

# Data modelleren in de praktijk



# Data modelleren in de praktijk

Bert Dingemans

Schrijver: Ir. Ing. Bert Dingemans

ISBN: 9789402187960

© Ir. Ing. Bert Dingemans

## Voorwoord

Een boek over data modelleren in dit digitale tijdperk, is dat niet vreemd? Als ik kijk naar mijn dagelijkse werkzaamheden waarin data modellering een centrale plek inneemt dan denk ik van niet.

Een naslagwerk met modelleervormen, toepassingsgebieden en een verwijzing naar relevante hyperlinks kan het modelleren eenvoudiger maken.

Daarnaast is het materiaal ook digitaal beschikbaar op de website van de architectuur assistent.

Dit boek bestaat uit twee delen, in ieder deel wordt vanuit een ander gezichtsveld gekeken naar data en modelleren. Ieder deel bestaat uit een overzicht van verschillende uitwerkingen opgedeeld in hoofdstukken. Ieder hoofdstuk heeft binnen een deel een vergelijkbare opbouw.

In het eerste deel wordt gekeken naar de verschillende modelleervormen, afgebeeld op een eenvoudig raamwerk om aan te geven welk aspect van data modellering ondersteund wordt.

In het tweede deel ligt de focus op toepassingsgebieden. Binnen een toepassingsgebied worden een aantal modelleervormen gecombineerd om de stakeholders binnen dit toepassingsgebied ondersteuning te bieden bij het modelleren van deze toepassing.

Het boek is beschikbaar in een papieren- en een elektronische versie. Daarnaast is er een ondersteunende website [www.architectuurassistent.nl](http://www.architectuurassistent.nl) waarop diverse achtergrondartikelen, hulpmiddelen, checklists en registers te vinden zijn. Bij het laatste deel waarin de praktijkvoorbeelden beschreven zijn horen een aantal webvideo's. Deze video's ondersteunen de tekst en worden daarom van harte aanbevolen.

Zonder hulp van een aantal mensen had dit boek niet de huidige vorm gekregen. Twee personen wil ik met name noemen. Allereerst Ria van Rijn en Eric van Wanrooij voor hun inhoudelijke en tekstuele suggesties. Ten tweede mijn partner Anneke Hubert voor haar steun. Rest mij iedereen veel leesplezier te wensen.

Culemborg, Januari 2019.

# Inhoud

Voorwoord	5
Inhoud	7
Deel 1: Modelleervormen	15
Informatie- en Datamodellering	16
Inleiding	16
Historie	16
Data levensloop	17
Data pipe	18
Raamwerk	20
Stakeholders	23
Modelleerwijzen	24
Modelleervormen	25
Begrippenboom	26
Doel	26
Raamwerken	26
Stakeholders	28
Concepten	29
Notatie	30
Kenmerken	31
Gebruikstoepassingen	31
Gerelateerde notatiewijzen	32
Tooling	32
Evaluatie	33
SIPOC	34
Doel	34
Raamwerken	34
Stakeholders	36

Concepten	36
Notatie	37
Kenmerken	38
Gebruikstoepassingen	38
Gerelateerde notatiewijzen	39
Tooling	39
Evaluatie	40
ArchiMate datamodeltering	41
Inleiding	41
Doel	41
Raamwerken	41
Stakeholders	43
Concepten	44
Notatie	45
Kenmerken	47
Gebruikstoepassingen	47
Gerelateerde notatiewijzen	48
Tooling	48
Evaluatie	48
ArchiMate data & bedrijfsmodellering	50
Doel	50
Raamwerken	50
Stakeholders	52
Concepten	53
Notatie	54
Kenmerken	55
Gebruikstoepassingen	55
Gerelateerde notatiewijzen	56
Tooling	56



Evaluatie	56
ArchiMate data- & Applicatiemodellering	58
Doel	58
Raamwerken	59
Stakeholders	60
Concepten	61
Notatie	62
Kenmerken	63
Gebruikstoepassingen	63
Gerelateerde notatiewijzen	64
Tooling	64
Evaluatie	64
ArchiMate data motivation modellering	66
Doel	66
Raamwerken	67
Stakeholders	68
Concepten	69
Notatie	71
Kenmerken	72
Gebruikstoepassingen	73
Gerelateerde notatiewijzen	73
Tooling	74
Evaluatie	74
RACI Matrix	75
Inleiding	75
Doel	75
Raamwerken	76
Stakeholders	78
Concepten	79

