

Data steward heeft belangrijke sleutelrol

# Op weg naar BI-wonderland

Joost Henskens

**Informatievoorziening, Business Intelligence en datawarehousing zijn alweer decennia geleden in zwang geraakt, maar steeds vaker worstelt men met het blijvend behalen van het rendement uit deze investeringen. Dit artikel geeft een inleiding in hoe rendement kan worden behaald door de invoering van data governance, life-cycle management voor datawarehouses en de nieuwe rol van de informatiebeheerder; de data steward.**

Data governance is een opkomend fenomeen binnen de wereld van Business Intelligence en information delivery. De aandacht voor data governance komt voort uit een nieuw tijdperk (Winston Chen, 25 juni 2010). Het *applicatietijdperk* liep van circa 1960 tot circa 1990. Daarna kwam het *repository-tijdperk*, van circa 1990 tot nu, waarin centralisatie en *enterprise views* centraal staan. Op dit moment staan we aan de vooravond van het volgende tijdperk de *'policy era'*, waarbij de focus ligt op de omgang met data. Nieuwe applicaties volgen elkaar steeds sneller op en er is steeds meer behoefte aan uitwisseling van gegevens. Dit typeert de steeds belangrijker wordende rol van data in de besluitvorming van organisaties. De beschikbaarheid en kwaliteit van de corporate informatie moet de snelheid van handelen van organisaties daarom nog beter gaan bijhouden.

Vroeger was het niet vreemd om Master Data Management (MDM) als het enige probleem te zien dat moest worden opgelost om data-uitwisseling en kwaliteitsborging mogelijk te maken. Er werd altijd gesproken over het creëren van het 'gouden record'. Deze werkwijze heeft eigenlijk meer met eenmalig goed MDM-ontwerp te maken dan met een voortdurende zoektocht en ambitie naar verbetering en goede datakwaliteit. Als gevolg van het ontstaan van grote consolidaties en repository's (datawarehouses enzovoort) ontstaat een aanvullende behoefte om deze beter te stroomlijnen en te beheersen. Data governance vormt hierin een belangrijke stap.

## Data governance gedefinieerd

Bij organisaties waar data-uitwisseling plaats vindt tussen systemen en processen, wordt het tijd om een data governance programma te starten. Hiermee zijn organisaties in staat om positieve invloed uit te oefenen op de processen en methoden die worden gebruikt bij het verwerken van hun data. Een set van

processen zorgt ervoor dat belangrijke gegevens formeel worden beheerd binnen de gehele onderneming en dat deze gegevens kunnen worden vertrouwd. De scope van data governance is de hele organisatie. Als bedrijven de controle over hun gegevens willen krijgen, deelt data governance de juiste bevoegdheid uit, maakt het duidelijk welke stappen moeten worden genomen en schakelt het de hulp in van technologie om dit te bereiken. Data governance is een doorlopend proces dat continu gericht is op de wijze hoe data worden beheerd en gebruikt. Het brengt datakwaliteit, data management, Business Process Management en risicomangement rondom de verwerking van gegevens in een organisatie bij elkaar.

In mijn definitie omvat data governance de mensen, processen en informatietechnologie die benodigd zijn voor een consistente en correcte afhandeling van de gegevens binnen een organisatie, met inbegrip van de volgende doelstellingen:

- Vergroting van het vertrouwen in de besluitvorming;
- Vermindering van het risico op boetes van de regelgevende instanties (SOX, Basel II, enzovoort);
- Verbetering van de beveiliging van gegevens;
- Maximalisatie (van de mogelijkheden) van het genereren van inkomsten uit/met gegevens;
- Beleggen van verantwoordelijkheid voor de kwaliteit van informatie.

## De rol van de data steward

Bij een doorlopend proces hoort een team van mensen die dit proces een succes maken. Dit team buigt het hoofd over de nauwkeurigheid, toegankelijkheid, consistentie, volledigheid en eventuele andere eisen die aan de data moeten worden gesteld. Het team bestaat daarom meestal uit gemandateerd leiderschap, projectmanagement, afdeling/unit managers en data stewards.

De rol van een data steward binnen een data governance initiatief is *key*.

