doen we dat voor de BIVP matrix ook. De matrix is daarmee een weergave van twee ArchiMate concepten over beide assen.

Entiteiten

Score bestaat uit het mappen van twee concepten en het geven van een score op het kruispunt van deze concepten. In de cellen waar de twee concepten elkaar snijden wordt vervolgens een score gegeven. Dat kan een getal zijn bijvoorbeeld tussen 1,2 en 3 waarmee aangegeven wordt in welke mate een entiteit aan een beveiligingsniveau dient te voldoen. Andere opties zijn om hier ordinale waarden te gebruiken zoals Hoog - Midden - Laag (of Brons - Zilver - Goud).

Belangrijk is hierbij wel dat deze score vanuit twee perspectieven gegeven kan worden. Allereerst vanuit de aanbod- en vraagzijde. Ten tweede als huidige- en gewenste situatie of baseline en target.

Naast de scores concepten wordt er in een matrix een koppeling gelegd met een tweede dimensie en dit zijn verschillende soorten beveiligings- en privacy eisen:

- **BIVP eisen**, kwantificeerbare requirements gemodelleerd als (ArchiMate) requirements of constraints
- **Conceptueel Data Model object** in conceptuele modellen

Secundaire modellering is mogelijk door modellen op te stellen van de verschillende kwaliteit verhogende maatregelen. Hierbij kunnen vanuit de ArchiMate modelleertaal de volgende concepten gekozen worden

- **Deliverable**, een oplevering vanuit een project dat er voor zorgdraagt dat een kwaliteit verhogende maatregel geïntroduceerd wordt
- **Core model entiteit**, elementen vanuit een architectuur laag (bedrijf-, applicatie-, infrastructuurlaag) dat de implementatie van de deliverable is.

Desgewenst is er nog een extra dimensie toe te voegen en dat is de dimensie tijd, de score acties kunnen namelijk per data entiteit en per gebruikersentiteit in de tijd veranderen, met name relevant bij een baseline en target aanpak.

Associaties

De volgende associaties kunnen toegepast voor de koppeling tussen de conceptuele data entiteiten en de kwaliteiten

- **Realisation**, Conceptuele data entiteit realiseert de data kwaliteit.
- **Influence**, Conceptuele data entiteit beïnvloedt de data kwaliteit, meestal in negatieve zin.
- **Association**, verder niet gestereotypeerde associatie tussen kwaliteit en data entiteit.

Notatie

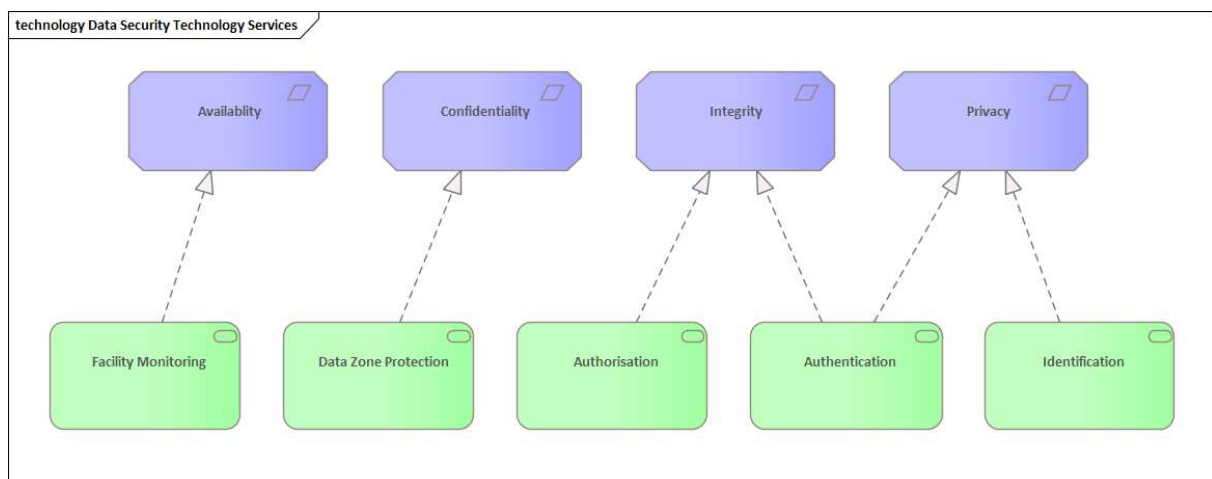
Notatie is relatief simpel maar daardoor eveneens krachtig. Notatie bestaat uit een matrix waarbij in de cellen een score getal of ordinale waarde komt te staan. Op de twee assen staat op de ene as de data entiteit zoals beschreven in de paragraaf concepten. Op de andere as wordt de kwaliteit getoond. Er wordt geen onderscheid gemaakt over welke as welk entiteitstype wordt weergegeven.

Afbeelding in de volgende paragraaf geeft een weergave van een score matrix. In de matrix zie je een voorbeeld op de business laag waarbij wordt weergegeven per bedrijfsobject hoe deze scoort BIVP categorieën. Let op dat een cel ook leeg kan zijn of een NaN waarde kan hebben.

AFBEELDING

Target +	Objects::Classroom training	Objects::Employee	Objects::Online training	Objects::Person	Objects::Registration	Objects::Trainee	Objects::Trainer	Objects::Training	Objects::Training Session
+ Source									
BIVP::Availability	1	2	1	1	2	1	1	1	1
BIVP::Confidentiality	1	3	1	3	1	3	3	1	1
BIVP::Integrity	2	3	2	3	3	3	3	3	2

Voorbeeld van de primaire modellering met een score matrix. Score is in dit geval een numerieke ordinale waarde maar kan ook hoog-midden-laag zijn. Houdt er rekening mee dat deze scorematrix zowel voor de huidige als de gewenste situatie opgesteld kan worden.



Secundair model met daarin BIVP verhogende maatregelen die een niveau van beveiliging realiseren.

