

Scheiding van Data Shop en Data Studio

# Gegevensexploitatie: voorbij Business Intelligence

Gerton Kuis

**E**r is een omvangrijke Business Intelligence (BI) industrie ontstaan rondom het feit dat organisaties niet in staat zijn om voldoende stuurinformatie te halen uit hun enorme, altijd groeiende hoeveelheid gegevens. Vanuit het productieproces zijn veel organisaties meestal moeilijk in staat om over de juiste informatie te beschikken. Trends, dwarsverbanden, segmentaties en aggregaties op bedrijfsniveau, alsmede geïntegreerde klantprofielen, zijn nu eenmaal moeilijk te maken vanuit een veelheid aan product- of dienstgerichte processen en systemen.

Het meest genoemde doel van BI is het verkrijgen van informatie die gebruikt

wordt ter ondersteuning van het nemen van strategische beslissingen. Hoewel dit een zinvol streven is, is het ook een beperkt streven, en wel beperkt op twee aspecten:

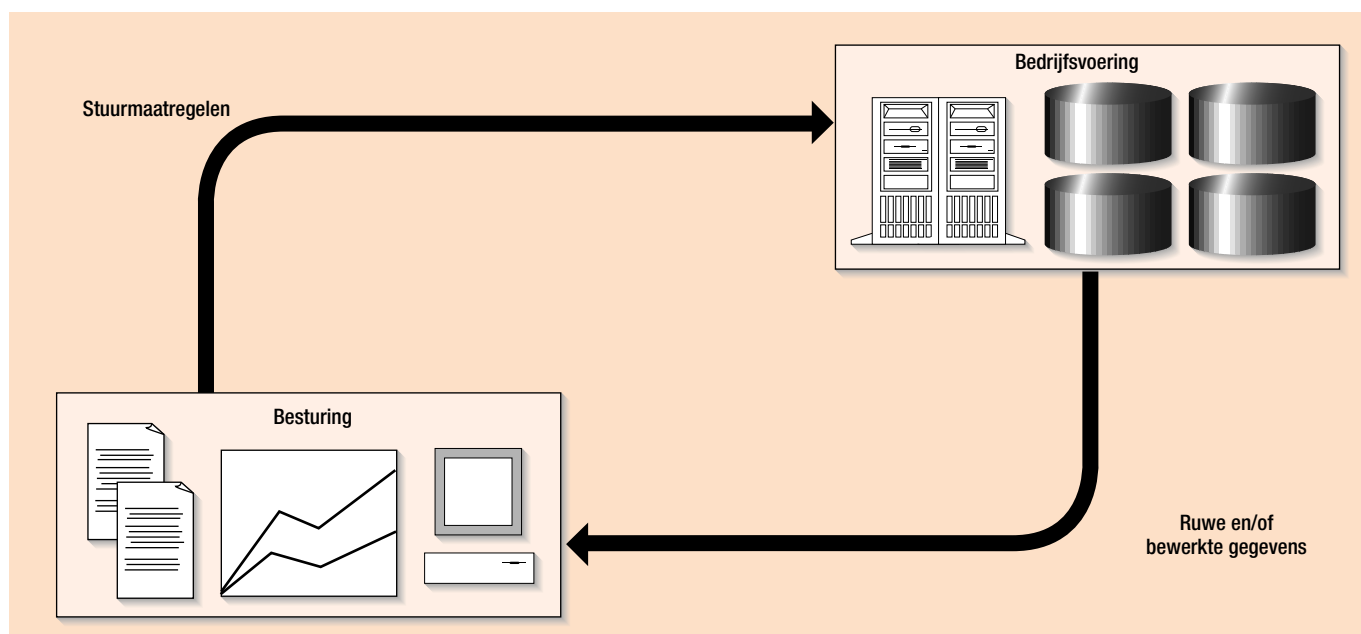
Het gaat om het verkrijgen van *informatie*, op basis waarvan beslissingen kunnen worden genomen

Het gaat om de behoeften op *strategisch* niveau. Hiermee wordt voorbij gegaan aan behoeften op tactisch of operationeel niveau, die hier meestal los van worden gezien. Het kan daarbij gaan om behoefte aan informatie, maar net zo goed om behoefte aan technische verbeteringen (bijvoorbeeld conversie of herstructurering van gegevens ten behoeve van productieprocessen) of kwalitatieve verbeteringen. Dit roept derhalve om een veel bredere

