

# **Tien tips voor canonieke datamodellering**

**Bert Dingemans**

## Abstract

---

Modelleren is een vakgebied gebaseerd op eenvoudige notaties. Echter op het moment dat en model opgesteld wordt blijkt de te modelleren werkelijkheid al snel complex. In dit whitepaper worden daarom een aantal handvatten geboden die ingezet kunnen worden om de complexiteit te verminderen.

De meeste van de tien tips zijn gericht op het reduceren van de modelleercomplexiteit door het opsplitsen van het totaal in kleinere gerubriceerde deelactiviteiten die minder complex zijn.

## Inhoudsopgave

---

Abstract .....	2
Inhoudsopgave .....	2
Inleiding .....	4
Werk iteratief .....	4
Volg de projectagenda van de organisatie .....	5
Analyseer aanwezige informatie in een vroeg stadium .....	6
Documenteer de gemaakte keuzen.....	6
Werk interactief .....	7
Hou rekening met doel en doelgroep van het model .....	8
Valideer je model regelmatig bij je stakeholders.....	8
Bepaal en beschrijf de notatiewijze .....	9
Werk met primaire en secundaire modellen .....	9
Zoek en hergebruik bestaande patronen .....	10
Tot slot .....	10
Literatuur.....	10



# Inleiding

---

In de voorgaande whitepapers zijn een aantal handvatten gegeven hoe een canonieke data architectuur gemodelleerd kan worden. Bijvoorbeeld op het vlak van notatiewijzen en op het vlak van stakeholder betrokkenheid.

Je wilt daarmee aan de slag om een datamodel op te starten. Echter waar begin je en hoe pak je je werkzaamheden aan om tot een goed model te komen. In dit whitepaper krijg je daarom een tiental tips waarmee je aan de slag kunt met het opstellen van een canoniek datamodel. Hierbij is onderstaande drieindeling toegepast

- Werkwijze
  - Werk iteratief
  - Volg de projectagenda van de organisatie
  - Analyseer aanwezige informatie in een vroeg stadium
  - Documenteer de gemaakte keuzen
- Gebruikersparticipatie
  - Werk interactief
  - Hou rekening met doel en doelgroep van het model
  - Valideer je model regelmatig bij je stakeholders
- Modelleren
  - Bepaal en beschrijf de notatiewijze
  - Werk met primaire- en secundaire modellen
  - Zoek en hergebruik bestaande patronen

De tien tips worden in de onderstaande hoofdstukken uitgewerkt.

## Werk iteratief

---

Het opstellen van een canoniek datamodel is een complexe activiteit. Het is daarom vrijwel onmogelijk om direct een compleet model op te stellen. Reden om te zoeken naar een werkwijze waarbij er gezocht wordt naar deelactiviteiten die na elkaar kunnen worden uitgevoerd.

Daar komt bij dat de betrokken stakeholders in het begin van het modelleertraject zich moeilijk een beeld kunnen vormen van het gewenste uiteindelijke resultaat. Door met iteraties te werken worden tussenproducten in een vroeg stadium ter validatie voorgelegd aan de stakeholders. Hiermee biedt iteratief werken de stakeholders in een vroeg stadium de mogelijkheid om het model bij te sturen.

Binnen de softwareontwikkeling is momenteel een trend gaande waarbij agile werkwijzen worden toegepast zoals Scrum of DSDM. Deze methoden kunnen ook voor het opstellen van een datamodel goed ingezet worden. Een datamodel is feitelijk ook een ontwikkelproces, echter niet van software maar van een model

Iteratief modelleren kan op meerdere wijzen worden uitgewerkt. In onderstaande opsomming een aantal werkwijzen:

- **Dakpansgewijs modelleren**, bij deze werkwijze wordt een deelmodel geheel uitgewerkt en gevalideerd bij de betrokken stakeholders. Op het moment dat er een akkoord is over

het deelmodel wordt een aanpalend model in een vervolgiteratie uitgewerkt en gevalideerd

- **Ketengewijs modelleren**, hierbij is het kenmerkend dat een model wordt opgesteld, vervolgens wordt het aan de stakeholders ter goedkeuring voorgelegd, is men akkoord dan wordt gestart met het volgende model, is men niet akkoord dan wordt slechts eenmaal een aanpassing gedaan. Is men daarna nog niet akkoord dan wordt het model geparkeerd tot een laatste iteratie waarin alle “discussiemodellen” besproken worden. Deze werkwijze werkt goed bij een traject onder tijdsdruk.
- **Top down modelleren**, in situaties waarbij men eerste een overzicht wil hebben van het domein is dit goed toepasbaar. Er wordt een model opgesteld van alleen de (hoofd)entiteiten en hun onderlinge relaties. In latere iteraties wordt detaillering toegevoegd op het vlak van attributen, relatie- en entiteitsoorten.

