

DATAMODELLERING DATA MAPPING MODEL

Inleiding

In dit whitepaper wordt de datamodelleervorm data mapping model beschreven. Deze modelleervorm staat in verhouding tot een aantal andere modelleervormen. Wil je een beeld krijgen van welke modelleervormen er zijn bekijk dan het whitepaper wat een introductie geeft tot datamodelleervormen en deze serie van whitepapers zie Inleiding datamodellering.

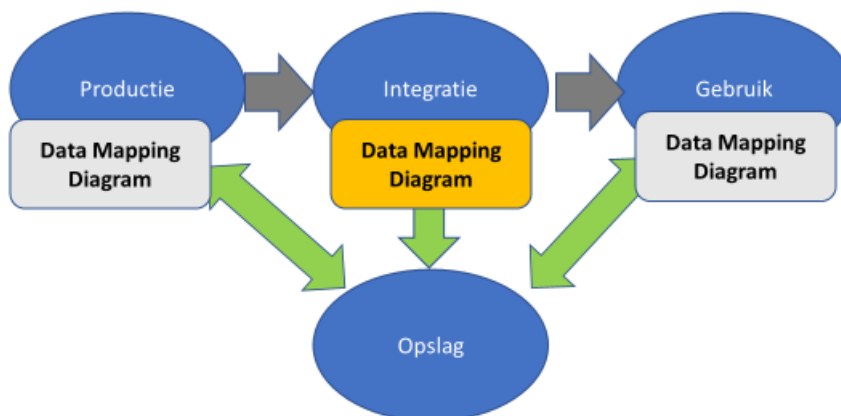
Doel

Data mapping diagrammen zijn een informele modelleerwijze die op attribuut of kolom niveau koppelingen in kaart brengt tussen twee entiteiten. Deze entiteiten zijn in hoofdzaak dataentiteiten zoals tabellen en klassen binnen de logische- en fysieke modellering. Echter ook koppelingen vanuit andere disciplines zoals data management en data security kunnen tot op attribuut niveau koppeling kennen. Deze willen we eveneens graag in een diagramvorm kunnen presenteren aan de stakeholders.

Bij data mappings wordt veelal de richting van de koppeling gemodelleerd in een verrijkte graaf. Hiermee wordt inzichtelijk gemaakt hoe de gegevens stromen van de ene attribuut/kolom naar de andere. Dit maakt het goed mogelijk om bron en doel van gegevensstromen op detail niveau inzichtelijk te maken.

Raamwerken

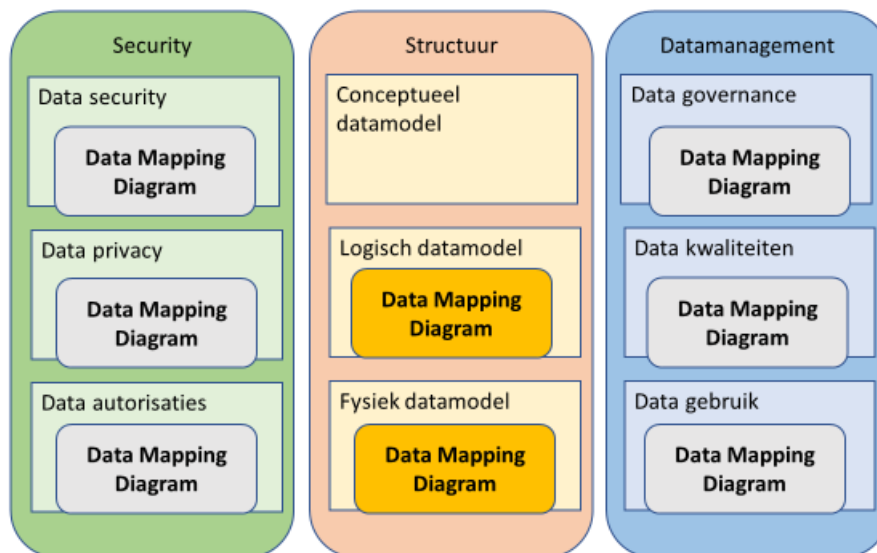
DATALEVENSLOOP



Binnen de data levensloop zijn data mappings met name van belang bij data integratie. Hier zijn gegevensstromen en de bron- en doel-dataentiteiten essentieel aspect van de modellering. Data mappings zullen in dit domein dan ook veelvuldig ingezet worden.

