

# DATAMODELLERING SCORE MATRIX

## Inleiding

In dit whitepaper wordt de datamodelleervorm Score Matrix 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 Modelleervormen

## Doel

De score matrix is een bijzondere vorm van data modelleren omdat het zowel voor de vraag- als aanbodkant van datasets gebruikt kan worden. Het doel is om eisen en requirements in relatie te brengen met data entiteiten en vervolgens een numerieke- of ordinale waarden worden toegekend.

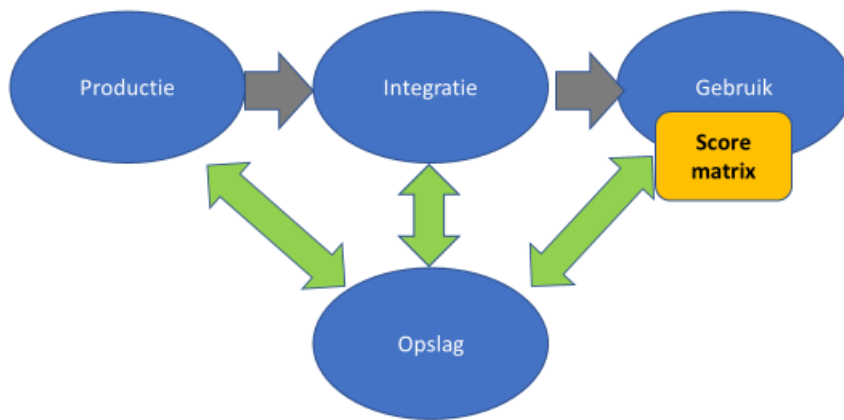
Score matrices zijn voor verschillende doeleinden te gebruiken, waarbij opvallend is dat dit zowel in de ontwikkelfase als in de beheerfase hulp biedt, hierbij komt wederom zowel de vraag- als aanbodzijde aan bod.

Als laatste is te noemen dat de score matrix op meerdere wijzen ingezet kan worden. Een kenmerkend werkveld zijn data kwaliteiten binnen een data management raamwerk. Tweede karakteristieke werkveld is requirements- en constraint modellering in relatie tot data entiteiten. Relevant bij met name aanbestedingstrajecten.

## Raamwerken

### DATALEVENSLLOOP

In onderstaande afbeelding wordt de score matrix notatie afgebeeld op de data levensloop.



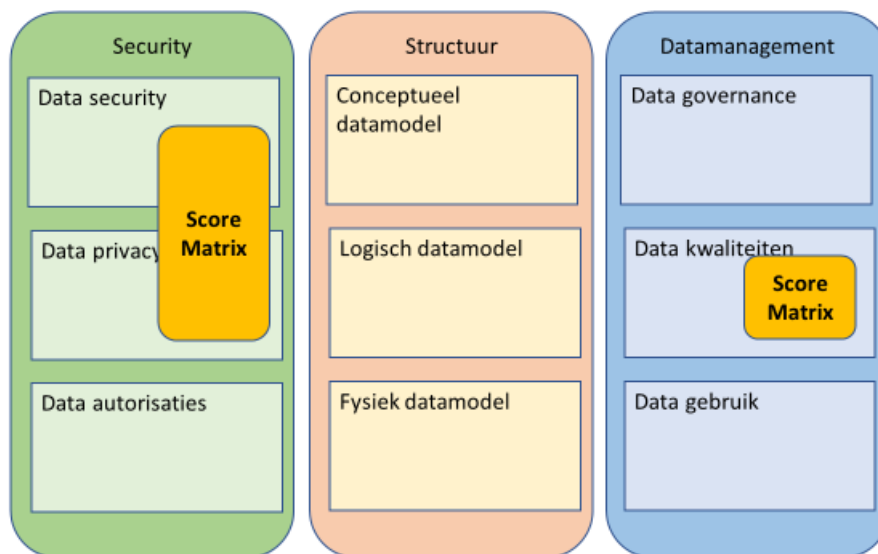
In de afbeelding is te zien dat de Score matrix voornamelijk bij datagebruik wordt toegepast. Dat komt met name door het feit dat de score matrix voornamelijk binnen data management processen wordt ingezet.

In de data gebruik fase wordt gekeken welke dataset op welke wijze aan eisen of kwaliteit voldoet. Dat kan enerzijds de huidige situatie (de aanbodzijde) als de gewenste situatie (de vraagzijde) zijn.

Bij de data productie, data integratie en data opslag kan een matrix opgesteld worden van bijvoorbeeld data kwaliteiten in relatie tot deze fasen in de levensloop, echter dit wordt zelden toegepast. Zie voor een uitwerking van maatregelen behorend bij deze fasen "Data architectuur in de praktijk".

## DATARAAMWERK

In de onderstaande afbeelding is te zien dat de score matrix meerdere doeleinden dient en daarnaast relevant is voor meerdere toepassingen.



De score matrix is een belangrijk hulpmiddel bij het inventariseren van eisen, requirements en kwaliteiten in relatie tot data management, waarbinnen bijvoorbeeld data kwaliteiten een eigen werkveld hebben. De score matrix kan ingezet worden om te bepalen wat het huidige- en gewenste kwaliteitsniveau is van bepaalde data entiteiten. Hiermee kan gezocht worden naar voor welke data entiteiten in relatie tot de meest relevante data kwaliteiten maatregelen genomen dienen te worden.

Deze kwaliteiten en eisen worden veelal geïnteriseerd met behulp van de lijst van relevante data kwaliteiten binnen het werkveld van de organisatie. Vandaar dat de scorematrix in het werkveld data kwaliteiten een eigen en specifieke plaats heeft.

