Grip op data modelleren

Grip op data modelleren

Bert Dingemans

Schrijver: Ir. Ing. Bert Dingemans
Coverontwerp: Rudy Barten
ISBN: 9789465125978
© Ir. Ing. Bert Dingemans 2025

Voorwoord

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

Dit boek bestaat uit twee delen. In ieder deel wordt vanuit een ander gezichtsveld gekeken naar data en modelleren. In het eerste deel worden de verschillende modelleervormen beschreven. Dit afgebeeld op een eenvoudig raamwerk om aan te geven welk aspect van datamodellering ondersteund wordt in de desbetreffende modelleervorm.

In het tweede deel wordt een datamodel uitgewerkt van een fictieve case de Alberto ijssalons. Het datamodel is een uitwerking van alle besproken datamodelleerwijzen. In het model is een hybride model uitgewerkt dat alle notatiewijzen met elkaar tot één gezamenlijk en geïntegreerd model maakt.

Zonder hulp van een aantal mensen had dit boek niet de huidige vorm gekregen. Een aantal personen wil ik met name noemen. Mijn dochter Karin heeft dit document gereviseerd en verbeteringen aangebracht. De leden van de werkgroep Metadata van DaMa-NL hebben de initiële whitepapers gereviewd en suggesties voor verbetering gedaan.

Verder de cursisten bij klassikale trainingen voor data- en informatiemodelleren. Zij hebben suggesties gedaan voor verbeteringen en een selectie van de te beschrijven modelleervormen.

Culemborg, januari 2025.

Combinatiepakket producten

Dit boek “Grip op data modelleren” is deel van een combinatie van producten. Hiermee is het voor een datamodelleerteam mogelijk om op basis van deze combinatie een gestructureerde en effectieve wijze een nieuwe datamodelleer werkwijze te introduceren. In onderstaande overzicht de producten behorend bij dit boek.

Voor de lezers van het papieren boek zijn de producten uit het combinatiepakket te vinden via:

https://data-docent.nl/datamodellerenlinksNL.aspx

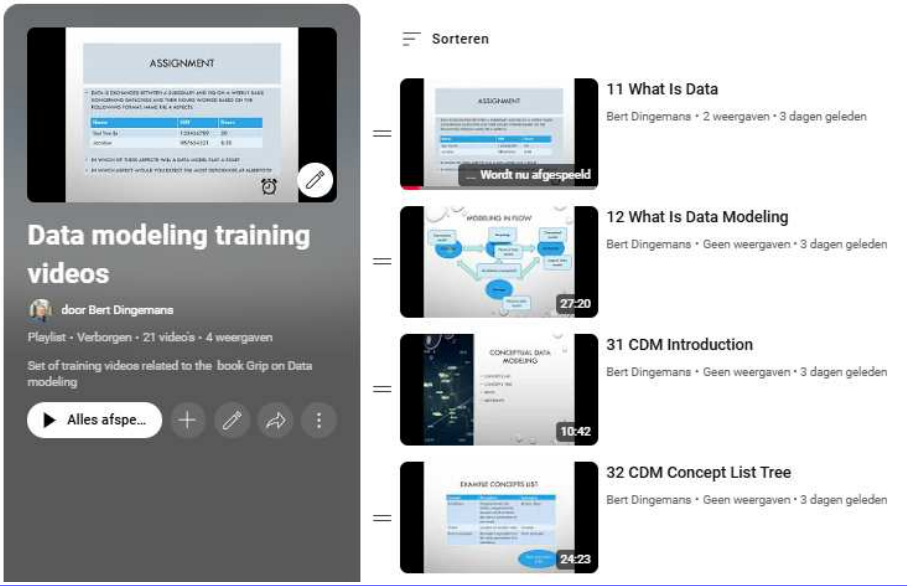
Voorbeeld datamodellen



Voorbeeld datamodellen met daarin een uitwerking van de content zoals uitgewerkt is in dit boek. De repository is een Sparx Enterprise Architect file die met deze tool kan worden geraadpleegd en desgewenst uitgebreid naar de eigen context. De modellen zijn ook beschikbaar in het XMI uitwisselformaat voor andere tools.

Online training data en informatie modelleren

Webbased cursusmateriaal behorend bij de modelleerwijzen over data en informatie modelleren (Engels). Van ieder onderwerp besproken in dit boek wordt in videolessen ingegaan op diverse aspecten van de afzonderlijke datamodelleervormen.



Modelleervormen

Informatie- en datamodellering is een belangrijk onderdeel van veel werkvelden in de bedrijfskunde en de informatiekunde. Denk bijvoorbeeld aan informatieanalyse, 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 c.q. 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.

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.

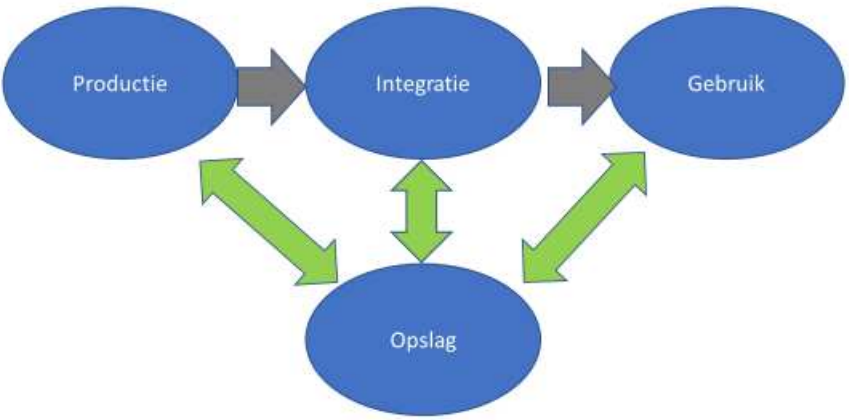
Historie

Data is al eeuwenoud, misschien zelfs al bestaand 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 zijn er meerdere ontwikkelingen te zien in het toepassen van data. Denk aan Data Analyse, Business Intelligence, Kunstmatige Intelligentie en Data Management. Deze maatschappelijke stromingen geven aan hoe omvangrijk datatoepassingen tegenwoordig zijn.
Door deze toename van de complexiteit in data 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. 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 flowdiagrammen zoals DFD. Zo is er een veelheid aan verschillende informatie- en datamodelleringswijzen ontstaan.