Kenmerken

De score matrix is al lang aanwezig in andere werkvelden maar in data managementland wordt deze notatie relatief weinig toegepast. Haar kracht is bewezen in meerdere verschijnings- en toepassingsvormen. Ik heb goede ervaring opgedaan in inventarisaties maar ook in interactieve workshops. De belangrijkste (positieve en negatieve) kenmerken:

- Krachtige notatiewijze waarin drie concepten in één weergave met elkaar worden gecombineerd, namelijk scores (cel) data entiteit (As 1) en eis/requirement of kwaliteit (As 2).
- Eenvoudig toepasbaar bij gebruik voor stakeholders zonder modelleerervaring.
- Kan gebruikt worden in interactieve workshops.
- Toepasbaar op meerdere abstractie niveaus, van logisch tot conceptueel.
- Geschikt voor notatie binnen data management en data kwaliteit context.
- Goed toepasbaar binnen data security en - privacy context.
- Modellen kunnen bij veel entiteiten en kwaliteiten lastiger leesbaar worden.
- Tijdsdimensie kan ingezet worden voor het modelleren van een baseline en target inventarisatie.

Toepassingen

Score matrix wordt in meerdere situaties toegepast maar de meest voorkomende zijn

- Data kwaliteit matrix (hoe scoort een data entiteit op een kwaliteit)
- Data Security en privacy, in welke mate wordt aan eisen vanuit deze werkvelden voldaan.

Gerelateerde domeinen

- **Conceptueel data model** waarin de data entiteiten gemodelleerd zijn waarop de BIV classificatie worden gescoord.
- **Kwaliteiten model**, zijn feitelijk een generalisatie van de hier genoemde BIVP requirements en kunnen daarom met dezelfde concepten gemodelleerd worden.
- **Maatregelen model**, secundair model voor kwaliteit modellering en bijbehorende maatregelen.

Toepassingsmodellen

Naast de datamodellen en de kaders rond data management wordt data ook op allerlei plaatsen geproduceerd en gebruikt binnen en buiten de organisatie. Ook hierover willen we graag metadata verzamelen. In dit hoofdstuk gaan we daarom in op databronnen, datagebruik en eigenaren en stewards van data.

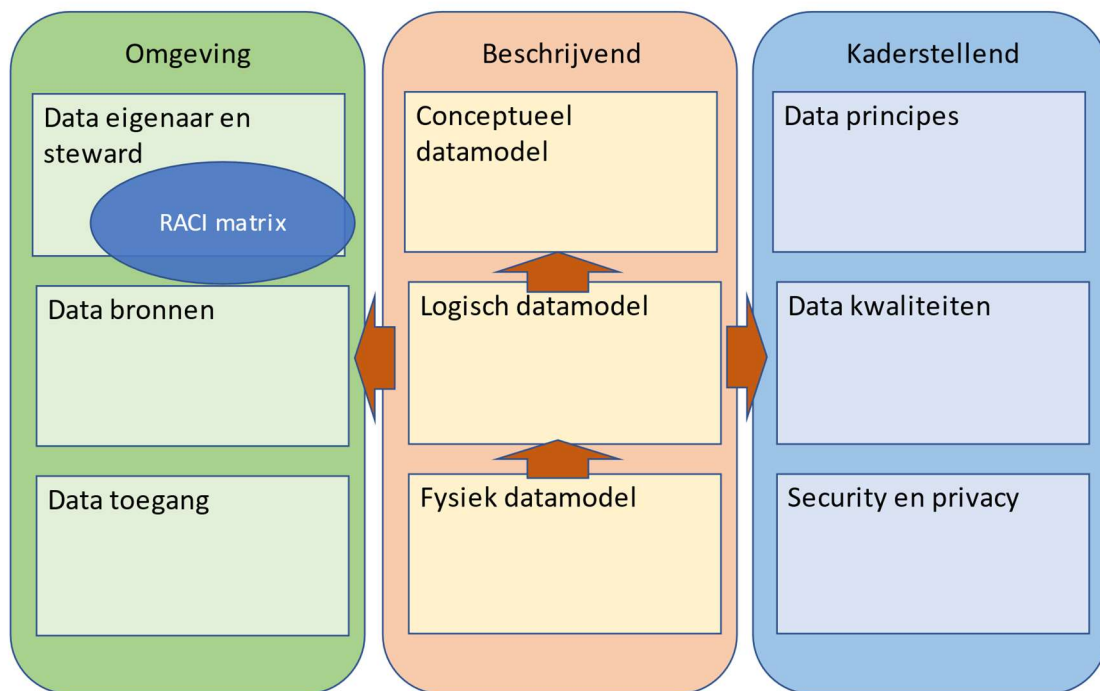
DATA EIGENAAR EN STEWARD

Data eigenaren en stewards spelen een rol bij het beleid rond de verschillende data entiteiten in de organisatie. Zij bepalen het beleid en nemen beslissingen omtrent de productie, het gebruik en de kwaliteiten van de data.

Een matrix waarmee deze betrokkenheid inzichtelijk wordt gemaakt is daarmee een eenvoudig hulpmiddel om inzichtelijk te maken wie welke rol speelt bij een data entiteit. Ook kan de matrix gebruikt worden wie op welke wijze geïnformeerd dient te worden bij wijzigingen cq aanpassingen in de structuur en toepassing van een data entiteit binnen een organisatie.

Hiervoor zijn verschillende soorten matrices te gebruiken. Bijvoorbeeld de Eigenaar matrix maar ook de RACI matrix. RACI staat voor Responsible, Accountable, Consulted en Informed en de matrix geeft aan wie op welke wijze betrokken is bij data entiteiten. RACI matrices zijn voor verschillende doeleinden te gebruiken, waarbij opvallend is dat dit zowel in de ontwikkelfase als in de beheerfase hulp biedt. Als laatste is te noemen dat de RACI matrix op meerdere abstractieniveaus toegepast kan worden, maar dat de data governance view op conceptueel niveau de meest gebruikelijke is.

De RACI matrix is een belangrijk hulpmiddel bij de werkprocessen rond data governance en in mindere mate bij datagebruik. De RACI matrix kan op een gedetailleerd niveau (desgewenst tot op attribuutniveau binnen objecten) aangeven welke stakeholder op welke wijze betrokken is bij de governance van de objecten. Zo zal de data eigenaar vrijwel altijd responsible zijn en in veel gevallen ook accountable. De data stewards zullen veelal consulted en informed worden bij het bepalen en vaststellen van het beleid rond data.