Notatie	80
Kenmerken	81
Gebruikstoepassingen	82
Gerelateerde notatiewijzen	82
Tooling	82
Evaluatie	83
Score Matrix	84
Doel	84
Raamwerken	84
Stakeholders	87
Concepten	88
Notatie	89
Kenmerken	89
Gebruikstoepassingen	90
Gerelateerde notatiewijzen	90
Tooling	90
Evaluatie	91
Basis UML Klasse model	92
Doel	92
Raamwerken	93
Stakeholders	95
Concepten	96
Notatie	96
Kenmerken	98
Gebruikstoepassingen	99
Gerelateerde notatiewijzen	99
Tooling	99
Evaluatie	100
Geavanceerd UML Klasse model	101

Doel	101
Raamwerken	102
Stakeholders	104
Concepten	105
Notatie	105
Kenmerken	107
Gebruikstoepassingen	107
Gerelateerde notatiewijzen	107
Tooling	108
Evaluatie	108
Data flow diagram	109
Doel	109
Raamwerken	109
Stakeholders	111
Concepten	111
Notatie	112
Kenmerken	112
Gebruikstoepassingen	113
Gerelateerde notatiewijzen	113
Tooling	114
Evaluatie	114
CRUD Matrix	115
Doel	115
Raamwerken	116
Stakeholders	118
Concepten	118
Notatie	120
Kenmerken	120
Gebruikstoepassingen	121

Gerelateerde notatiewijzen	121
Tooling	122
Evaluatie	122
ER diagram	123
Doel	123
Raamwerken	124
Stakeholders	126
Concepten	126
Notatie	127
Kenmerken	129
Gebruikstoepassingen	129
Gerelateerde notatiewijzen	130
Tooling	130
Evaluatie	131
Data Mapping model	132
Doel	132
Raamwerken	133
Stakeholders	135
Concepten	136
Notatie	136
Kenmerken	138
Gebruikstoepassingen	138
Gerelateerde notatiewijzen	138
Tooling	139
Evaluatie	139
XML Schema Definitions	140
Inleiding	140
Doel	140
Raamwerken	140

Stakeholders	142
Concepten	143
Notatie	144
Kenmerken	146
Gebruikstoepassingen	147
Gerelateerde notatiewijzen	147
Tooling	147
Evaluatie	148
Deel 2: Modelleer toepassingsgebieden	149
Service Oriented Architecture (SOA)	150
Doel	150
Context	150
Notatiewijzen	153
Kenmerken	159
Producten	160
Tooling	160
Evaluatie	161
Data analytics	162
Doel	162
Context	162
Analytics en data modellering	164
Notatiewijzen	165
Data autorisaties	172
Kenmerken	174
Producten	175
Tooling	175
Evaluatie	175
Data Governance	177
Inleiding	177

Doel	177
Context	178
Notatiewijzen	181
Kenmerken	184
Producten	185
Tooling	185
Evaluatie	185
Data quality	187
Doel	187
Context	187
Notatiewijzen	189
Kenmerken	193
Producten	194
Tooling	194
Evaluatie	194
Security en Privacy	196
Doel	196
Context	197
Notatiewijzen	200
Data autorisaties	204
Kenmerken	205
Producten	206
Tooling	206
Evaluatie	206
Referenties	207
Hyperlinks	209

## **Deel 1: Modelleervormen**

Informatie- en datamodellering is een belangrijk onderdeel van veel werkvelden in de bedrijfskunde en de informatiekunde. Denk bijvoorbeeld aan informatie analyse, data of enterprise architectuur, systeemontwerp, gegevensuitwisseling, service oriëntatie, database ontwerp en datastroommodellen.

Door deze verschillende verschijningsvormen is er in de loop der jaren een groot aantal modelleerwijzen ontstaan rond informatie- en datamodellering. Elke modelleerwijze heeft eigen specifieke kenmerken waarmee ze enerzijds inzetbaar zijn voor bepaalde groepen stakeholders en anderzijds inzetbaar zijn om een bepaald aspectgebied van data inzichtelijk te maken.

Bij het uitwerken van informatie- of datamodellen kan daarom gezocht worden naar welke modelleervorm het beste aansluit bij de doelgroep cq stakeholders. Maar ook het duidelijkst de relevante aspecten helder in kaart brengt. Het kiezen van de juiste modelleerwijze kan in deze bijdragen aan het op adequate wijze overdragen van informatie omtrent het domein dat gecommuniceerd wordt aan de stakeholders.