Voor de data productie en het data gebruik zijn data mappings minder relevant maar kunnen ze zeker toegevoegde waarde bieden. Bijvoorbeeld bij het in kaart brengen van attributen of kolommen getransformeerd naar besturingselementen in een grafische user interface is een vorm van mappings die in deze domeinen toegevoegde waarde heeft.

DATARAAMWERK



Data Mapping Diagrammen zijn met name relevant voor de logische en de fysieke modellering. In deze twee modelleervormen spelen attributen en kolommen een essentiële rol. In een aantal scenario's wil je dan ook inzichtelijk maken welke attributen/kolommen gerelateerd zijn aan kenmerken van andere data entiteiten en op welke wijze. Daarnaast worden in de logische- en fysieke modellering van mappings ingezet om de relatie tussen deze twee modellen in kaart te brengen. Dus hoe zijn de attributen uit de logische modellering gekoppeld aan de kolommen in de fysieke modellen.

Mappings tussen attributen/kolommen en kenmerken uit andere modelleerdomeinen worden minder frequent toegepast maar kunnen in bepaalde situaties zeker toegevoegde waarde bieden. Denk aan het in kaart brengen van data gebruik en -autorisatie op attribuutniveau in plaats van op entiteitniveau. Maar ook in situaties waar privacy of security maatregelen, zoals anonimiseren, tot op attribuutniveau beschreven dienen te worden.

Stakeholders

Data Mapping richt zich op stakeholders die zich voornamelijk op het raakvlak van binnen de business en ICT bevinden. Data Mappings brengen veelal verbanden in kaart rond data integratie en rollen in dit werkveld zullen vaak gebruik maken van data mappings. Echter worden data mappings ingezet voor visualisaties in data management en security dan wordt de groep van betrokken stakeholders natuurlijk groter.

Onderstaande opsomming geeft de belangrijkste stakeholders:

- **Functioneel (Applicatie) Beheerders**, een groep stakeholders die diepgaande kennis heeft van (standaard) informatiesystemen en het onderliggende logische datamodel inclusief de door de leverancier gebruikte datastructuren en eventueel opslagmechanismen. Mappings tussen deze verschillende modellen zijn voor hen noodzakelijk.
- **Applicatie- of Informatiearchitect**, zal veelal kennis hebben van het data integratie modellen huidige- en de gewenste situatie van een verandering. Ze hebben daarnaast veelal diepgaande kennis van data patronen, data structuren en logische data modellen.
- **Open standaarden**, ontwikkelen van logische informatiemodellen voor open standaarden worden veelal uitgewerkt in mappings waarbij de logische modellen uit de open standaarden worden gemapt naar de fysieke modellen in de eigen organisatie cq informatiesystemen.
- **Ketenintegraties**, binnen ketenintegratie is een overeenstemming rond de data die uitgewisseld wordt van groot belang. Vaak wordt begonnen met een logisch model opgesteld. Tussen de logische modellen en de fysieke implementaties wil men veelal tot op attribuutniveau een beeld hebben hoe de integratie verloopt
- **Ontwikkelaars**, met name op het vlak van data integratie. Denk hierbij aan ETL ontwikkelaars en XSD schemaontwerpers. Zij maken op fysiek niveau de koppelingen tussen de verschillende entiteiten en dragen zorg voor transformaties en dergelijke. Hiervoor is een goed inzicht in de mappings onontbeerlijk.