Modelleerwijze

De modelleerwijze voor de eigenaar of de RACI matrix is zoals de naam al aangeeft een matrix. Waarbij over de ene as de conceptuele data entiteiten en op de andere as een bedrijfsrol of bedrijfsactor gemodelleerd wordt.

Modelleertaal

Voor de RACI matrix is het mogelijk om een willekeurige modelleertaal te kiezen. Omdat andere onderdelen van ons model worden gemodelleerd in ArchiMate doen we dat voor de RACI matrix ook. De matrix is daarmee een weergave van twee ArchiMate concepten over beide assen.

Entiteiten

Naast de RACI concepten wordt er in een matrix een koppeling gelegd met een tweede dimensie en dit zijn verschillende soorten gebruiksentiteiten bijvoorbeeld:

- **Data eigenaar en - steward** van de data entiteit (RA en CI)
- **Gebruiker**, degene die een CI relatie heeft met een data entiteit
- **Applicatie of proces eigenaar** van een informatiesysteem of bedrijfsproces dat een RACI bewerking uitvoert op een (data) object
- **Management vertegenwoordiging** met name voor de Accountable aspecten

Naaste de tweede dimensie is er nog een derde die een verband legt tussen de dynamische of gebruikersentiteiten naar de data entiteiten zoals:

- **Conceptuele data objecten als business object** in conceptuele modellen
- **Applicatieve Data object**, in een detailuitwerking binnen het metamodel op basis van de ArchiMate lagen. Deze is optioneel en relevant voor alleen grotere metamodelen in complexe organisaties.

Associaties

Leggen van verbindingen tussen de bovengenoemde entiteiten. Hierbij zijn er meerdere ArchiMate relaties mogelijk:

- **Association**, een niet gestereotypeerde associatie, in de naam van de associatie wordt aangegeven wat de soort relatie is (R, A, C, I of een combinatie daarvan).
- **Serving** (vanuit de rol of actor). Dit is een interessante stereotypering die zelden gebruikt wordt maar in deze extra informatie geeft. Je geeft hierbij aan dat de eigenaar of steward een dienst levert omtrent de data entiteit.

Access associatie lijkt in deze ook interessant maar geeft feitelijk een andere relatie aan, het gebruik van de data en is daarmee voor een relatie met de eigenaar of de steward minder relevant.

Notatie

RACI bestaat zoals de afkorting al laat zien uit vier basis concepten. Deze basisconcepten zullen worden weergegeven in de cellen van de matrix en is daarmee de eerste dimensie. Of dit vanuit datagebruik- of dataproductie perspectief gemodelleerd wordt is niet relevant

- **Responsible**, wie is verantwoordelijk voor de data entiteit. Met andere woorden, wie neemt beslissingen over het doen van aanpassingen aan objecten of records. Dit heeft zowel betrekking op de structuur van een data entiteit als op de inhoud.
- **Accountable**, wie is verantwoordelijk voor de financiële aspecten voor een data entiteit. Met andere woorden wie betaalt voor het beheer van de data entiteit en financiert de activiteiten waarmee data entiteiten verbeterd worden (bijvoorbeeld structuur of data kwaliteit verhogende activiteiten).
- **Consulted**, wie dient geraadpleegd te worden bij activiteiten rond de data entiteiten, bijvoorbeeld wijzigingen in structuur en inhoud. Veelal zijn dit afnemers van datasets die betrokken zijn in een (integratie)keten.
- **Informed**, wie dient geïnformeerd te worden bij activiteiten rond de data entiteiten, bijvoorbeeld wijzigingen in structuur en inhoud. Veelal zijn dit afnemers van datasets.

Eventueel is ook een matrix te definiëren voor de specifieke rol vanuit data management perspectief. Denk hierbij bijvoorbeeld aan:

- Data owner (O)
- Data steward (S)
- Security/privacy officer (B)

- Architect etc (A)

Je kunt hierbij eenvoudig een lijst samenstellen van rollen die in de eigen organisatie aanwezig zijn rond de conceptuele data entiteiten.

AFBEELDING

Target +	Classroom training	Employee	Online training	Person	Registration	Trainee	Trainer	Training	Training Session
+ Source									
Backoffice	CI	CI	CI	CI	CI	CI	CI	CI	CI
Management	RA	RA	RA	RA	RA	RA	RA	RA	RA
Trainee						CI		CI	
Trainer	CI	CI	CI				RCI	CI	CI

De afbeelding geeft de RACI matrix. Let op dat een rol meerdere elementen uit de RACI categorieën kan omvatten dus responsible en accountable.

Een mooie controle is of er voor elke conceptuele data entiteit een complete set van RACI elementen voorkomt binnen de bedrijfsrollen.

Kenmerken

De RACI of datarol matrix is al lang aanwezig in data managementland en heeft haar kracht bewezen in meerdere verschijningsvormen. De belangrijkste (positieve en negatieve) kenmerken:

- Krachtige notatiewijze waarin drie concepten in één weergave met elkaar worden gecombineerd, namelijk RACI bewerkingen (cel) data entiteit (As 1) en gedragsentiteit (As 2).
- Eenvoudig toepasbaar bij gebruik voor stakeholders zonder modelleerervaring.
- Kan gebruikt worden in interactieve workshops.
- Toepasbaar op meerdere abstractie niveaus, van fysiek tot conceptueel.
- Geschikt voor notatie binnen data management en met name de data governance context.
- Goed toepasbaar binnen data security en - privacy context.
- Modellen kunnen bij veel entiteiten en attributen en gedragsentiteiten lastiger leesbaar worden.

Toepassingen

RACI matrix wordt in meerdere situaties toegepast maar de meest voorkomende zijn

- Data Governance matrix (wie is op welke wijze betrokken bij een data entiteit).
- Data Management beschrijving van de betrokkenheid die door een gedragsentiteit wordt uitgevoerd op een gegevens entiteit (dus een algemener perspectief dan de governance).