Datalevensloop

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

datamodellering af te beelden. In dit boek gebruiken we er drie die we in deze en komende paragrafen nader zullen toelichten. We gebruiken deze gestandaardiseerde wijze om iedere de modelleervorm te beschrijven.
Het eerste raamwerk is de datalevensloop. Deze levensloop toont hoe data in een beperkt aantal stappen verschijnt. Onderstaande afbeelding toont het raamwerk.



Toelichting op het raamwerk:

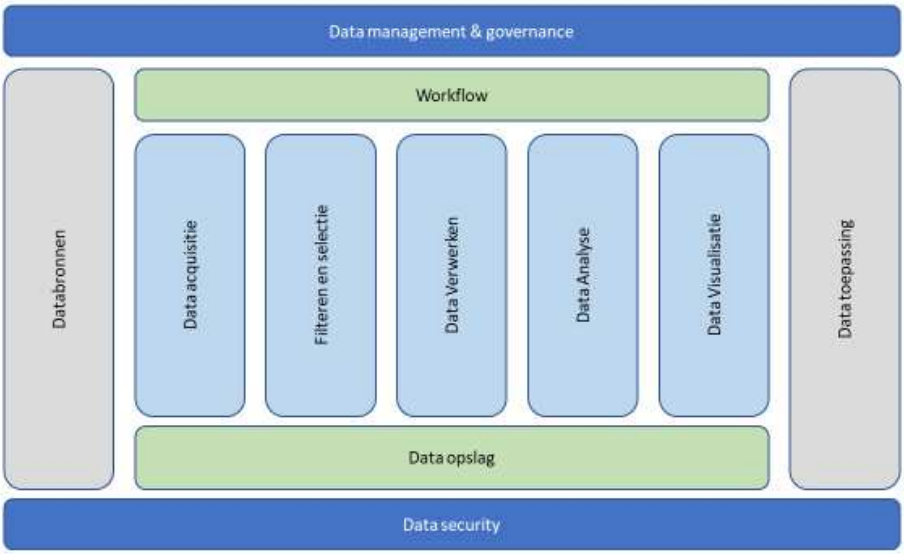
- **Productie:** data wordt geproduceerd en ontstaat daardoor. Dat kan op vele manieren zijn. Denk hierbij aan personen die gegevens invoeren via formulieren, logs en dergelijke in informatiesystemen of devices zoals smart meters en mobiele telefoons die gegevens produceren.
- **Gebruik:** data die geproduceerd wordt zal op zeker moment gebruikt worden, bijvoorbeeld bij het nemen van beslissingen op basis van een dataanalyse, alerting bij afwijkingen in de dataproductie 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 dataproductie 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 datalevensloop wordt wel eens verondersteld dat deze 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 wordt veel toegepast in (big) data-integratieprojecten. 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 datatoepassing (datagebruik). Daarnaast zijn er een aantal extra dimensies toegevoegd namelijk data management en security aspecten. In onderstaande opsomming een korte toelichting op de onderdelen:

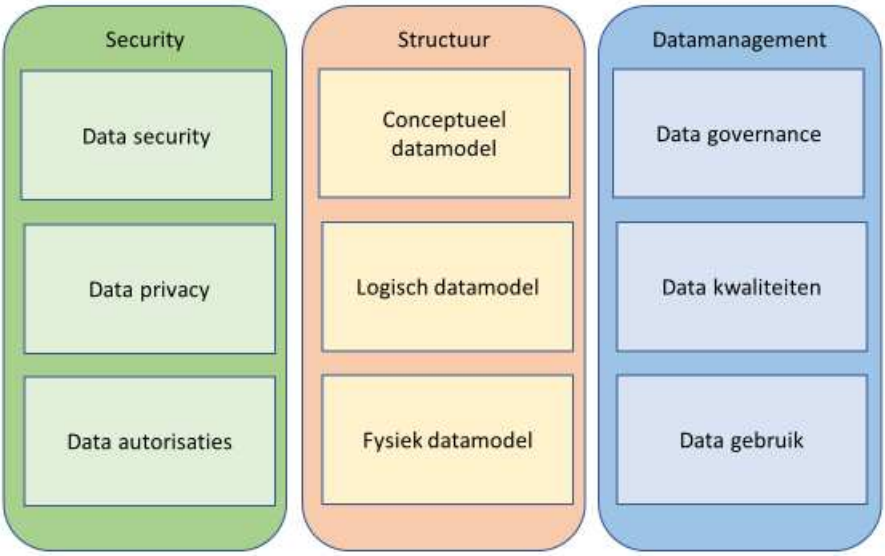
- **Databronnen:** gegevensverzameling die gebruikt worden als grondstof voor de data toepassing.
- **Data acquisitie:** activiteiten die de verkrijging van de relevante databronnen bewerkstelligen.
- **Filteren en selectie:** bewerken van de gegevens vanuit de databronnen tot die datasets die relevant 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:** het zichtbaar maken van de analyseresultaten 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 datastromen.
- **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:** het 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.