Naast deze drie vormen zijn er vele andere iteratieve werkwijzen. Houd bij het kiezen van een werkwijze altijd rekening met een aantal aspecten zoals:

- **Stakeholder**, denk hierbij aan het kennisniveau, het aantal en de diversiteit van de verschillende stakeholders
- **Politiek**, rond het model dient besluitvorming plaats te vinden. Dat kan als de belangen groot zijn bij de stakeholders lastig zijn. Is dit het geval zorg dan dat het besluitvormingsproces voor iedereen vooraf duidelijk is.
- **Tijdsaspect**, is de tijdsdruk hoog dan zal een andere werkwijze gevolgd moeten worden dan in situaties waarbij tijdsdruk geen rol speelt
- **Doel van het model**, waar wordt het model voor ingezet, is dit voor een softwareontwikkeltraject, een ketenintegratie of een besluitvormingsproces

Het kan interessant zijn, zeker als je regelmatig een (deel)model dient uit te werken een eenvoudig register op te stellen van toe te passen werkvormen voor iteratief en interactief werken. Deze twee vormen kunnen namelijk erg goed gecombineerd worden.

## Volg de projectagenda van de organisatie

---

Het opstellen van een canoniek datamodel dient geborgd te zijn in de organisatie. Het ontwikkelen van een dergelijk model zonder betrokkenheid van bijvoorbeeld ambassadeurs is kansloos. Je krijgt geen commitment, men heeft geen tijd of betrokkenheid bij de activiteiten. Echter een project starten specifiek voor het opstellen van een canoniek datamodel is alleen mogelijk als door de stakeholders het belang van een dergelijk model gezien wordt. Dat zal alleen mogelijk zijn als je stakeholders een goed beeld hebben van het belang van data, ICT en informatievoorziening. Andere mogelijkheid is dat je als data-architect voldoende overtuigingskracht hebt om de stakeholders enthousiast te krijgen voor je canonieke datamodel. Beide scenario's kom ik in mijn dagelijkse praktijk niet veel tegen. Een werkwijze waarbij aangesloten wordt op de projectagenda zal in veel situaties goed werken. Projecten sluiten meestal aan bij knelpunten in de organisatie en hebben daardoor een zekere urgentie. Dat betekent dat er daardoor focus is bij een aantal betrokkenen. Niet onbelangrijk projecten hebben veelal budget. Bijkomend voordeel is dat de producten van het datamodel direct ingezet kunnen worden voor de producten die binnen het project ontwikkeld worden.

Bijvoorbeeld een datamodel dat in een ketenintegratie ingezet wordt voor de uit te werken koppelvlakspecificaties.

Nadeel van deze werkwijze is dat het uitwerken van het datamodel minder efficiënt is. Doordat projecten elkaar op het vlak van een datamodel kunnen overlappen zal het model door nieuwe inzichten vanuit de projectuitwerking bijgesteld, uitgebreid of herzien moeten worden. Hierbij is het mogelijk dat de aanpassingen invloed hebben op modellen van projecten die reeds afgerond zijn. Gevolg is dat er wijzigingsverzoeken gedaan moeten worden die impact kunnen hebben op de informatievoorziening van de organisatie.

## Analyseer aanwezige informatie in een vroeg stadium

---

Rond het canonieke datamodel is in de organisatie bijna altijd al heel veel informatie aanwezig. Om effectief en efficiënt te werken en de stakeholders te ontlasten loont het om voorafgaand aan modellersessies een concept model op te stellen op basis van de aanwezige informatie.

Informatie is te vinden in de vorm van:

- **Procesbeschrijvingen**, is binnen de organisatie het procesmatig werken ingebed en zijn de processen beschreven dan is hier goede informatie te vinden over het datamodel. Processen veranderen tenslotte de status van gegevensentiteiten
- **Formulier- en documentsjablonen**, formulieren en sjablonen van uitvoerdocumenten geven al een goed beeld van de informatiebehoefte en kunnen daarom gebruikt worden voor het concept datamodel
- **Open standaarden** zijn op het vlak van inhoud gebaseerd op een informatiemodel dat veelal al door een aantal betrokken partijen is gevalideerd en goedgekeurd.
- **Berichtenboeken en koppelvlakspecificaties**, zijn er in het domein reeds koppelvlakken beschikbaar en zijn deze beschreven met behulp van een berichtenboek dan is dit een goede informatiebron voor het concept datamodel.