Gerelateerde domeinen

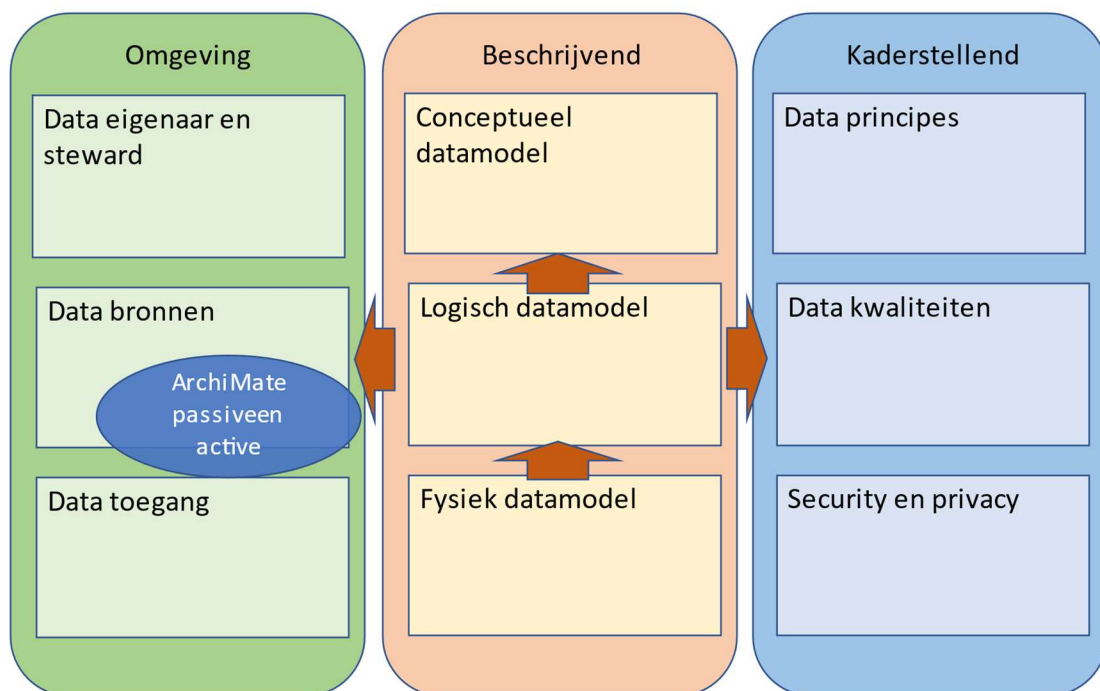
- De RACI matrix heeft een directe relatie met het conceptuele data model.
- De RACI matrix kan desgewenst gecombineerd worden met het logisch en fysiek datamodel om hiermee de details van de entiteiten te modelleren.
- Een vergelijkbare notatiewijze uit data gebruik is de CRUD matrix die veelal op meerdere modelleerniveaus wordt toegepast.

DATA BRONNEN

ArchiMate is een architectuur-modelleertaal waarmee enerzijds verschillende lagen van een organisatie gemodelleerd kunnen worden, namelijk, bedrijf-, applicatie- en technische laag. Daarnaast zijn er een aantal extensies zoals motivation en implementation. Naast de lagen zijn er een drietal concept categorieën, te weten gedrag, actieve structuur en passieve structuur. Binnen al deze aspectgebieden spelen data entiteiten een rol en worden er associaties gelegd van en naar deze entiteiten.

Datagebruik modellering in ArchiMate heeft tot doel om de indeling van bedrijfsentiteiten, data entiteiten en artefacten te beschrijven. Hierbij wordt een duidelijk onderscheid gemaakt tussen de entiteiten in de lagen en worden er meestal per laag modellen opgesteld. De bronmodellering bevindt zich in alle drie de concept categorieën waarbij de passieve structuur concept categorie het startpunt van onze modellen is. Vervolgens wordt er een model opgesteld welk deel van de actieve structuur of het gedrag de bron is van deze data entiteiten

Het doel van deze modelleerwijze is om de verbanden tussen conceptuele of applicatieve data entiteiten en de verschillende concepten in de applicatie laag te beschrijven. Deze notatie is daarmee zeer geschikt om grondplaten of blauwdrukken op te stellen van een data- en applicatielandschap. Met de andere ArchiMate viewpoints wordt met taal gestandaardiseerde viewpoints verbanden gelegd naar andere aspectgebieden zoals requirements of de bedrijfsarchitectuur.



Modelleerwijze

Voor het modelleren van conceptuele data objecten en de bronnen van deze objecten kunnen gestereotypeerde ArchiMate grafen gebruikt worden. Zeker in de situatie waarin niet alleen de actieve concepten maar ook het logische applicatie landschap gemodelleerd.

Modelleertaal

Voor de databronnen is er binnen ArchiMate een aantal elementen, associaties en afgeleide associaties beschikbaar. Hierin zitten onder andere de voor ons relevante elementen zoals applicatie component of applicatie functie.

Binnen de ArchiMate modellering wordt gebruik gemaakt van grafen bestaande uit entiteiten en associaties. De relevante concepten worden afzonderlijk beschreven:

Entiteiten

- **Conceptuele Data objecten**, zijn als business object het centrale punt in ons model en dienen op enigerlei wijze verbonden te zijn met het applicatielandschap. Dat kan direct met een afgeleide associatie of via applicatieve data objecten die een business object realiseren.
- **Applicatiefuncties (inclusief events en processen)**, beschrijven het applicatielandschap op een logische wijze, dus zonder de fysieke applicaties waarin de data objecten beheerd worden.
- **Applicatiecomponenten**, beschrijven het applicatielandschap op een fysieke wijze, dus de fysieke applicaties waarin de data objecten beheerd worden. Via de applicatiefuncties of -services wordt de koppeling gelegd met deze data objecten.
- **Applicatieve data objecten**, beschrijven de data entiteiten waaruit een business object is opgebouwd. Een business object kan namelijk zijn opgebouwd uit meerdere data objecten die vervolgens door verschillende logische- en fysieke applicaties worden benaderd.
- **Applicatieservices en interfaces**, zijn een bijzonder concept en geven aan hoe er logische- en fysieke verbindingen worden gelegd tussen applicaties onderling en tussen applicaties en de entiteiten in de bedrijfsarchitectuur.