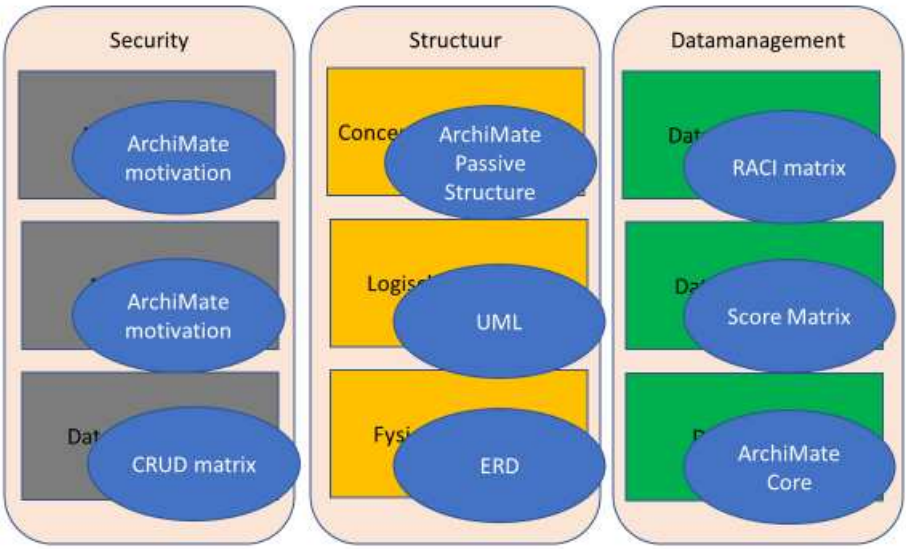


Dit raamwerk gaat uit van een drietal kolommen:

- **Structuur:** het 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**: een platform-onafhankelijk model dat meestal gebruikt wordt voor een abstracte weergave van een fysiek datamodel.
- **Conceptueel model**: 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
- **Security:** data is de drager van de informatie-inhoud en security is daarom een belangrijk aspect ter bescherming van deze informatie.

- **Data-autorisaties**: modellen om te beschrijven op welke wijze en voor welke persoon of toepassing data-entiteiten toegankelijk zijn.
- **Dataprivacy**: privacy is een bijzondere vorm van security gericht op het beschermen van persoonsgegevens.
- **Security**: maatregelen en risico's rond het gebruik en met name de misbruik van gegevensverzamelingen.
- **Data Management**, data is een asset voor veel organisaties en heeft daarmee waarde. Data Management zijn de activiteiten die zorgdragen dat deze waarde van data behouden of verhoogd wordt.
- **Datagebruik**: op welke wijze en door welke personen en toepassingen wordt de data gebruikt.
- **Datakwaliteiten**: data heeft bepaalde kwaliteiten waarmee de waarde van data bepaald kan worden. Op basis hiervan kunnen data verhogende maatregelen genomen worden.
- **Data governance**: inrichting van de organisatie ten behoeve van het bewaken van de kenmerken die data tot een asset maken voor de organisatie.

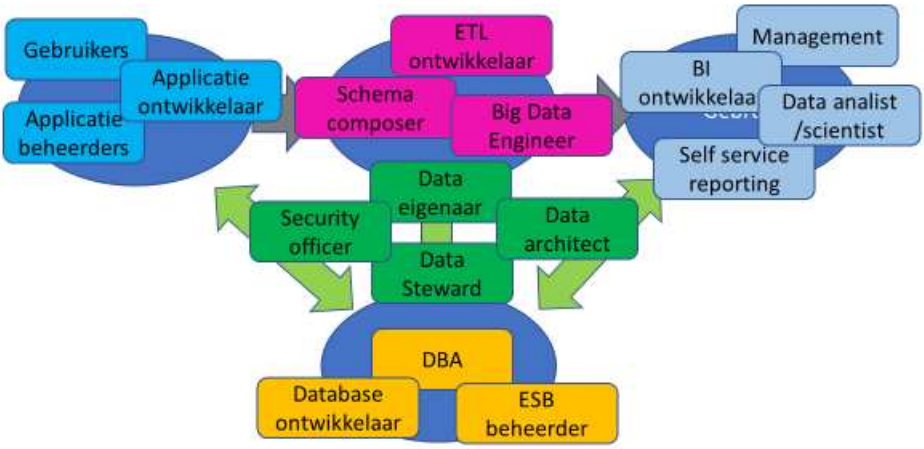
Het model is weergegeven als een drie bij drie matrix. Feitelijk zijn in dit model ook een drietal lagen te definiëren. De indeling van de concepten in de kolommen houdt hier reeds rekening mee. Echter, de dimensies van de lagen in de verschillende kolommen zijn verschillend. Vandaar dat er geen lagen als bedrijfs- en applicatie laag in deze situatie niet toepasbaar zijn. We gebruiken daarom de lagen wel maar beelden deze niet af in het raamwerk.
Dit raamwerk kan goed gebruikt worden om de verschillende modelleervormen tot elkaar in verhouding te brengen. Onderstaande afbeelding toont op welke wijze dit gedaan kan worden. Dit voorbeeld toont slechts een deel van de modelleervormen maar geeft een goed beeld van het toepassen van het raamwerk.