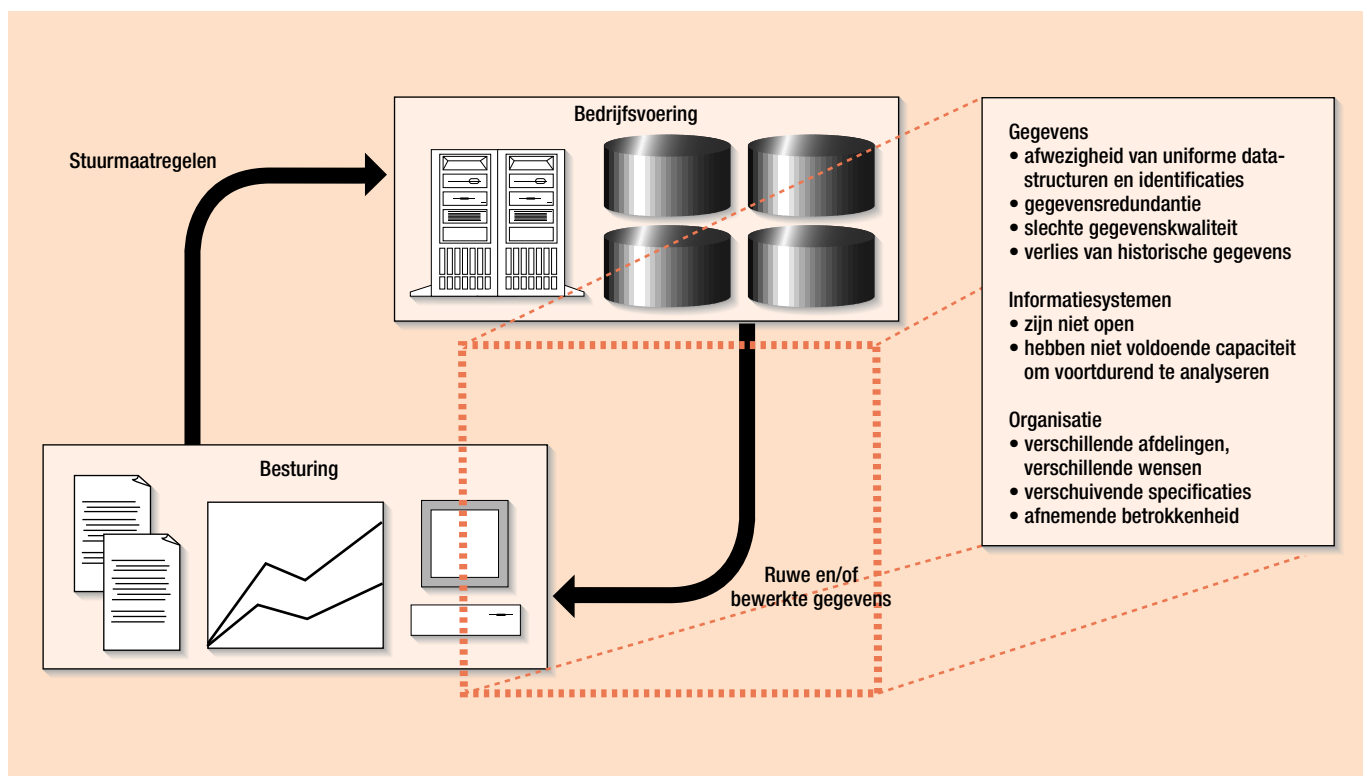
visie op het uitnutten van gegevens, die we zullen aanduiden met de term *gegevensexploitatie*. Het nastreven van gegevensexploitatie zoals hier gedefinieerd betekent echter dat een organisatie moet omschakelen: van het incidenteel oplossingen genereren voor specifieke gegevensbehoeften naar een vaste, heldere structuur waarin continue gegevensexploitatie bedreven kan worden. De vraag is dus: hoe wordt gegevensexploitatie een integraal onderdeel van de bedrijfsvoering?

## GEGEVENSEXPLOITATIE ALS CYCLUS

De primaire (productie-) processen in een organisatie zijn bij uitstek de plaats waar



FIGUUR 1: GEGEVENSEXPLOITATIE.



FIGUUR 2: EEN TE GROTE STAP.

ruwe gegevens worden geproduceerd: de grondstoffen voor gegevensexploitatie. De informatie en kennis die op basis van deze gegevens verkregen worden, en de besturing (op alle niveaus) die op deze informatie en kennis gebaseerd worden, hebben op hun beurt weer invloed op de primaire processen, eventueel via ondersteunende processen als marketing of account management. Er is dus sprake van een bepaalde cyclus van gegevens in ruwe of voorbewerkte vorm naar stuurmaatregelen. Figuur 1 geeft deze cyclus in een simpele vorm weer.

In feite is hier sprake van het klassieke besturingsparadigma, waarbij gegevens staan voor het "model van het bestuurde systeem".

Figuur 1 stelt de zaken nogal simpel voor. Enerzijds omdat het hier niet gaat om twee eenvoudige stappen; anderzijds omdat de richting van de pijlen uitsluitend de fysieke afhankelijkheden weergeeft. Gegevensexploitatie als geheel wordt echter gestuurd vanuit de bedrijfsvoering: er is een vraag naar stuurmaatregelen ter verbetering van bestaande processen, dan wel om vernieuwingen (klantgroepen, marktsegmenten) in de bedrijfsvoering aan

te brengen. Deze vraag doorloopt de cyclus de andere kant op: welke doelstellingen worden nagestreefd, welke maatregelen moeten daarvoor getroffen worden en welke informatie is er nodig om dit effectief aan te sturen.