Associaties leggen verbindingen tussen de bovengenoemde entiteiten:

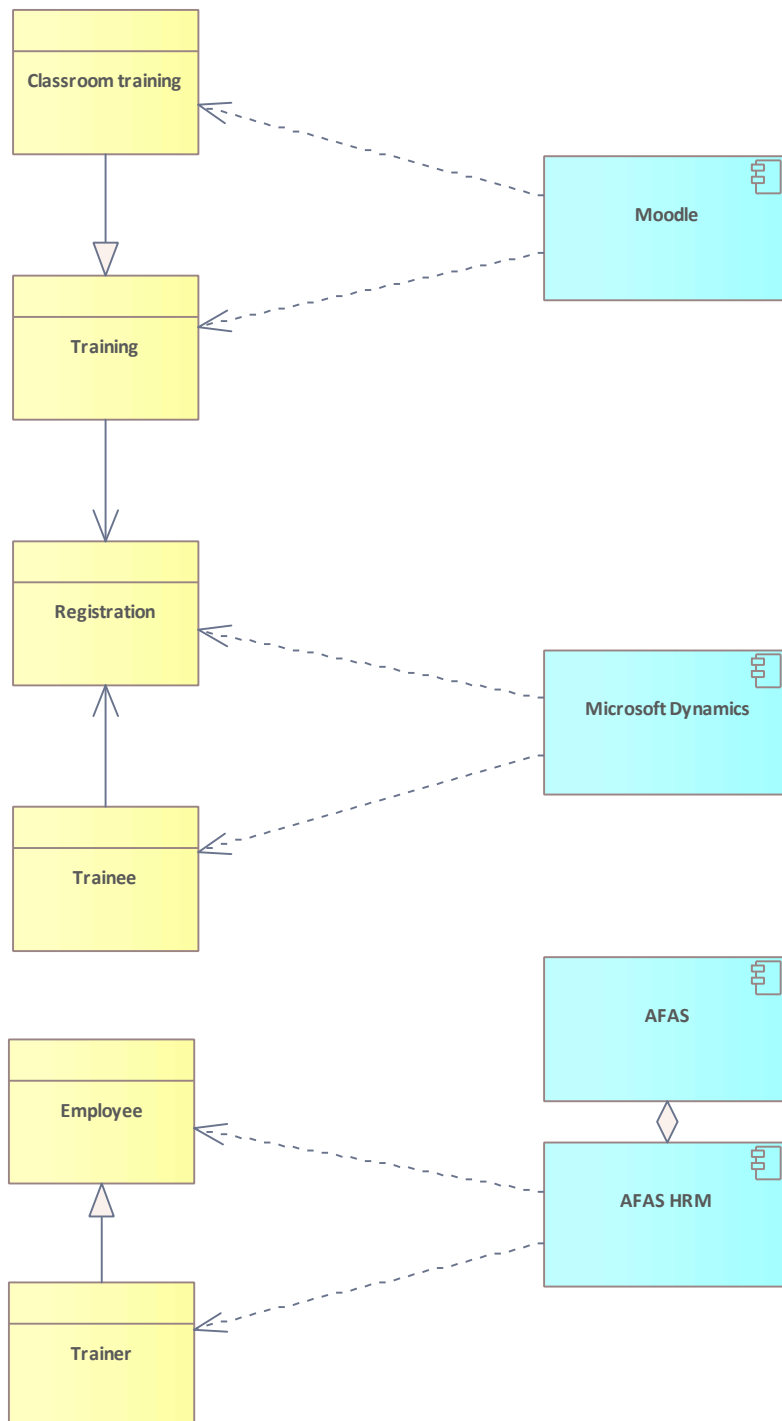
- **Access** waarmee een applicatiefunctie toegang heeft tot een data object. Dit kan in bi-directionele richting gemodelleerd worden, dus lezen en/of schrijven. Dit is de voorkeursrelatie.
- **Assignment**, hiermee worden de concepten uit de actieve structuur categorie toegewezen aan de dynamische categorie concepten.
- **Associatie**, tussen al deze entiteiten kan de meest generieke verbinding gelegd worden. Deze kan gebruikt worden om een associatie op basis van een rol te leggen.

Notatie

In de notatie van de ArchiMate data- en applicatiemodellering wordt een onderscheid gemaakt tussen modellen met één soort entiteiten en met gecombineerde entiteiten. Deze modellen zijn echter vrijwel altijd gebaseerd op samengestelde entiteiten. Dit omdat je hier feitelijk de koppeling legt tussen data objecten en het applicatielandschap. Alle modellen zijn gebaseerd op grafen.

Extra aandachtspunt is dat de conceptuele data entiteiten in de bedrijfslaag van ArchiMate en de applicatie componenten in de applicatielaag. Daarnaast is het van belang te beseffen dat het gedrag (logisch applicatie model) niet altijd gemodelleerd wordt. Hiermee ontstaat er een vereenvoudiging van het ArchiMate model. Echter door gebruik van afgeleide associaties is dit een prima toe te passen werkwijze.

AFBEELDING



In de afbeelding is te zien hoe een aantal applicatie componenten de bron zijn voor een aantal conceptuele data entiteiten. Dit is een voorbeeld van het combineren van twee architectuur lagen en het niet uitwerken van de gedrag kenmerken.

Dit is daarmee het eenvoudigste data bron model dat gebruikt kan worden. Is binnen de architectuur van de organisatie reeds een logisch applicatie model, bijvoorbeeld met architectuur functies, uitgewerkt, dan is dit een mooie detaillering voor ons databronnen model.