Stakeholders

Opstellen van (data)modellen doe je hopelijk niet zonder reden. Veelal is het gericht om betrokkenen of stakeholders te informeren, te overtuigen of besluiten te laten nemen over een bepaald aspect van het toepassen van data. Dit toepassen kan vanzelfsprekend zijn op vele gezichtspunten betrekking hebben. Denk hierbij bijvoorbeeld aan data-transport en -integratie, opslag, en ook op het vlak van data toepassen in werkprocessen of informatiesystemen. Stakeholders hebben kenmerken die van invloed zijn op de keuze van de juiste modelleervorm. Bijvoorbeeld, ontwikkelaars of beheerders hebben behoefte aan meer details op het fysieke of implementatievlak. Echter zijn gebruikers of vertegenwoordigers van de business meer geïnteresseerd in hoe de concepten toegepast worden in een organisatie. Reden om stakeholders in kaart te brengen en te benoemen voor de verschillende modelleerwijzen. Stakeholders kennen veel verschijningsvormen en het maken van een indeling of raamwerk is daarom niet goed mogelijk. De eerder

benoemde indelingsvormen zijn hulpmiddelen die daarbij kunnen helpen. Onderstaande afbeelding toont een voorbeeld van hoe stakeholders afgebeeld worden op de datalevensloop.



Modelleerwijzen

In dit deel wordt een groot aantal modelleervormen beschreven. Modelleervormen zijn gebaseerd op modelleerwijzen. De modelleerwijzen zijn feitelijk modelleerpatronen op basis waarvan een modelleervorm wordt uitgewerkt.
Binnen datamodelleren worden een beperkt aantal modelleerwijzen toegepast. Sommige zijn veelvoorkomend, anderen worden slechts sporadisch toegepast. De belangrijkste modelleerwijzen zijn:

- **Lijsten:** weergave van data-entiteiten in opsommingen en lijstweergaven. Een veel toegepaste modelleerwijze binnen modelleertools om snel overzicht te krijgen van de beschikbare data-entiteiten. Binnen modelleervormen zijn de lijsten weinig toegepast omdat er aan de vorm weinig toegevoegd kan worden vanuit modelleerperspectief.

- **Bomen:** weergave van data-entiteiten in een hiërarchische boomstructuur met veelal één “root” of begin entiteit waar vervolgens nul, één of meerdere “child” of kind entiteiten zijn gekoppeld. Kenmerkend is dat een kind entiteit slechts één ouder entiteit heeft. Voorbeeld van een boom is de begrippenboom of thesaurus.
- **Matrices:** weergave van de relaties tussen twee soorten entiteiten in een tabel of matrix weergave. Matrix weergave is goed bruikbaar in situaties waar tussen twee unieke entiteitsoorten slechts een koppeling kan bestaan. Vaak wordt de koppeling (lees matrix cel) verrijkt met extra selecties om een vierde dimensie toe te voegen die de koppeling typeert. Denk aan de CRUD matrix.
- **Grafen:** gebaseerd op de wiskundige grafentheorie waarbij entiteiten (nodes) via associaties of koppelingen (edges) aan elkaar gekoppeld zijn. Grafen notatie maakt het mogelijk om enerzijds meerdere associaties tussen twee entiteiten te modelleren. Daarnaast worden de entiteiten en de associaties veelal verrijkt met extra notaties zoals associatietypen en de cardinaliteiten van een koppeling. Voorbeeld van een graaf modelleerwijze is het UML klasse diagram.
- **Predicaten:** een op taal gebaseerde modelleerwijze waarbij veelal op basis van drie aan elkaar gerelateerde entiteiten de modellen opbouwen. Bijvoorbeeld aan het predicaat “Student volgt Cursus”. Voorbeeld van predicaatnotaties is NIAM.

Modelleervormen

In dit hoofdstuk hebben we een introductie gegeven van datamodellering en een aantal raamwerken geschetst op basis waarvan de verschillende datamodelleerwijzen in verband gebracht kunnen worden met elkaar.

Begrippenboom

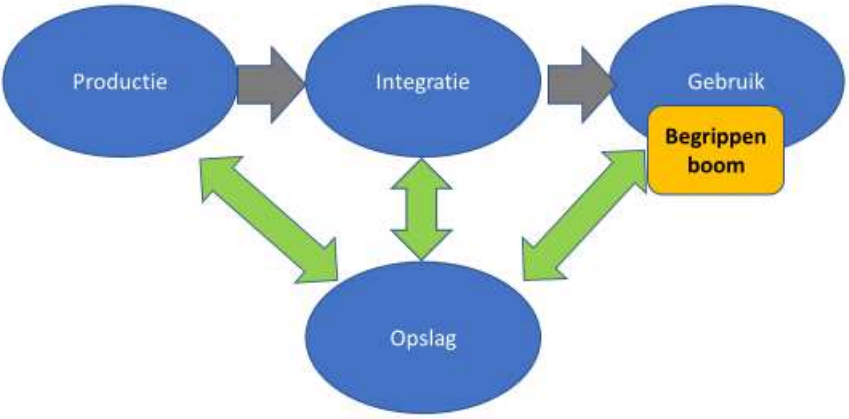
Doel

De begrippenboom heeft tot doel om op semantisch niveau de data- of bedrijfsconcepten binnen een domein in kaart te brengen. 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 de begrippenlijst opgesteld dan kan vervolgens een begrippenboom opgesteld worden waarbij verschillende hiërarchieën en verbanden tussen de begrippen gemodelleerd kunnen worden.
Het doel van een begrippenboom 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 begrippenboom en een begrippenlijst.