Gezien de vele modelleervormen, aspectgebieden en stakeholders is het vinden van een goede modelleervorm niet altijd eenvoudig. Reden om binnen de architectuur assistent een boek te schrijven waarin de belangrijkste datamodelleervormen worden behandeld op basis van een eenvoudig raamwerk.

Naast dit boek is er een training ontwikkeld waarin de verschillende modelleervormen op praktische wijze worden toegelicht:

<https://www.itmg.nl/training/masterclass-data-en-informatie-modellering/>

# Informatie- en Datamodellering

## Inleiding

Dit hoofdstuk is een introductie op een aantal hoofdstukken met datamodelleervormen en beschrijft de kaders en uitgangspunten die we gaan gebruiken bij het beschrijven van de verschillende modelleervormen.

## Historie

Data is al eeuwenoud, misschien zelfs al vanaf het eerste moment dat we zaken gingen opschrijven. Door de jaren heen is de hoeveelheid data en de complexiteit van datastructuren steeds groter geworden. Door de komst van geautomatiseerde toepassingen is deze ontwikkeling explosief toegenomen.

Sinds enige jaren is het concept Big Data geïntroduceerd. Dit concept geeft aan hoe omvangrijk data tegenwoordig is. Echter naast de hoeveelheid zijn ook de snelheid en de veelvormigheid van data complicerende factoren geworden (3Vs). Via de link

<https://www.winshuttle.com/big-data-timeline/> krijg je een aardig beeld van de ontwikkelingen rond data en big data door de tijd heen te zien.

Door deze verhoging van de complexiteit is de behoefte aan het opstellen van informatie- en datamodellen meegegroeid. De eerste vormen van data modellering stammen uit de jaren 60. Zo is door Codd in 1969 het relationeel datamodel geïntroduceerd

([https://en.wikipedia.org/wiki/Relational\\_model](https://en.wikipedia.org/wiki/Relational_model)). Dit model wordt nog steeds toegepast in het database werkveld.

Naast de relationele modellen zijn er door de jaren heen veel nieuwe datamodelleertechnieken ontstaan. Denk aan object georiënteerde modellen zoals UML of flow diagrammen zoals DFD. Zo is er een veelheid aan verschillende informatie- en datamodelleringswijzen ontstaan.

## Data levensloop

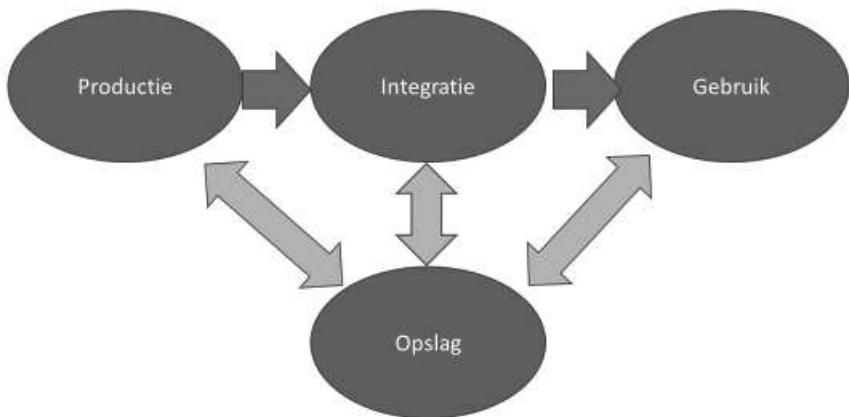
Naar informatie en data kan op veel manieren gekeken worden. Veel modelleervormen gaan uit van een bepaald gezichtspunt. Echter in dit boek willen we graag raamwerken gebruiken die in elke situatie gebruikt



kunnen worden om de verschillende concepten in een modelleerwijze op te plotten.

In de afgelopen jaren ben ik een aantal raamwerken gaan inzetten die het mogelijk maken om verschillende aspecten van data modellering af te beelden. In deze serie artikelen gebruiken we er drie die we in deze en komende paragrafen nader zullen toelichten. Bij het uitwerken van een modelleerwijze gebruiken we op een gestandaardiseerde wijze de modelleervorm te beschrijven.