Kenmerken

De databron modellen op basis van ArchiMate is een abstracte en krachtige notatiewijze met de volgende kenmerken:

- Eenvoudige notatiewijze waarin begrippen met een beperkt aantal associatietypen aan elkaar gerelateerd kunnen worden
- Viewpoint met een focus op het applicatielandschap in relatie tot conceptuele data objecten
- Eenvoudig toepasbaar bij stakeholders zonder modelleerervaring
- Kan goed gebruikt worden in interactieve workshops
- Toepasbaar op hoge abstractie niveaus, voornamelijk conceptueel
- Mogelijkheid om verbanden binnen meerdere enterprise architectuur lagen te leggen voor de conceptcategorieën (kolommen)
- Goed model om discussie op gang te brengen tussen applicatiebeheerders, data stewards, owners en architecten
- Belangrijk hulpmiddel bij het opstellen van datamodellen binnen complexe applicatielandschappen en bij ketenintegraties.
- Modellen kunnen complex worden als er een divers applicatielandschap is.

Toepassingen

ArchiMate data- en applicatiemodellering worden vooral toegepast om het applicatielandschap vanuit data perspectief te modelleren. Het biedt een goed startpunt voor het in kaart brengen van de bedrijfsapplicaties en dataentiteiten. In complexe applicatiedomeinen is data veelal een verbindend aspect en kan een analyse en een model van de gegevensstromen helderheid bieden.

Houdt er rekening mee dat ondanks de eenvoud van de notatiewijze het opstellen van een ArchiMate databronnenmodel een complex traject kan zijn, zeker bij een complex domein of binnen een organisatiecontext waar rond applicatiebeheer weinig volwassenheid is. Denk bijvoorbeeld aan het ontbreken van documentatie van koppelingen en integraties.

Gerelateerde domeinen

Deze ArchiMate databron modellering heeft met veel notatiewijzen een relatie. In onderstaande opsomming een overzicht:

- **Conceptuele Datamodellering**, primaire viewpoint waarmee de verbanden gelegd worden vanuit het applicatie landschap naar het data landschap
- **ArchiMate Data- & Bedrijfslaag**, hiermee worden de verbanden gelegd vanuit de data objecten via het data gebruik in concepten als bedrijfsproces of -functie etc
- **ArchiMate Data Kaders**, verbanden leggen vanuit de databronnen en conceptuele data entiteiten naar kwaliteiten, requirements, principes, stakeholders en constraints

DATA TOEGANG

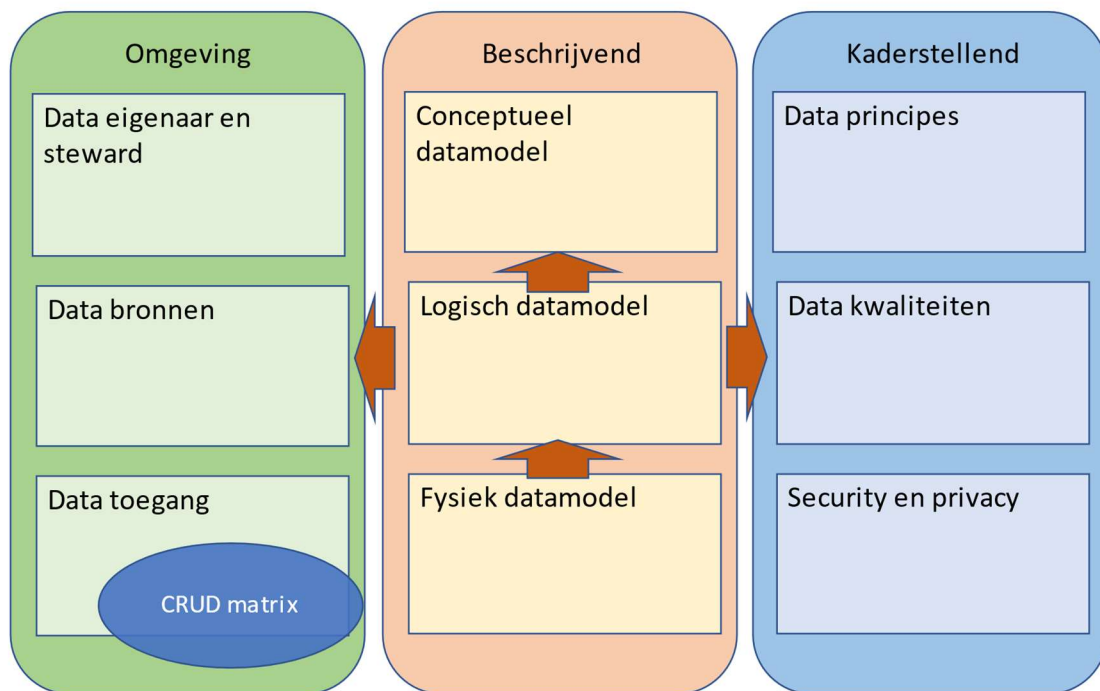
Wie heeft toegang tot de data binnen de organisatie is vanuit veel data management dimensies van groot belang. Het gaat daarbij om het gebruik en de productie van de data. Hiervoor wordt veelal een CRUD matrix gebruikt.

CRUD matrix is een belangrijk hulpmiddel bij de werkprocessen rond informatiebeveiliging en -privacy. De CRUD matrix kan op een gedetailleerd niveau (desgewenst tot op attribuutniveau binnen objecten) aangeven welke stakeholder toegang heeft tot de data en wie mutatierechten heeft op de data entiteiten.

CRUD staat voor Create, Read, Update en Delete en de matrix geeft aan wie deze bewerkingen op een data entiteit mag uitvoeren. Het is daarmee een eenvoudig hulpmiddel om inzichtelijk te maken wat enerzijds de autorisaties zijn van bepaalde entiteiten zoals rollen, actoren maar ook applicaties, bedrijfsfuncties en -processen. Anderzijds kan de matrix gebruikt worden welke bewerking door een bepaalde entiteit wordt uitgevoerd, waarbij het niet de autorisatie aspecten belicht maar meer ingaat op de dynamische kenmerken van gedragsentiteiten op de conceptuele gegevensentiteiten.