Raamwerken

Datalevensloop

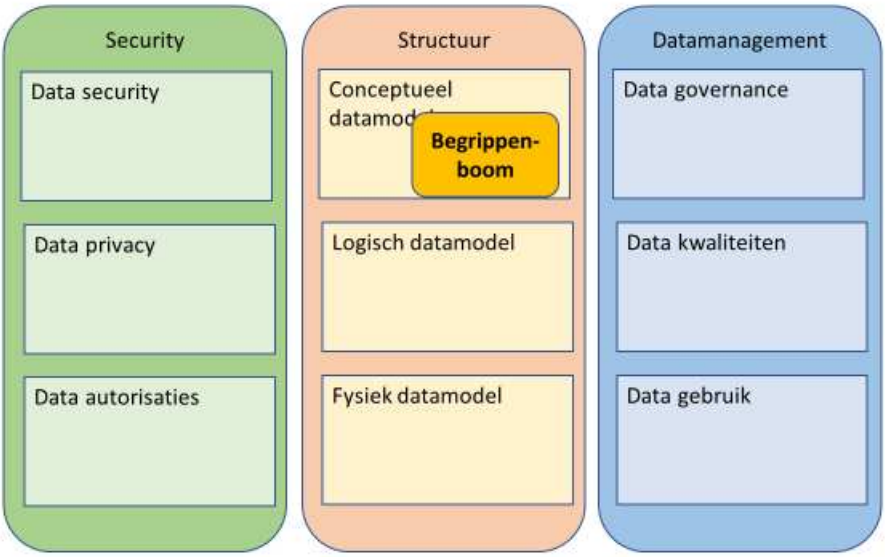
In de datalevensloop wordt de begrippenboom met name gebruikt binnen het datagebruikperspectief. Hier dient overeenstemming te zijn over de begrippen die bij het gebruik (en met name bij het nemen van besluiten op basis van de gebruikte datasets) relevant zijn. Is er geen overeenstemming over de begrippen dan zal op zeker moment in het meest positieve geval discussie ontstaan rond de gegevens die gebruikt worden om data om te zetten naar kennis. In het meest negatieve geval worden de begrippen verkeerd geïnterpreteerd en ontstaat de kans dat er verkeerde beslissingen genomen worden. Stel bijvoorbeeld dat in een ziekenhuis het begrip patiënt en de definitie niet duidelijk zijn, de gevolgen kunnen dan verstrekkend zijn.



In de afbeelding is te zien dat de begrippenboom slechts in een klein deel, met name datagebruik, een rol speelt. Afgeleid kan data een rol spelen in productie omdat dit regelmatig beïnvloed wordt door het datagebruik. Dit komt doordat het gebruik eisen stelt aan de data, maar in deze situatie wordt dit niet nader uitgewerkt.

Dataraamwerk

Ook in het dataraamwerk wordt de begrippenboom beperkt ingezet, alleen in de conceptuele modellering. Veelal wordt eerst begonnen met het opstellen van een begrippenboom als conceptueel model, bijvoorbeeld bij een top down aanpak.



In een aantal gevallen wordt de begrippenboom breder ingezet, bijvoorbeeld als glossary voor projectdocumentatie. In dat geval zal de begrippenlijst meer begrippen bevatten dan vanuit datamodelleringsperspectief noodzakelijk is. Echter, ook in dat geval blijven de begrippen op conceptueel niveau uitgewerkt.

Stakeholders

Bij het opstellen van een begrippenboom is vertegenwoordiging vanuit de business noodzakelijk. Data-architecten kunnen ondersteunen bij het opstellen van een begrippenboom maar zullen nooit in staat zijn om zelf een adequate definitie op te stellen. Onderstaande opsomming geeft de belangrijkste stakeholders:

- **Domeinexperts:** de belangrijkste stakeholder, zij dienen aan te geven welke begrippen relevant zijn en met elkaar in discussie gaan om tot een geaccepteerde definitie van een begrip en de onderlinge verhouding van de begrippen te komen.
- **Gebruikers:** in kleinere domeinen, bijvoorbeeld waarbij rapportages en business intelligence ingezet wordt kunnen

gebruikers van bijvoorbeeld rapportages ingezet worden voor het verfijnen en uitbouwen van een begrippenboom.

- **Data-analisten en scientists:** een relatief nieuwe groep stakeholders, op basis van datasets gaan deze stakeholders analyses doen met datasets en de gebruikte begrippen. Zij bouwen hierdoor diepgaande kennis op van de begrippenboom en de achterliggende objectmodellen.
- **Functioneel (Applicatie) Beheerders:** een groep stakeholders die diepgaande kennis heeft van (standaard) informatiesystemen en het onderliggende datamodel, inclusief de door de leverancier gebruikte begrippen.
- **Data- of Informatiearchitect:** zal doorgaans beperkt kennis hebben van de inhoudelijke definities van de in het domein gebruikte begrippen. Echter, deze stakeholders hebben veelal wel diepgaande kennis rond data patronen, diagnostische technieken of het opstellen van modellen rond begrippenbomen.

Concepten

Binnen de begrippenlijst en de begrippenboom staan twee zaken centraal, namelijk:

- **Begrip:** de naam zoals dit binnen de domeinpopulatie gebruikt wordt.
- **Definitie:** de 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.

Naast deze basisconcepten worden, zeker bij omvangrijke domeinen, een aantal extra concepten toegevoegd, zoals:

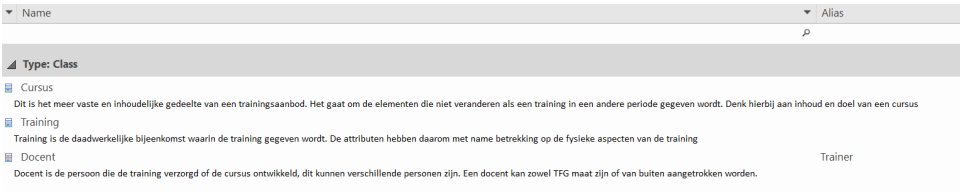
- **Synoniemen of aliassen:** begrippen met dezelfde betekenis binnen het domein maar met een ander begrip.
- **Generalisaties of Broader terms:** worden met name binnen begrippenboom veel toegepast. Hiermee wordt hiërarchie aangebracht in de begrippen. Bijvoorbeeld vervoermiddel –

fiets – racefiets is een hiërarchie waarbij vervoermiddel een generalisatie is van een fiets.