Zorg daarom als data-architect dat je een aantal stappen voorloopt op het feitelijke ontwerptraject voor wat betreft de analyse van de aanwezige informatie. Eventueel kan je op basis hiervan voor de bijeenkomsten ook al een aantal zaken voorbereiden zodat de bijeenkomst direct aan de slag kan met het aanwezige concept. Let er hierbij wel op dat betrokkenen voldoende kritisch zijn op de aangeleverde modellen en dit niet als geldend gaan zien

## Documenteer de gemaakte keuzen

---

Het lange termijngeheugen van organisaties is veelal minder ontwikkeld dan het korte termijngeheugen. Gevolg is dat de reden waarom bepaalde keuzen gemaakt zijn verdwijnen naar de achtergrond. Bij de stakeholders van een dataontwerp traject is daardoor niet meer bekend waarom bepaalde onderdelen wel of niet uitgewerkt worden en waarom een bepaalde werkwijze wordt toegepast.

Draag tenminste zorg voor documentatie van de volgende keuzen:

- **Toepassen werkwijze**, beschrijf welke werkwijze toegepast wordt, bijvoorbeeld of iteratief ontwikkeld wordt of een methode als scrum toegepast wordt en wie bepaald wat

de volgorde van de deelmodellen is. Als laatste is het van belang om de inspraak te beschrijven.

- **Notatiewijze**, beschrijf op welke wijze het datamodel gemodelleerd wordt, welke architectuurmethode gebruikt wordt en welke delen hiervan gebruikt worden
- **Producten**, beschrijf de op te leveren producten qua inhoud maar ook qua vorm.
- **Inzet tooling**, wordt er tooling ingezet, zo welke, wordt aangesloten op de aanwezige standaarden. Moet een tool aangeschaft of ontwikkeld worden geef dan aan wat de acceptatiecriteria zijn.
- **Verslaglegging van bijeenkomsten**, is wat mij betreft een optionele uitwerking maar kan zeker bij een grote groep stakeholders of bij politiek gevoelige trajecten meerwaarde bieden.

Kies bij het uitwerken van de documentatie wel voor een lichtgewicht vorm. Zelf heb ik goede ervaring met het inzetten van powerpoint als documenteerwijze. Dat nodigt uit om de besluitvorming kort en krachtig te formuleren en biedt de mogelijkheid om dia's eenvoudig te herschikken.

## Werk interactief

---

Bij het uitwerken van een datamodel is interactie met de stakeholders van groot belang. Enerzijds om input te vragen over het model, anderzijds om het model continu te kunnen valideren. Daarnaast is het datamodel in een aantal gevallen de start van een veranderingstraject. In een vroeg stadium betrokkenheid krijgen rond het model door een interactieve werkwijze kan bijdragen aan de acceptatie van de aanstaande veranderingen.

Interactief werken kan op vele manieren, hieronder een aantal voorbeelden van interactieve vormen:

- **Interactieve workshops**, bij deze werkwijze wordt in een bijeenkomst met een aantal stakeholders het model uitgewerkt of gevalideerd. Afhankelijk van de groepsgrootte en de eventuele tegenstrijdigheden in belangen van betrokkenen moet gezocht worden naar een geschikte werkvorm. Om workshops efficiënt te laten verlopen is het van belang om aan de voorbereiding voldoende tijd te besteden. Zie bijvoorbeeld de tip **Analyseer aanwezige informatie in een vroeg stadium**
- **Één op één sessies**, in situaties waarbij een deelmodel in de scope van één stakeholder ligt of in situaties waar samenwerking tussen de stakeholders gevoelig ligt kunnen één op één sessies een goede oplossing zijn. Voordeel van deze werkwijze is dat het efficiënt is. Belangrijk blijft wel om het gehele model bij alle betrokkenen te valideren. Dit kan ook weer één op één of in een hybride vorm
- **Pair modelling**, in situaties waar tooling ingezet wordt voor het opstellen van datamodellen kan gezamenlijk met een aantal stakeholders een model uitgewerkt worden. Hierbij wordt het resultaat direct getoond via een beamer oid en de betrokkenen kunnen direct input leveren op het resultaat. Deze werkvorm werkt goed als de stakeholders de gebruikte notatiewijze goed kunnen lezen en als de groep niet te groot is
- **Combinaties of hybride vormen**, vaak is een combinatie van werkvormen bijvoorbeeld één op één en workshops een prettige werkwijze. Eerst wordt één op één

een conceptversie van het model uitgewerkt. Vervolgens wordt het model in een groepsessie gevalideerd. Interessant is de werkvorm waarbij de validatiesessie naast de deelnemers van de één op één sessies er een aantal deelnemers zijn die geen bijdrage hebben geleverd aan het te valideren concept.