Daarnaast is de score matrix goed inzetbaar binnen data security en data privacy. Hierbij wordt veelal niet uitgegaan van kwaliteiten maar voornamelijk van eisen en requirements die aan data entiteiten gesteld worden. Ook hierbij kan zowel de huidige als de gewenste score matrix opgesteld worden (eventueel inclusief een verschillen matrix). Op basis van deze matrices kunnen vervolgens maatregelen bepaald worden die zorg moeten dragen aan een lager scoreverschil tussen de huidige en gewenste beveiliging/privacy situatie van de geanalyseerde data entiteiten.

## Stakeholders

De score matrix komt met name terug in de data management en data kwaliteiten dimensies van het data raamwerk. Ook binnen data security en data privacy komt de score matrix voor, dat is logisch omdat dit feitelijk verbijzonderingen zijn van datamanagement processen.

- **Data Owners**, eigenaren van data entiteiten zijn de key stakeholders van de score matrix, zij zijn enerzijds geïnteresseerd in de huidige scores omdat dit aangeeft wat het kwaliteitsniveau is van de data entiteiten. Anderzijds is de gewenste score voor hen van belang omdat dit aangeeft in welke mate maatregelen nodig zijn om de scores voor bepaalde data kwaliteiten te verhogen. zitten.

- **Data Stewards** zullen voor kleine aanpassingen binnen de operaties van data entiteiten behoefte hebben aan de huidige- en de gewenste kwaliteitsscore en of de operaties die uitgevoerd worden invloed hebben op deze scores
- **Data Security Officers en IB experts**, willen in kaart kunnen brengen met een score matrix in welke mate data entiteiten voldoen aan de security eisen die gesteld worden aan de data entiteiten en welke maatregelen nodig zijn indien de datasets onvoldoende aan de gestelde security eisen is voldaan
- **Data Privacy Officers of Data Controllers** willen in kaart kunnen brengen met een score matrix in welke mate data entiteiten voldoen aan de privacy eisen die gesteld worden aan de data entiteiten en welke maatregelen nodig zijn indien de datasets onvoldoende aan de gestelde eisen is voldaan vanuit bijvoorbeeld AVG perspectief
- **Data architecten, informatie analisten en procesmodelleurs**. Ieder op een verschillend abstractieniveau hebben behoefte aan de huidige en gewenste requirement- of kwaliteitsniveaus van data entiteiten.
- **Data Management en Data Operators**, in de beheerfase van het data landschap kunnen wijzigingen doorgevoerd moeten worden in het data gebruik en de daarbij horende structuur- en inrichting aanpassingen. Deze stakeholders willen daarom op eenvoudige wijze inzage kunnen krijgen in wat de score wijzigingen zijn door de benoemde wijzigingen.

## Concepten

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 DaMa raamwerk
- **Requirements en/of Constraints** waaraan in meer of mindere mate voldaan dient te worden
- **Eisen**, bijvoorbeeld vanuit perspectief van data security of -privacy

De derde dimensie legt een verband van de eisen/requirements naar de data entiteiten zoals:

- **Data object** in logische modellen en bij de modellering van applicatie gedrag op gegevens entiteiten
- **Bedrijfs- of business object** in conceptuele modellen

Desgewenst is er nog een vierde 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.

## 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, eis of requirement getoond. Er wordt geen onderscheid gemaakt over welke as welk entiteitstype wordt weergegeven.

Onderstaande afbeelding 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 DaMa. Let op dat een cel ook leeg kan zijn of een NaN waarde kan hebben

Target Source	Accuraatheid	Actualiteit	Compleetheid	Consistentie	Precisie	Privacy	Redelijkheid	Referentiele integriteit	Tijdigheid	Uniekheid	Validiteit
Cursus	8	8	8	8	NaN		NaN	NaN	8	9	6
Docent	8	8	8	8	NaN	6	NaN	NaN	8	6	6
Training	8	8	8	8	NaN		NaN	NaN	8	9	6

## 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 modellerervaring.
- 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.

## Gebruikstoepassingen

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 notatiewijzen

De Score matrix wordt vaak gecombineerd met de ArchiMate Motivation extensie, bijvoorbeeld bij aanbestedingstrajecten.

## Tooling

Voor score matrices is algemene tooling beschikbaar variërend van heel eenvoudig tot zeer geavanceerd.

Hieronder een opsomming van de meest bekende tooling:

- Office Excel of Open Office Calc
- Sparx Enterprise Architect
- Mogelijk zijn de andere veel toegepaste tools eenvoudig configureerbaar voor een scorematrix

## Evaluatie

Score matrix is een datamodellering notatie waarmee een score, bijvoorbeeld van 0 - 10 worden gecombineerd met Data entiteiten en eisen, requirements of kwaliteiten. De notatie wordt toegepast op met name de conceptuele datamodellering.

Naast toepassingen in de datamodellering wordt de score matrix met name gebruikt binnen data management, data kwaliteiten, data security en data privacy. Hierbij gaat het veelal om twee perspectieven, bijvoorbeeld de huidige en de gewenste situatie.

Door de eenvoudige opzet van de notatiewijze is de score matrix ook goed toepasbaar voor betrokkenen met weinig data modelleer ervaring. De notatie is goed inzetbaar bij inventarisaties of binnen interactieve workshops. Er is voldoende tooling beschikbaar voor het samenstellen van score matrices.

## 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](mailto:bert@interactory.nl)