Ook op het vlak van data modellering is de CRUD matrix goed inzetbaar. Dat kan op alle drie de niveaus. Voor fysieke modellering zal het gericht zijn op het modelleren van de gebruikersrollen, users en autorisaties, op een hoger niveau zoals logische en conceptuele datamodellering wordt hiermee gedocumenteerd welke stakeholder, applicatie, bedrijfsproces of -functie CRUD rechten heeft op data- en of business objecten.

Binnen data management tenslotte zijn CRUD matrices eveneens inzetbaar om in kaart te brengen welke stakeholders tijdens de beheerfase van een datalandschap CRUD rechten hebben op de data entiteiten.



Modellerwijze

De modellerwijze voor de CRUD matrix is zoals de naam al aangeeft een matrix. Waarbij over de ene as de conceptuele data entiteiten en op de andere as een bedrijfsrol of bedrijfsactor gemodelleerd wordt. Eventueel is op andere abstractieniveaus een andere indeling mogelijk. Bijvoorbeeld toegang tot de fysieke tabellen door gebruikers of gebruikersgroepen. Of het modelleren welke applicatie lezen en schrijven naar logische data entiteiten.

Modelleertaal

Voor de CRUD matrix is het mogelijk om een willekeurige modelleertaal te kiezen. Omdat andere onderdelen van ons model worden gemodelleerd in ArchiMate doen we dat voor de CRUD matrix ook. De matrix is daarmee een weergave van twee ArchiMate concepten over beide assen. Welke concepten dat zijn is vrij te kiezen maar Conceptuele data entiteit (Business Object) en bedrijfsrol (Business Role) is een minimum combinatie.

Entiteiten

CRUD bestaat zoals de afkorting al laat zien uit vier basis concepten. Deze basisconcepten zullen worden weergegeven in de cellen van de matrix en is daarmee de eerste dimensie. Of dit vanuit gebruiks- of autorisatieperspectief gemodelleerd wordt is niet relevant

- **Create**, aanmaken van objecten of records
- **Read**, lezen van de inhoud van een object, record of een dataset
- **Update**, mutatie van de inhoud van een data entiteit

- **Delete**, verwijderen of vernietigen van een data entiteit

Echter naast de CRUD concepten wordt er in een matrix een koppeling gelegd met een tweede dimensie en dit zijn zeer verschillende soorten gebruiks- en gedragsentiteiten bijvoorbeeld

- **Gebruiker**, degene die de CRUD acties zal of mag uitvoeren
- **Applicatie**, informatiesysteem dat een CRUD bewerking uitvoert op een (data) object
- **Bedrijfsproces** dat een CRUD bewerking uitvoert, bijvoorbeeld op een business object
- **Bedrijfsfunctie** dat een CRUD bewerking uitvoert, bijvoorbeeld op een business object
- **Bedrijfsrol** dat een CRUD bewerking uitvoert, bijvoorbeeld op een business object
- **Actor** die een CRUD bewerking uitvoert, bijvoorbeeld op een business object

Naaste de tweede dimensie is er nog een derde die een verband legt tussen de dynamische of gebruikersentiteiten naar de data entiteiten zoals:

- **Conceptueel data of business object** in conceptuele modellen
- **Logisch data object**, in logische modellen en bij de modellering van applicatie gedrag op gegevens entiteiten
- **Tabel**, met name bij fysieke modellering voor relationele databases relevant
- **Bericht of Interface**, met name bij fysieke modellering van interfaces of berichten relevant. Zoals in een SOA register

Desgewenst is er nog een vierde dimensie toe te voegen en dat is de dimensie tijd, de CRUD acties kunnen namelijk per data entiteit en per gebruikersentiteit in de tijd veranderen.

Associaties

Leggen van verbindingen tussen de bovengenoemde entiteiten uitgaande van de ArchiMate relaties:

- **Access**, is de meest gebruikelijke associatie voor de CRUD matrix. Hou er rekening mee dat je desgewenst ook de richting kunt aangeven bij een access associatie in ArchiMate
- **Associatie**, zonder verdere kwalificatie, desgewenst wel uit te breiden met de richting van de associatie, access heeft de voorkeur.

Notatie

Notatie is relatief simpel maar daardoor eveneens krachtig. Notatie bestaat uit een matrix waarbij in de cellen een combinatie van nul of meer letters CRUD komt te staan. Op de twee assen staat op de ene as de data entiteit zoals beschreven in de paragraaf concepten. Op de andere as wordt de gedrag- of dynamische entiteit weergegeven. Er wordt geen onderscheid over welke as welk entiteitstype wordt weergegeven

Afbeelding in de volgende paragraaf geeft een weergave van een CRUD matrix. In de matrix zie je een voorbeeld op de business laag waarbij wordt weergegeven welke bedrijfsrol welke business objecten binnen de CRUD acties kan uitvoeren. Let op dat een cel ook leeg kan zijn dan heeft deze bedrijfsrol geen enkele CRUD rechten op het business object.

AFBEELDING

Target +	Classroom training	Employee	Online training	Person	Registration	Trainee	Trainer	Training	Training Session
+ Source	Backoffice	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD
	Management		R	R	R				
	Trainee	R		R		CRU	R	R	
	Trainer	R	R	R	R	R	R	CRU	R

De afbeelding geeft de CRUD matrix. Let op dat een rol meerdere elementen uit de CRUD categorieën kan omvatten dus zowel lezen als schrijven gecombineerd.

Een mooie controle is of er voor elke conceptuele data entiteit een complete set van CRUD elementen voorkomt binnen de bedrijfsrollen. Is dat niet het geval dan is het mogelijk dat een entiteit geen complete levensloop kan doorlopen. Zo is een entiteit zonder R bijzonder, het wordt wel aangemaakt en gewijzigd maar nooit gelezen.