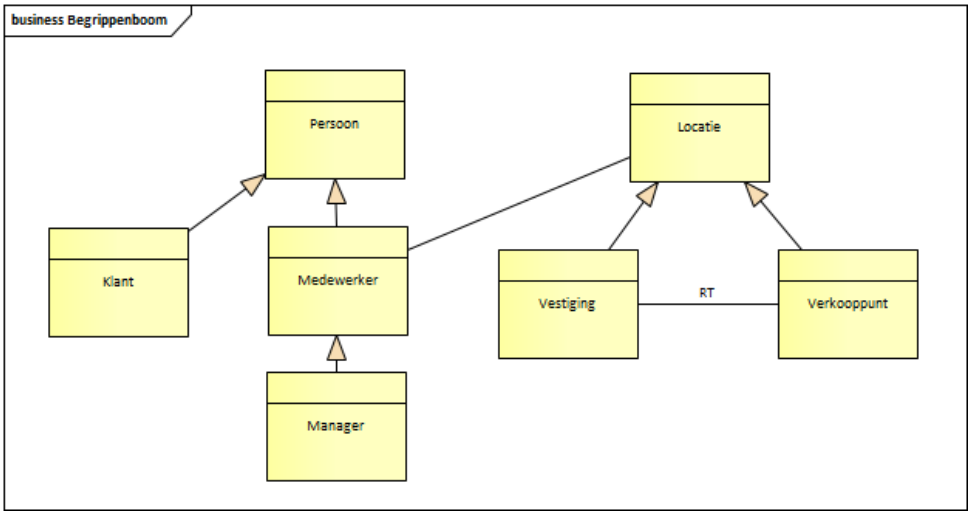
- **Specialisatie of Narrower term:** vergelijkbaar met een generalisatie alleen dan in de richting van de specialisatie. Dus fiets is een specialisatie van vervoermiddel.
- **Relatie of related term:** een associatie tussen twee begrippen die aan elkaar gerelateerd zijn. Bijvoorbeeld doordat zij hetzelfde begrip als generalisatie hebben. Bijvoorbeeld Mountainbike is gerelateerd aan Racefiets.

Notatie

Vooral de notatie van de begrippenlijst is relatief eenvoudig. Het is veelal een overzicht (op alfabetische volgorde) van de begrippen en hun definitie. Onderstaande afbeelding toont een voorbeeld hiervan.



In het voorbeeld zie je de begrippen met daaronder kort de definitie die bij het begrip hoort. Er is bij een begrip een alias, een docent is in dit domein hetzelfde als een trainer.
Voor begrippenbomen kunnen bomen als modelleerwijze gekozen worden. Ook is het mogelijk om gebruik te maken van een eenvoudig grafen model, waarbij de begrippen met gekwalificeerde associaties gekoppeld zijn. In onderstaand voorbeeld zie je een voorbeeld hiervan. De associatie met een driehoekje is een generalisatie – specialisatie associatie, de gerelateerde termen hebben geen extra symbool op de associatie maar de rolnaam RT.



Kenmerken

De begrippenlijst en de begrippenboom hebben 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 abstractieniveaus, 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.

Gebruikstoepassingen

Begrippenbomen worden vooral toegepast op een hoog abstractieniveau van datamodellering. Een begrippenboom 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.
Houd er rekening mee dat ondanks de eenvoud van de notatiewijze het opstellen van een begrippenboom een complex traject kan zijn, zeker bij een complex domein of binnen een organisatiecontext waar rond datamodellering weinig volwassenheid is.

Gerelateerde notatiewijzen

Gezien het hoge abstractieniveau van de begrippenlijst en begrippenboom zijn er weinig modelleerwijzen die overeenkomsten hebben. Binnen ArchiMate kan met behulp van business objecten een begrippenboom opgesteld worden, deze objecten worden vervolgens als koppelpunt naar bijvoorbeeld bedrijfsprocessen of -functies ingezet. Dit wordt nader uitgewerkt binnen de ArchiMate datamodellering hoofdstukken.
Ook wordt de begrippenboom veelvuldig gecombineerd met een UML klasse diagram. Hiermee ontstaat een hybride datamodel waarbij de begrippenboom als startpunt dient op een hoog abstractieniveau en de UML klassen de detaillering van deze begrippen uitwerken in de klassen, attributen en associaties.

Tooling

Voor de begrippenboom en lijst zijn verschillende tools aanwezig:

- MS-Visio
- Powerpoint
- Excel en Word (voor de boomnotatie)
- Sparx Enterprise Architect
- PowerDesigner
- ERWin
- BizzDesign
- ADO-IT
- ARIS

Evaluatie

De begrippenlijst en begrippenboom zijn krachtige notatiewijzen voor het op hoog abstractieniveau in kaart brengen van begrippen en hun onderlinge relaties. De notatie is eenvoudig en kan daardoor

ingezet worden bij stakeholders met weinig modelleerervaring en weinig affiniteit met ICT.
De notatiewijze is eenvoudig, maar het opstellen van een begrippenboom kan een langdurig en complex traject zijn als het domein complex is.
Gezien de eenvoud van de notatie is er voldoende tooling aanwezig, van eenvoudige kantoorautomatisering tot geavanceerde architectuur tools.