Het eerste raamwerk is de data levensloop. Deze levensloop toont hoe data in een beperkt aantal stappen verschijnt. Onderstaande afbeelding toont het raamwerk.



Toelichting op het raamwerk:

- **Productie**, data wordt geproduceerd en ontstaat daardoor. Dat kan op vele manieren zijn. Denk hierbij aan personen die gegevens invoeren via formulieren, logs en dergelijke in informatiesystemen of devices zoals smart meters en mobiele telefoons die gegevens produceren.

- **Gebruik**, data die geproduceerd wordt zal op zeker moment gebruikt worden, bijvoorbeeld bij het nemen van beslissingen op basis van een data analyse, alerting bij afwijkingen in de data productie of het gebruik in werkprocessen (op basis waarvan beslissingen genomen worden).
- **Integratie of transport**. Soms is er een fysieke of modelmatige afstand tussen de data die geproduceerd wordt en de data die gebruikt wordt. Denk bijvoorbeeld aan de verschillen tussen transactionele- en DWH datamodellen. Maar ook aan de plaats waar de data geproduceerd wordt (smart meter in het veld) en de plaats waar het gebruikt wordt (besturingscentrum).
- **Opslag**, Opslag maakt data persistent bijvoorbeeld wanneer er een tijdsverschil is tussen data productie en gebruik of als de data op een later tijdstip of andere context opnieuw gebruikt kan worden in een andere gebruiksvorm. Daarnaast kan opgeslagen data gebruikt worden om efficiëntie van productie en gebruik te verhogen (hergebruik van data). Reden om vanuit iedere stap in de levensloop een pijl te tonen waarbij de data beide richtingen op kan stromen (van en naar de opslag).

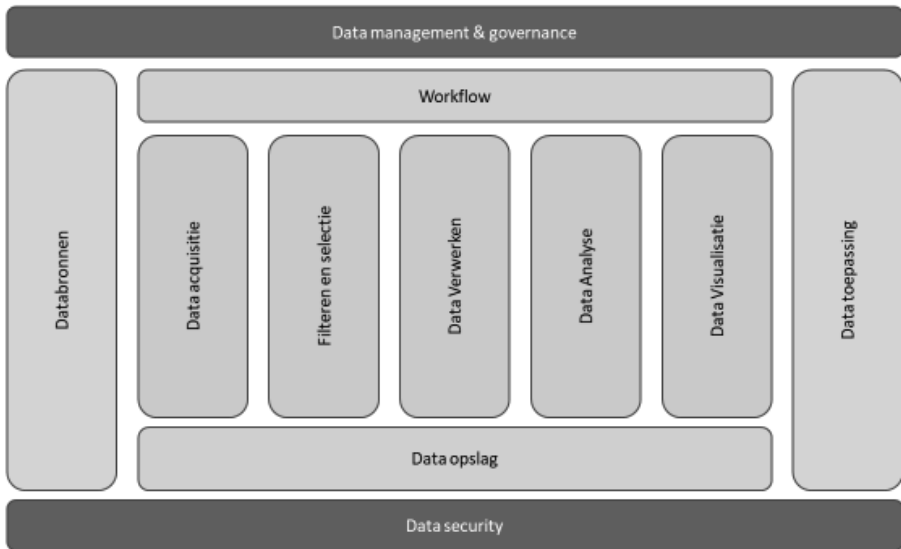
Bij een data levensloop wordt wel eens verondersteld dat het ontstaan, mutatie en einde van data entiteiten weergeeft. Dat is niet het geval. Ook al is dat een interessante modelleervorm (zie bijvoorbeeld de Object Gebeurtenis Tabel) deze levensloop heeft een beperktere scope. Desgewenst is het wel mogelijk meerdere levenslopen achter elkaar te plaatsen als dit relevant is.

In de paragraaf over stakeholders wordt een voorbeeld getoond hoe de levensloop toegepast kan worden bij het toelichten van de datamodelleringsvormen

## Data pipe

De data pipe is een detaillering van de datalevensloop en is wordt veel toegepast in (big) data integratie projecten. Het is feitelijk een raamwerk waarin je verschillende projectactiviteiten, deliverables en modelleervormen kunt afbeelden. Dit helpt om de complexiteit op

eenvoudige wijze in kaart te brengen. Onderstaande afbeelding toont de data pipe.