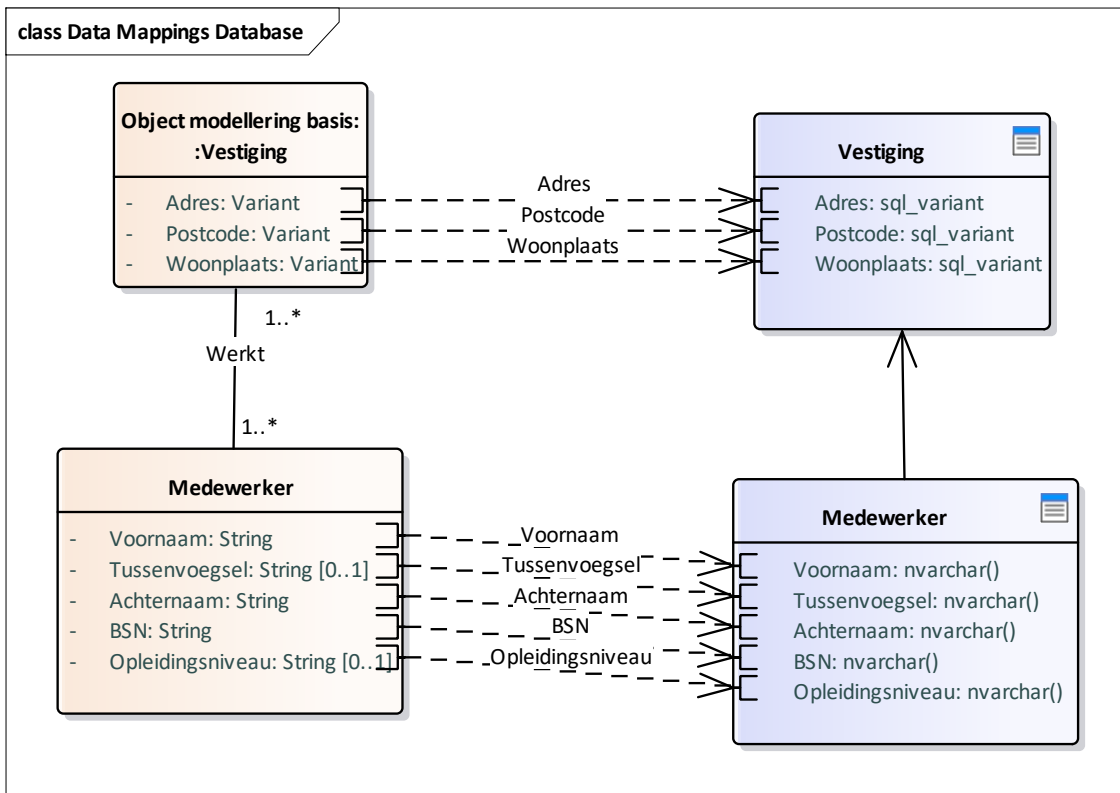
Concepten

Vanuit andere notatiewijzen zoals het UML klassediagram, het ER diagram of het XSD diagram worden een aantal concepten in deze notatiewijze met elkaar in verband gebracht. Daarnaast wordt er feitelijk maar één extra concept aan toegevoegd namelijk de mapping. Onderstaande opsomming geeft een overzicht:

- **Klasse/Tabel/Complextype**, entiteit waarvan de eigenschappen die aan elkaar gemapped worden onderdeel zijn.
- **Attribuut/Kolom/XSDElement**, eigenschap die met een eigenschap in een andere data entiteit gekoppeld of gemapped wordt
- **Mapping**, daadwerkelijke koppeling tussen twee of meer eigenschappen in de bovengenoemde data entiteiten
- **Overige datamanagement en security entiteiten**, denk hierbij aan data kwaliteiten, security maatregelen, gebruikers waarbij mappings worden gebruikt om verbanden tussen eigenschappen in data entiteiten met data management entiteiten inzichtelijk te maken.

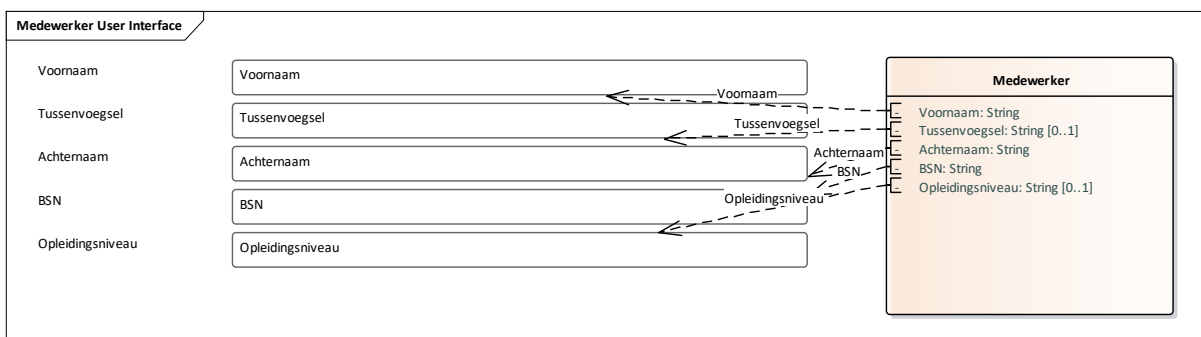
Notatie

De notatie is relatief eenvoudig en informeel, er wordt een verbinding gelegd tussen eigenschappen van data entiteiten. In onderstaande afbeelding een voorbeeld van een mapping tussen UML klassen en tabellen in een ER notatie