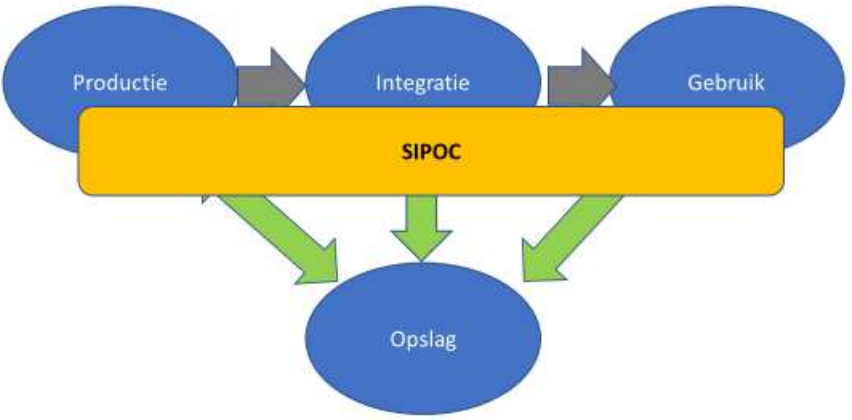
SIPOC

Doel

SIPOC is een afkorting voor Suppliers, Input, Process, Output en Consumer. Het model geeft aan wat de inkomende en uitgaande (informatie)producten zijn voor een proces of processtap. Het biedt hiermee op een hoog abstractieniveau een goed overzicht van wat de input en output is van een proces.
SIPOC is afkomstig van de Lean Six Sigma methode en is eenvoudig in te zetten bij niet-ICT stakeholders. Daarnaast is het eenvoudig in te zetten in interactieve workshops omdat het naast eenvoudige tooling ook gebruikt kan worden in situaties waar alleen pen/papier c.q. stift/whiteboard aanwezig zijn.

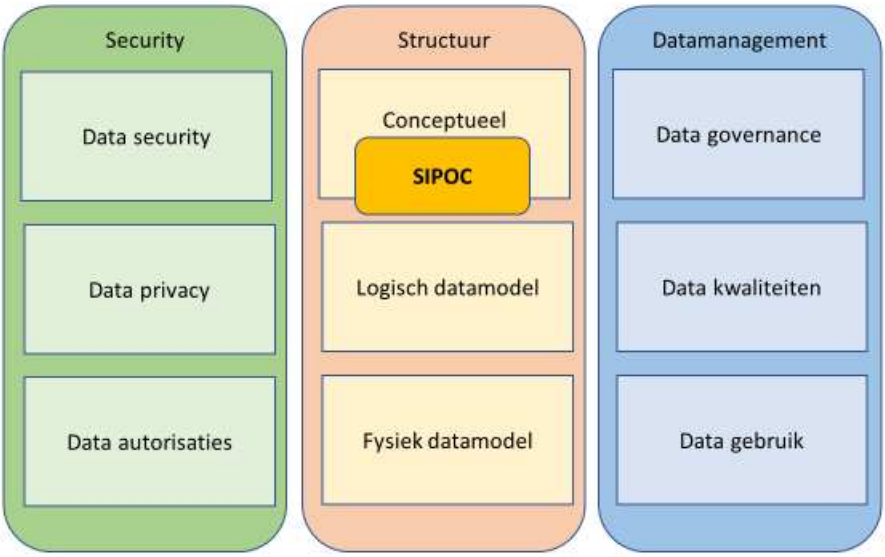
Raamwerken

Datalevensloop



In de datalevensloop wordt SIPOC afgebeeld op de gehele levensloop met uitzondering van de data-opslag. SIPOC laat zien wat de input is die door suppliers wordt geproduceerd en laat vervolgens de output zien van het proces en de bijbehorende consumers. Het model heeft daarmee een aantal details toegevoegd aan de datalevensloop stappen.

Dataraamwerk



In het dataraamwerk wordt SIPOC afgebeeld op het conceptuele datamodel. SIPOC heeft een hoog abstractieniveau qua data modellering en details rond de data-entiteiten worden meestal niet binnen de SIPOC uitgewerkt. Als details relevant zijn dan wordt veelal verwezen naar andere diagrammen en notaties met meer details (bijvoorbeeld UML Klasse diagrammen).
Indien wenselijk kan SIPOC afgebeeld worden op het datagebruik, omdat het beschrijft wat de inkomende informatie is. Dit is feitelijk een vorm van data gebruik.

Stakeholders

Omdat SIPOC heel open is in het gebruik van concepten is het op veel manieren inzetbaar bij het uitwerken van een data-architectuur. Veelal wordt SIPOC echter ingezet in interactieve workshops met business vertegenwoordigers. Stakeholders in deze zijn dan:

- Uitvoerders in werkprocessen.
- Deelnemers in interactieve workshops rond proces evaluaties.
- Teamleden van afdelingen etc.
- Proces experts.

SIPOC is ook te gebruiken op applicatieniveau waarbij de suppliers en consumers geen bedrijfsrollen of -functies zijn maar logische of fysieke applicaties. In die situatie zullen de stakeholders anders in te delen zijn en vervullen zij rollen als gebruiker of functioneel/technisch applicatiebeheerder van de te evalueren informatiesystemen.