Kenmerken

De CRUD matrix is al lang aanwezig in datamodelleerland en heeft haar kracht bewezen in meerdere verschijningsvormen. De belangrijkste (positieve en negatieve) kenmerken:

- Krachtige notatiewijze waarin drie concepten in één weergave met elkaar worden gecombineerd, namelijk CRUD bewerkingen (cel) data entiteit (As 1) en gedragsentiteit (As 2)
- Eenvoudig toepasbaar bij gebruik voor stakeholders zonder modelleerervaring
- Kan gebruikt worden in interactieve workshops
- Toepasbaar op meerdere abstractie niveaus, van fysiek tot conceptueel
- Geschikt voor notatie binnen data management context
- Goed toepasbaar binnen data security en - privacy context
- Kan omgezet worden in sourcecode bijvoorbeeld bij fysieke modellering in autorisaties en users voor tabellen in fysieke relationele modellen
- Modellen kunnen bij veel entiteiten en attributen en gedragsentiteiten lastiger leesbaar worden.

Toepassingen

CRUD matrix wordt in meerdere situaties toegepast maar de meest voorkomende zijn

- Autorisatie matrix (wie mag wat op welke data entiteit)

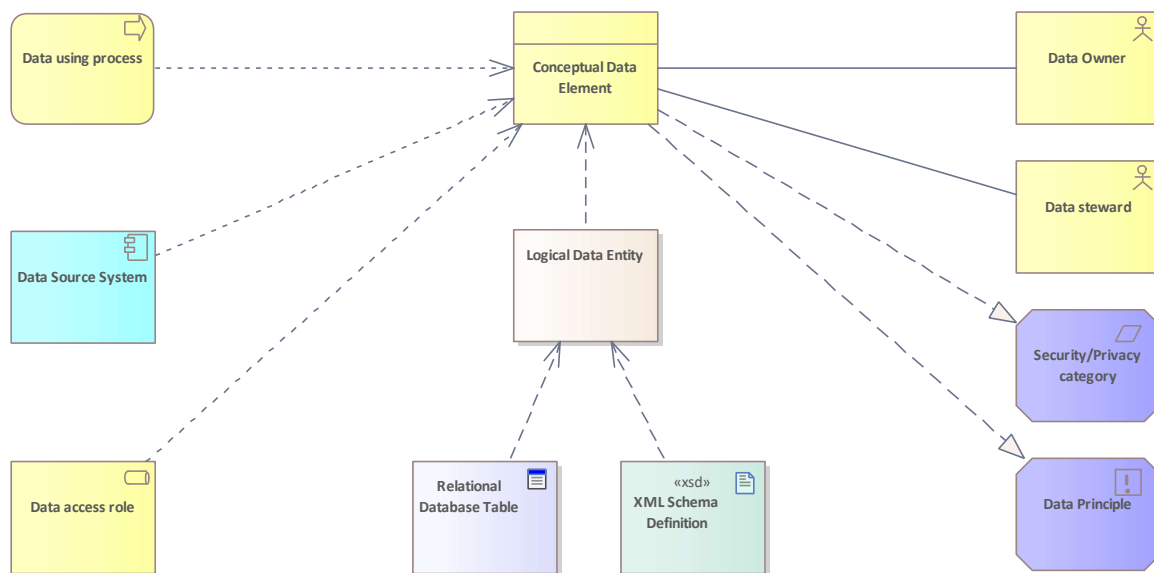
- Beschrijving van de bewerkingen die door een gedragsentiteit wordt uitgevoerd op een gegevens entiteit (dus een algemener perspectief dan de autorisaties)

Gerelateerde domeinen

- De CRUD matrix wordt vaak gecombineerd met logische en fysieke data modellen om hiermee de details van de entiteiten te modelleren. De CRUD wordt dan vervolgens gebruikt om autorisaties of acties op de entiteiten te modelleren.
- Een vergelijkbare notatiewijze uit de object modellering is de Object Gebeurtenis Tabel die veelal op een conceptueel model toegepast worden.
- Een reeds behandelde notatiewijze uit data governance is de RACI matrix die veelal op een conceptueel model toegepast wordt.

Meta Meta Model

In de uitwerking van het metamodel hebben we de negen domeinen van het metadata model afzonderlijk van elkaar beschreven. Echter de samenhang tussen de verschillende domeinen is voor een register vanzelfsprekend ook belangrijk. Daartoe is een eenvoudig meta meta model opgesteld. Hieronder een afbeelding met de opzet van dit model.



In de afbeelding is te zien hoe de domeinen aan elkaar gerelateerd zijn. Het conceptuele model is de centrale spil. Het logische en fysieke model zijn feitelijk detailleringen van het conceptuele model.

De andere domeinen zijn gerelateerd aan het conceptuele model. Voor deze domeinen en de verbindingen naar het conceptuele data model is gebruik gemaakt van de modelleertaal ArchiMate. Dus ondanks dat we gebruik maken van matrices in ons metamodel is ArchiMate het onderliggende model van het meta data register.

Tooling & aanvullende producten

Dit whitepaper is uitgewerkt in een generiek modelleertool Sparx Enterprise Architect. Reden waarom voor deze tool is gekozen is dat het meerdere modelleertalen ondersteunt waaronder ArchiMate, UML en ER en de matrix modellen. Neemt niet weg dat een dergelijk meta data register ook in andere tools ingericht kan worden. Bijvoorbeeld Visual Paradigm, Aris of Blue Dolphin.

Bij dit whitepaper horen een aantal aanvullende producten. Belangrijkste in deze is dat er een voorbeeld repository beschikbaar is waarin alle modellen uitgewerkt zijn. De afbeeldingen in dit whitepaper zijn gemaakt van de inhoud van deze voorbeeld repository. Heb je interesse in dit voorbeeld model neem dan contact met me op via email. Dan kan ik het voorbeeldmodel naar je toe zenden.

Evaluatie

In de voorgaande paragrafen is een metadata model uitgewerkt bestaande uit een combinatie van modelleertechnieken en -talen. Uit het voorgaande blijkt dat relatief eenvoudig een meta data register ingericht kan worden met modelleertalen als ArchiMate en UML binnen een generiek modelleertool.

Met name ArchiMate blijkt binnen het ingerichte model een centrale rol te spelen en zorg te dragen voor het verbinden van de verschillende meta data domeinen. UML klasse diagrammen en Fysieke datamodellering zijn feitelijk de enige twee domeinen waar een detaillering in het model gewenst is wat in ArchiMate niet gemodelleerd kan 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