In het diagram is te zien hoe er tussen eigenschappen een koppeling wordt gelegd en dat iedere koppeling een naam heeft. Hier is het informele karakter van de notatie te zien, op dit vlak heeft iedere modelleur namelijk de vrijheid om een eigen inrichting te kiezen. Hetzelfde geldt voor de opzet van de associaties zelf, dat kunnen pijlen zijn (een- of twee richting), kleuren krijgen of andere verrijkingen van de graafnotatie voor data mappings.

Daarnaast is het mogelijk om ook andere vormen van mappings te kiezen waarbij een koppeling gelegd wordt tussen de eigenschappen van een data entiteit en een andersoortige entiteit binnen het te modelleren domein. In de afbeelding een voorbeeld van een mapping tussen eigenschappen en een user interface. Echter er zijn veel andere vormen van mappings mogelijk, naar data kwaliteiten, security maatregelen of bedrijfsrollen zijn willekeurige voorbeelden van mappings.



Kenmerken

Onderstaande opsomming geeft een aantal kenmerken van deze notatie:

- Eenvoudige en vrijwel zelf verklarende notatiewijze
- Niet geformaliseerd rond de verrijkingen van de associaties
- Diagrammen worden snel vol en moeilijk te interpreteren door de vele lijnen
- Inzetbaar voor het verbinden van eigenschappen met diverse data modelleringsconcepten (niet alleen beschrijvend)
- Zijn eenvoudig om te zetten naar gegenereerde programmacode

Gebruikstoepassingen

Data mappings worden in veel domeinen van de data modellering. Het meest kenmerkend zijn ze in de fysieke modellering voor ETL procesmodellering waarbij transformaties en gegevensstromen tussen bron- en doeltabellen worden gemodelleerd. Hetzelfde geldt voor het modelleren van XSD schema's en de mappings van de berichtonderdelen naar data entiteiten binnen applicaties of werkprocessen.

Kenmerkend en een noodzakelijk onderdeel van een gelaagd datamodel is de datamapping tussen de klassen in de logische laag en de fysieke data entiteiten zoals tabellen en berichten in de fysieke laag. Zie bijvoorbeeld de afbeelding in de voorgaande paragraaf.

Laatste toepassing is het leggen van mappings vanuit de data management perspectieven voor het leggen van koppelingen tussen management concepten en de eigenschappen van data entiteiten.

Gerelateerde notatiewijzen

Mappings is een veelgebruikte notatie als alternatief voor matrices. Ze hebben een vergelijkbare inrichting en zeker bij het gebruik van een verrijkte graaf in mappings zijn de overeenkomsten groot.

Mappings zijn nauw gerelateerd aan alle notatiewijzen waarbij de mapping een verbinding legt tussen de concepten van de verschillende notatiewijzen. De belangrijkste zijn:

- UML klassediagram
- ER diagram
- XSD schema

Tooling

Voor data mappings in data modellering zijn met name de tools ontstaan vanuit data perspectief van belang. Denk hierbij aan:

- Powerdesigner
- ERWin
- Sparx Enterprise Architect (met extensie)

Daarnaast zijn er een aantal specifieke tools op het vlak van data integratie die specifieke mapping functionaliteit bieden voor een bepaald werkveld. Bijvoorbeeld:

- Altova Mapforce (XSD)

- KNime (Data integratie en analyse)

Evaluatie

Data mappings zijn een krachtige, en daarom veelvuldig toegepaste, notatiewijze in de data modellering. Omdat het eigenschappen van verschillende data entiteiten met elkaar verbindt biedt het op een detailniveau de mogelijkheid om koppelingen te documenteren.

De notatiewijze is eenvoudig te begrijpen en informeel. De notatie kan uitgebreid worden met eigen interpretatie of verrijking van de associaties. Daarnaast kunnen associaties gelegd worden tussen willekeurige soorten data entiteiten en andere concepten in het domein. Dat is krachtig maar kan ook een risico vormen.

Voor data mappings is voldoende tooling aanwezig, echter het leggen van associaties op eigenschap niveau is niet altijd even intuïtief en kan behoorlijk arbeidsintensief zijn. Evalueer daarom tools voorafgaand op deze requirements.

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