Voor de interactieve workshops en modelleersessies kunnen allerlei creatieve werkvormen ingezet worden. Kijk eens op [www.werkvormen.info](http://www.werkvormen.info) voor een overzicht..

## Hou rekening met doel en doelgroep van het model

---

Bij het uitwerken van een datamodel is het goed om voorafgaand aan het modelleren na te denken over wat je met het model wilt bereiken en wie het model gaat gebruiken. Een model ter ondersteuning van besluitvorming door een managementteam dan een datamodel waarmee je een softwareontwikkelaar duidelijkheid wilt geven over de te ontwikkelen software. Modelleren ter overtuiging van betrokkenen zal veelal een vereenvoudiging inhouden van het model. Dat kan zelfs zover gaan dat de weergave meer een “artist impression” is dan een feitelijke weergave van het datamodel. Mijn ervaring in deze is dat eerst het gedetailleerde model opstellen en vervolgens het vereenvoudigde model eenvoudiger realiseerbaar is dan omgekeerd. Heb je te maken met heel verschillende doelen of met heel verschillende stakeholders schroom dan niet om meerdere modellen op te stellen.

## Valideer je model regelmatig bij je stakeholders

---

Zowel bij de hoofdstukken **werk interactief** en **werk iteratief** is reeds aandacht besteed aan dit onderwerp. Modelvalidatie is vanwege een aantal redenen van groot belang:

- **(Bij)sturing van de modellering**, bij het uitwerken van een datamodel zonder validatie loop je de kans dat de uitwerking onvoldoende aansluit bij de perceptie van de stakeholders. Hierdoor kunnen in een later stadium grote problemen ontstaan. Bijvoorbeeld wanneer integratieoplossingen of applicaties ontwikkeld worden. Als deze producten op een verkeerd of onvolledig datamodel gebaseerd zijn is het doen van correcties een kostbare activiteit
- **Betrokkenheid van de stakeholders**, is al ruimschoots aan de orde gekomen in voorgaande hoofdstukken. Samenvattend betrokken stakeholders zijn goede ambassadeurs van je datamodel.
- **Homoniemen en synoniemenprobleem**, zeker bij modellering op conceptueel niveau is het risico van homoniemen en synoniemen altijd aanwezig. Door regelmatig te valideren en daarbij tijdens de validatie aandacht te besteden aan de eventuele homoniemen en synoniemen kan dit probleem gereduceerd worden.

Validatie kan gedaan worden door een werkwijze zoals beschreven in **werk interactief**. Andere vormen zijn eveneens goed toepasbaar. Denk bijvoorbeeld aan het beschikbaar stellen van het model met behulp van een wiki of een modelleertool. Veel modelleertools zijn in staat om statische websites, word- en pdf documenten te genereren of de repository real time te ontsluiten voor gebruikers. Door deze werkwijze kun je enerzijds je model valideren. Anderzijds kunnen deze producten later gebruikt worden om anderen te informeren over het gepubliceerd model.



## Bepaal en beschrijf de notatiewijze

---

Voor het opstellen van een datamodel zijn meerdere notatiewijzen beschikbaar. Hierover is reeds een whitepaper opgesteld zie [1]. Onder andere kunnen de volgende modelleertechnieken worden ingezet:

- **ER diagrammen**, modelleren van entiteiten, attributen en relaties (inclusief de cardinaliteiten). Notatiewijze is traditioneel gerelateerd aan database ontwikkeling en kan daardoor voor niet technische stakeholders lastig te lezen zijn
- **UML klassediagrammen**, modelleren van objecten, attributen en hun relaties (inclusief de cardinaliteiten). Notatiewijze is sterk gerelateerd aan softwareontwikkeling en kan eenvoudig uitgebreid worden met aan software gerelateerde notatiewijzen binnen de UML modelleertaal.
- **Archimate**, is een modelleertaal voor enterprise architectuur en kent daarbinnen een belangrijke data- en informatiemodellering op basis van objecten. Kan objecten en relaties modelleren (zonder de cardinaliteiten) Krachtig is dat het hiermee mogelijk is om de modellen uit te breiden met andere concepten zoals componenten, werkprocessen, actoren etc voor het completeren van een enterprise architectuur model.