Concepten

Belangrijkste concepten zitten al in de afkorting verborgen:

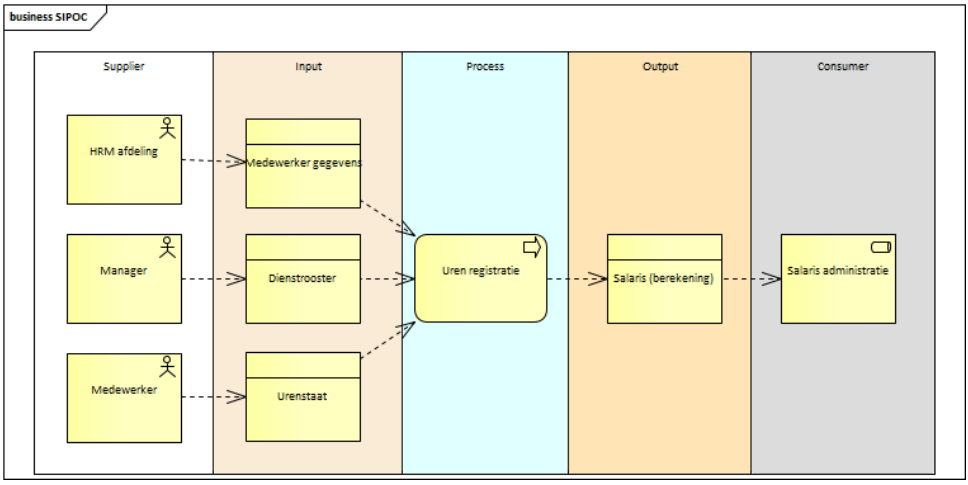
- **Supplier (leverancier):** de rol, persoon of functie die de informatieproducten aanlevert als input voor het proces dat geanalyseerd wordt.
- **Input:** informatie of data die van de leverancier naar het proces stroomt. Let op: input hoeft niet altijd een informatieproduct te zijn. Het kunnen ook fysieke producten zijn zoals grondstoffen of halffabricaten.
- **Proces:** het proces dat in deze SIPOC analyse geïnventariseerd wordt en dat zorgdraagt voor de transformatie van de input naar de output.
- **Output**: de uitkomst van de transformatie in het proces in de vorm van informatieproducten. Ook hierbij kan de output fysieke producten zoals halffabricaten en eindproducten of een combinatie van fysieke en informatieproducten zijn. Denk hierbij aan een eindproduct in combinatie met een pakbon.
- **Consumer (consument):** afnemer(s) van de output. Dit zijn veelal bedrijfsrollen en functies maar kunnen ook een vervolgstap in de procesketen zijn.

Houd er rekening mee dat de scope van een proces maar ook van de SIPOC op verschillende niveaus uit te werken zijn, zoals:

- **Bedrijfsprocessen** binnen een organisatie of afdeling (zie hierboven).
- **Ketenprocessen** over meerdere organisaties heen. In dit geval kunnen meerdere SIPOC ketens achter elkaar geplaatst worden.
- **Applicatiestromen**: stromen van data tussen logische en/of fysieke informatiesystemen.

Notatie

De notatie is relatief eenvoudig en bestaat feitelijk alleen uit de vijf concepten naast of onder elkaar geplaatst. Dit kan eenvoudig op papier of een whiteboard gedaan worden. Bij de inzet van ICT hulpmiddelen wordt veelal gebruik gemaakt van het concept swimming lanes.
Onderstaande afbeelding toont een eenvoudig SIPOC diagram afkomstig uit een ICT hulpmiddel.



In het voorbeeld is te zien hoe de concepten in verticale swimminglanes worden geplaatst. In dit geval wordt een bedrijfsproces binnen een organisatie afgebeeld. In het voorbeeld is de SIPOC opgebouwd uit ArchiMate bedrijfsconcepten. Dit is een hybride aanpak die als voordeel heeft dat de definities van de ArchiMate typen en associaties ook binnen de SIPOC gelden.

Naast de concepten zijn in dit SIPOC diagram ook de associaties tussen de verschillende concepten opgenomen. Zo zien we welke suppliers welke input leveren en welke consumers welke output ontvangen. Een extra detaillering die veelal meerwaarde heeft. Echter, deze toevoeging is niet noodzakelijk en kan ook worden weggelaten of met behulp van een andere notatiewijze worden uitgewerkt.

Kenmerken

SIPOC kenmerkt zich door de eenvoud in het gebruik. In onderstaande opsomming de kenmerken (zowel positief als negatief):

- Zeer eenvoudige- en niet gestandaardiseerde notatie (enkel de SIPOC stappen zijn gedefinieerd).
- Eenvoudig in te zetten bij betrokkenen met weinig affiniteit met modelleren en (enterprise) architectuur.
- Aanpasbaarheid van concepten. Meestal gebruikt om bedrijfsprocessen en de relaties te modelleren maar ook inzetbaar voor informatiesystemen en datastromen tussen applicaties.
- Uitbreidbaar met associaties tussen de SIPOC onderdelen waarmee een detaillering toegevoegd wordt waarmee informatie- en goederenstromen tussen concrete leveranciers en ontvangers van het proces worden verklaard.
- Modellen kunnen zeker bij omvangrijke processen complex worden. Dit kan ondervangen worden door een SIPOC op te splitsen in meerdere diagrammen, echter, de complexiteit wordt hierdoor maar beperkt verminderd.
- Eenvoudig te combineren met bijvoorbeeld ArchiMate voor het definiëren van de gebruikte concepten of met bijvoorbeeld UML voor de details van de input en/of output.

Gebruikstoepassingen

Vanwege de eenvoud van de methode en de notatie is SIPOC in veel situaties toe te passen. In voorgaande paragrafen is hier reeds op ingegaan. Meest kenmerkende gebruikstoepassing is de inzet bij interactieve workshops met werkgroepen zonder modelleerervaring.