De datapipe wordt gelezen van links naar rechts en toont de stappen die genomen moeten worden om de productie van data in te kunnen zetten in een data toepassing (datagebruik). Daarnaast zijn er een aantal extra dimensie toegevoegd namelijk data management en security aspecten. In onderstaande opsomming een korte toelichting op de onderdelen:

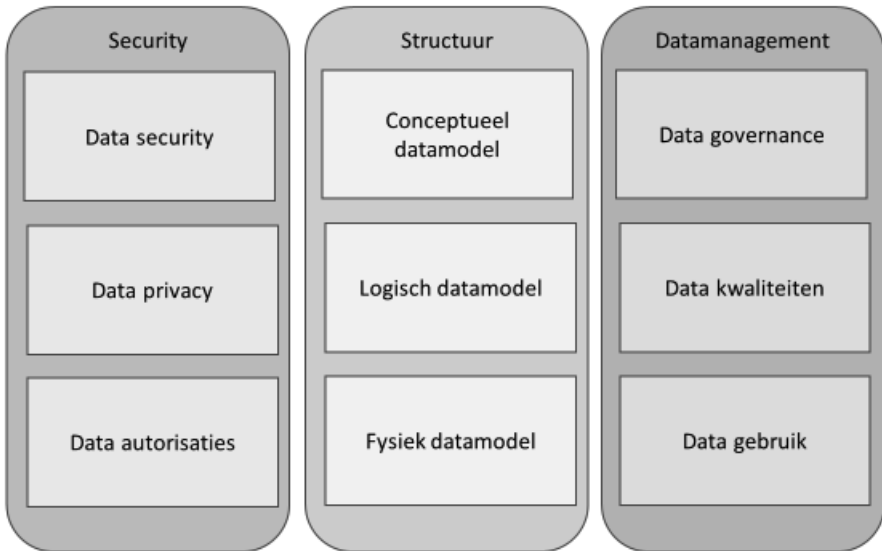
- **Databronnen**, gegevensverzameling die gebruikt worden als grondstof voor de data toepassing
- **Data acquisitie**, activiteiten die de verkrijging van de relevante databronnen bewerkstelligen
- **Filteren en selectie**, bewerken van de gegevens vanuit de databronnen tot die datasets die relevant zijn voor de toepassing
- **Verwerken**, transformatie, manipulatie en verrijking om het datamodel geschikt te maken voor een adequate data analyse, visualisatie en toepassing

- **Analyse**, activiteiten waarbij de getransformeerde data gebruikt wordt voor het zoeken naar verbanden, patronen of statistische verhoudingen
- **Visualisatie**, zichtbaar maken van de analyse resultaten ter ondersteuning van de analisten of voor presentatie aan andere stakeholders
- **Toepassen**, inzet van data in verschillende vormen van besluitvorming
- **Dataopslag**, opslag van de data en tussenproducten voor later gebruik in vervolgstappen of andere datapipes
- **Security**, beveiligings- en privacy aspecten van data
- **Data management**, data is een asset en daarom dient er management op plaats te vinden om de waarde van data te verhogen of behouden
- **Workflow**, automatiseren en standaardiseren van bewerkingsstappen op de data

Zoals reeds genoemd is dit een detaillering van de datalevensloop en gaat dit met name in op de integratie aspecten. Dit kan betekenen dat dit raamwerk niet bij elke datamodelleervorm relevant is.

## **Raamwerk**

Naast het meer dynamische levensloop aspect is het mogelijk om datamodelleervormen op een architectuurraamwerk af te beelden. Dit raamwerk is specifiek voor datamodellering en bestaat (zoals wel vaker) uit een drie maal drie raamwerk. Onderstaande afbeelding toont dit raamwerk.



Dit raamwerk gaat uit van een drietal kolommen:

- **Structuur**, modelleren van de structuur van data zoals het opgeslagen, of toegepast wordt.
  - Fysiek datamodel, modellering van data gebaseerd op het technische platform of implementatie, bijvoorbeeld bij data opslag en transport. Dit is veelal specifiek en gebaseerd op een gekozen technologie
  - Logisch model is een platform onafhankelijk model en wordt meestal gebruikt voor een abstracte weergave van een fysiek datamodel
  - Conceptueel model is het meest abstracte model voor het beschrijven van datastructuren. Het laat veelal details van structuur weg maar gaat in op hoe data zich verhoudt tot andere concepten als bedrijfsprocessen, informatiesystemen of organisatie eenheden