Welke modelleerwijze je ook kiest het is belangrijk om de gemaakte keuzes te documenteren. Denk daarbij ook aan het beschrijven waarom bepaalde elementen wel of niet beschreven worden en als ze beschreven worden op welke wijze dit gedaan dient te worden in de modellen. Door deze beschrijving ontstaat duidelijkheid over hoe een model (in een bepaalde diagramtechniek) gelezen dient te worden. Helaas bestaat niet alleen in natuurlijke taal een homoniemen en synoniemen probleem.

In een aantal van bovengenoemde modelleertechnieken wordt gebruik gemaakt van zogenaamde viewpoints. Dit is goed toepasbaar als hulpmiddel bij het beschrijven van de notatiewijze. Soms zijn er vooraf gedefinieerde viewpoints maar het staat je natuurlijk vrij om deze zelf op te stellen en te beschrijven.

Is de beschrijven van voldoende kwaliteit dan is het mogelijk om de beschrijving in te zetten als instructiemateriaal rond het lezen van de gebruikte datamodellen. Zeker in situaties waarbij er veel stakeholders betrokken zijn bij de modellen of waarbij niet iedere stakeholder rechtstreeks en persoonlijk benaderd kan worden voor een instructie.

## Werk met primaire en secundaire modellen

---

In het hoofdstuk **Houd rekening met doel en doelgroep van je model** is aangegeven dat er meerdere modellen ingezet kunnen worden bij verschillende doelen of doelgroepen. Bij het uitwerken van de modelleervormen kan een getrapte uitwerking gebruikt worden. Primaire modellen is de minimale set van notatiewijzen waarmee het datamodel wordt opgesteld. Secundaire en eventueel zelfs tertiaire modellen worden ingezet voor aanvullende notatie in situaties waarin deze toegevoegde waarde hebben. Veelal zullen deze secundaire modellen ingezet worden voor een nadere verklaring bij complexe modellen of als extra illustratie voor specifieke doelen of doelgroepen

Bij het beschrijven van de notatiewijze kan aangegeven worden of een modelleerwijze primaire of secundaire is. Beschrijf daarbij ook in welke situatie een model ingezet wordt en hoe deze zich verhoudt tot andere primaire en secundaire modellen.

## Zoek en hergebruik bestaande patronen

---

Patronen zijn bewezen oplossingen voor regelmatig terugkerende problemen binnen een bepaalde context. Patronen zijn krachtige hulpmiddelen die op vele werkerreinen worden ingezet. Denk bijvoorbeeld aan design patterns en interactiepatronen. Ook voor het opstellen van datamodellen zijn datapatronen beschikbaar die behulpzaam kunnen zijn om modelleren te vereenvoudigen en consistent te maken [2].

Ook binnen eigen modellen kunnen eenvoudige patronen je behulpzaam zijn om modellen te vereenvoudigen en te standaardiseren. Hiermee worden de modellen consistent en eenvoudiger te interpreteren door de verschillende stakeholders.

Ga je echt aan de slag met patronen dan kun je een catalogus van patronen opstellen en deze beschrijven wanneer een patroon ingezet wordt. Daarnaast kun je eventueel de notatiewijze uitbreiden met symbolen voor bepaalde datapatronen. In combinatie met **primaire en secundaire modellen** en de **beschrijving van de notatiewijze** ontstaat hiermee een krachtige dataontwerpmethode.

## Tot slot

---

Met de bovenstaande tien tips hoop ik dat je aan de slag kunt zonder modellers-block! Neem als uitgangspunt “verdeel en heers”. Hak voorbereiding en uitvoering in meerdere kleine stappen en zorg dat je na iedere stap je model valideert. Heb je vervolgens je werkwijze gevonden kijk dan naar opties om het modelleerproces te optimaliseren.

**Veel succes met modelleren!**

## Literatuur

---

[1] Dingemans, Bert, *Canonieke datamodelering in de praktijk* (2013), [www.architectuurassistent.nl](http://www.architectuurassistent.nl)

[2] Microsoft corporation, *Data Patterns (Patterns & Practices)*, z.j., **Microsoft Press**.