Data stewards hebben een belangrijke rol bij de definitie van de data. Zij bepalen namelijk in grote mate de waarde van de informatie. Een data steward is een persoon die verantwoordelijk is voor de opgeslagen gegevens en rekenregels van een bepaald (primair) administratief ICT-systeem. Zij dragen zorg voor een goede 'flow' van de data door het systeem en interactie met andere systemen en ze kennen de data, de logica en kunnen deze op waarde schatten in relatie tot de dagelijkse operatie en werkprocessen. Ook ondersteunen ze eindgebruikers bij het gebruik van de gegevens in hun dagelijks werk. Binnen het data governance initiatief, waarbij data, gegevens en informatie moeten worden gekoppeld aan de business processen en de ondersteuning hiervan, zijn de data stewards leidend in de relaties tussen IT en de business en dragen ze de verantwoordelijkheid voor de datadefinities, data management procesdefinities en de datakwaliteit (en bijbehorende eisen). Data stewardship gaat gepaard met het verbeteren en het beheer van de gegevens in termen van herbruikbaarheid, toegankelijkheid en kwaliteit. De stewards zorgen voor de aansluiting op de veranderende informatie vraag van organisaties en hun afdelingen.


Probleem is echter dat data stewards vaak verantwoordelijk zijn voor slechts één bepaald business segment of systeem (financieel, logistiek enzovoort). Het sturen op de bedrijfsprocessen is echter systeemoverstijgend en dus zal een Data Steward dikwijls

in conflict zijn met de informatieafnemer. Hij is tenslotte slechts verantwoordelijk voor de data uit één systeem. Steeds meer ondernemingen maken daarom onderscheid tussen technische data stewards en functionele (business) data stewards. Dit beperkt het probleem enigszins omdat de business data steward in dat geval de verantwoordelijkheid draagt voor gegevensdefinities binnen de gehele onderneming.

### Levenslange zorg

Nu de veranderingen in de informatie vraag van gebruikers door data stewards en data governance in goede banen wordt geleid moet er alleen op technisch vlak nog structureel invulling worden gegeven aan de gewenste veranderingen. Dat is mogelijk vanuit een life-cycle model. Hoe kan de veranderende vraag daarmee dan actief worden beantwoord? Simpelweg door de motor onder deze informatievoorziening (het datawarehouse) een periodieke keuring (APK) te geven.

Niet zelden werd een datawarehouse ontwikkeld om er vervolgens alleen nog maar zaken aan 'vast te plakken', zonder eens kritisch te kijken naar het geheel. Een datawarehouse dient systeemafhankelijk en tijdsloos te zijn. Een goed datawarehouse kan vele jaren mee en wordt rijker naarmate de tijd verstrijkt. Het ondersteunt de primaire en secundaire processen van een onderneming. Datawarehouse Life-cycle Management (DWLM) is de juiste oplossing voor het up-to-date houden van alles rondom het principe 'de juiste informatie op het juiste moment'. DWLM grijpt in waar Data Governance stopt en is een



# Sterren en Dimensies

*Ontwerp en onderhoud van datawarehouses*

ONGEWIJZIGDE HERDRUK

DB/M Essay Een bundeling van artikelen verschenen in Database Magazine in de periode 1998-2002

**Sterren en Dimensies**  
*Ontwerp en onderhoud van datawarehouses*

Dr. H. van der Lek  
F. Habers M. Schmitz

Herziene en uitgebreide tweede druk 2003

Sterren en Dimensies is vanwege grote belangstelling in een ongewijzigde derde druk verschenen. Het boek uit de welbekende DB/M Essay reeks bevat een bundeling van artikelen uit DB/M over het ontwerpen en onderhouden van datawarehouses. Deze artikelen zijn gepubliceerd in de periode 1998 – 2002. De experts Harm van der Lek, Frank Habers en Michael Schmitz geven principes voor het gebruik van sterschema's en laten zien hoe de 'sterren' uitblinken in eenvoud.

**Wilt u de inherente kracht van het dimensionale denken volledig benutten? Dan kunt u niet zonder dit boek!**

**Ga snel naar [www.array.nl](http://www.array.nl) en bestel Sterren en Dimensies!**

**Array** PUBLICATIONS

combinatie van processen (denk aan masterdata management en metadata management) en technische verbeteringen en aanpassingen (producten, tools en infrastructuur), en soms zelfs het gedrag van mensen die werken met informatie (besluitvormingsprocessen). Hoe pak je dit aan?

Conventionele ontwikkelmethodieken zijn gebaseerd op de watervalmethode. Toch is dit niet de juiste wijze. Frank Buytendijk, voorheen Vice President van Gartner, Inc., licht dit als volgt toe: "De meeste datawarehouseprojecten zijn alleen gefocust op de technische aspecten voor de eerste implementatie en veranderingen. Daarna worden ze als probleem voor beheer en onderhoud gezien. Dit is niet de realiteit. Enterprises moeten kiezen voor een methodologie waarin zowel business als IT samenwerken gedurende de gehele life-cycle van het datawarehouse." Dit agile denken kan ook worden toegepast op datawarehouses in combinatie met de governance problematiek.

## De data stewards zijn de gidsen door het landschap

Datawarehouse Life-cycle Management (DWLM) kent dan een serie 'loops' (sprints) binnen een grote 'loop' (lange termijn planning of release). Deze serie loops zijn stappen op een meer korte termijn, omdat men in staat moet zijn snel en adequaat te reageren op nieuwe informatie-eisen. In een aantal kortcyclische stappen (Plan, Run, Evaluate) wordt alignment tussen business en IT life-cycles gewaarborgd. Daartussen vindt voorbereiding, implementatie en uitvoering plaats. Door kleinere releases te laten plaatsvinden, wordt het grote datawarehouse uiteindelijk vaker up-to-date gemaakt in relatie tot de bedrijfscontext. Bij grotere organisaties als Shell en Unilever heeft men hier al ervaring mee.

### Langdurig rendement

Conclusie is dat bij een langdurig rendement op een investering in informatievoorziening er oog moet zijn voor brede aansluiting met de dagelijkse bedrijfsvoering. Dit wordt gerealiseerd door het inrichten van gedegen governance processen, waarbij de (technisch en functionele) data stewards een kritieke rol in de uitvoering en richting spelen. Zij slaan de brug tussen de informatievrage en de mogelijkheid om deze te beantwoorden. De ICT-oplossing (datawarehouse, repository) moet voldoen aan de inrichtingseisen om blijvend aan de vraag te kunnen voldoen. Het ontwerp van het datawarehouse moet ondersteunend zijn aan een 'agile'-werkwijze. Bij elk nieuw initiatief (of uitbreiding op bestaand datawarehouse) moet dus worden stil gestaan bij het feit dat er vaak wijzigingen zullen worden doorgevoerd.

## Gevaren bij de bouw van een datawarehouse

Een gevaar bij de bouw van een datawarehouse is dat er slechts oog is voor de initiële oplevering. Daarmee wordt gewerkt aan de bouw van een 'dead-end' oplossing, welke een kort leven beschoren zal zijn. We onderkennen grofweg twee verschijningsvormen van een dead-end oplossing:

1. Een maatwerkoplossing zonder oog voor toekomstige veranderingen;
2. 'Silo' oplossingen die als standaard worden meegeleverd met een administratief ICT-systeem.

Als een systeem van de eerste vorm vastloopt omdat het niet flexibel blijkt te zijn, vindt men de oplossing niet in het maken van nieuwe technologische keuzes. Dan is er namelijk slechts sprake van vervanging van een 'dead-end' BI-omgeving, met louter nieuwe technologie. Bovendien blijken de kosten dan ook te stijgen en wordt het probleem van slechte aansluiting op de bedrijfscontext en -veranderingen niet opgelost, maar alleen vooruitgeschoven. De vernieuwingsvraag komt op termijn weer terug. Er verandert feitelijk niets. De tweede 'dead-end' oplossing ontstaat door silo's of standaardoplossingen van administratieve systemen te implementeren. Deze vorm mist elke interoperabiliteit met andere systemen en business processen. Het verleidt om af te stappen van goede besluitvorming met als doel concurrentievoordeel te behalen en is derhalve ook geen verstandige keuze. Het lijkt een eenvoudige oplossing om grote maatwerk datawarehouses, datamarts en risico's te mijden met standaarden, maar feitelijk introduceert men halve oplossingen. Met zo'n oplossing is de informatievoorziening gedicteerd door de externe leverancier in plaats van de eigen besluitvormingsprocessen. Laat staan de beperkte mate van verandering die hierop kan plaats vinden, zodat de informatiebehoefte wijzigt.

De data steward heeft een leidende rol in het data governance- en DWLM-proces, omdat de steward de *linking pin* is tussen de gebruikersvragen en de ICT-oplossing. Door uiteindelijk de processen van Datawarehouse Life-cycle Management te koppelen aan de data governance processen wordt die brug volledig geslagen en is de investering langdurig gewaarborgd.

De goede weg naar duurzame informatievoorziening begint dus bij een doel aan de start van deze weg. Daar wordt het einddoel van het project of initiatief bepaald en wordt de weg er naar toe grotendeels uitgestippeld. Data governance en DWLM zijn de hulpmiddelen die ondersteuning bieden om het pad te kunnen blijven volgen. De data stewards zijn de gidsen door het landschap en zij sturen bij op basis van veranderende omstandigheden. Maar nog belangrijker: voor deze reis geldt dat de weg naar de bestemming zeker zo interessant is als de uiteindelijke bestemming zelf.

**Joost Henskens** is Directeur Consulting bij Kadenza.