### BREDE VISIE

Het kenmerk van gegevensexploitatie als bredere visie op het uitnutten van gegevens komt naar voren indien we inzoomen op de linker pijl van figuur 1. In de inleiding is reeds gesteld dat BI-oplossingen zich met name richten op het voorzien in informatiele behoeften op strategisch niveau. Het is hiervoor noodzakelijk om ruwe gegevens uit de productieprocessen te integreren, uniformeren en aggregeren. Dit kan alleen op een zinvolle manier gebeuren indien er afspraken zijn over de semantiek van de gewenste gegevens, en de afbeelding van ruwe gegevens naar deze semantiek. Kortom: de specificatie van het transformatieproces. Ongeacht of de gewenste gegevens een hoog aggregatieniveau hebben zullen deze specificaties gedefinieerd moeten zijn op het laagste

detailniveau van de brongegevens. Een aggregaat ontleent zijn betekenis namelijk aan hetgeen geaggregeerd wordt.

Deze specificaties bevatten een schat aan herbruikbare kennis. Dit zal echter alleen onderkend worden indien we de oplossing voor informatiele behoeften op strategisch niveau niet los zien van bijvoorbeeld gegevensconversies of het opzetten van een basisregistratie uit ver-

**BI-oplossingen richten zich met name op het voorzien van informatiele behoeften op strategisch niveau**

schillende 'oude' bronsystemen. Hiervoor zijn namelijk dezelfde afspraken nodig over semantiek, integratie en uniformering, zeker indien op het gebied van gegevens zowel intern als extern een consistent beeld moet bestaan.

Kortom: als de in figuur 1 genoemde "stuurmaatregelen" breed worden opgevat, en kunnen variëren van concrete ope-

rationele veranderingen tot strategische besturing, dan wordt daarmee de basis gelegd voor een efficiënt (her)gebruik van kennis en ervaring op het gebied van gegevenstransformaties. Een gevolg hiervan is bijvoorbeeld dat een geslaagd Data Warehousing project niet een eindpunt hoeft te zijn, maar juist het begin van veel mogelijkheden. Op basis van het hergebruik van de opgedane kennis en ervaring kunnen, tegen relatief geringe inspanning, allerlei acties uitgezet worden die in de kern gebaseerd zijn op integratie, uniformering en transformatie van gegevens.

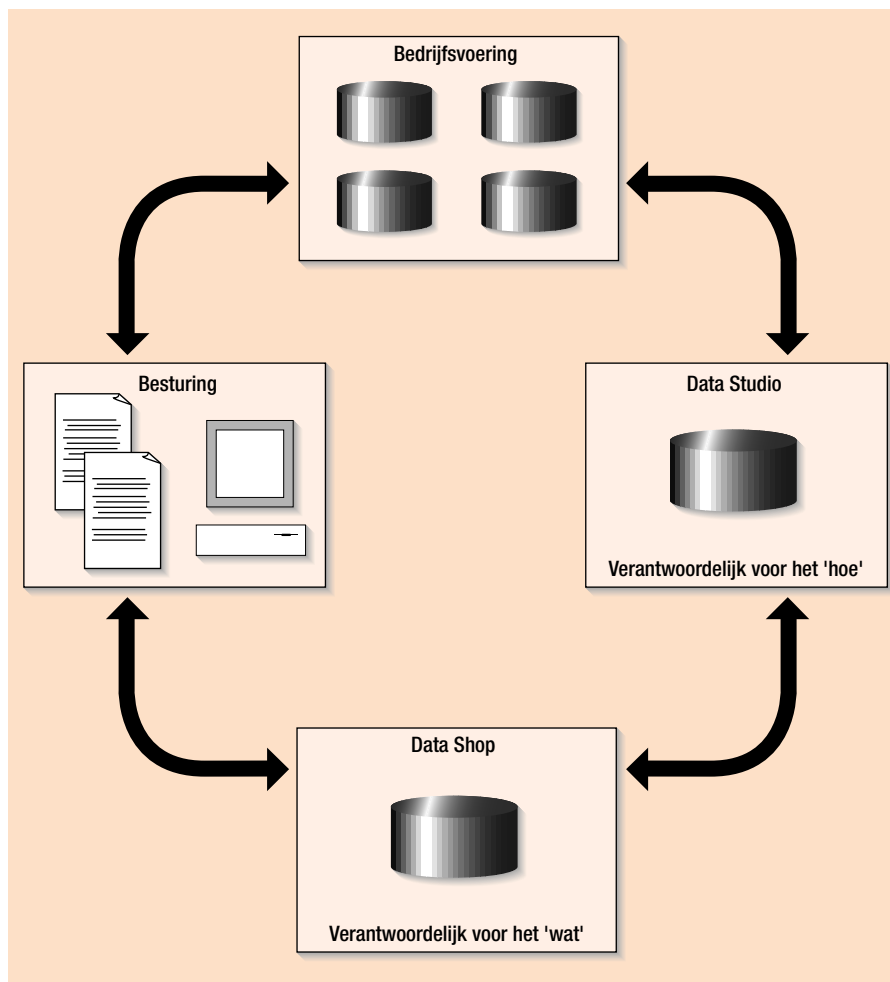
Hiermee is een veel bredere visie op het uitnutten van gegevens geïntroduceerd, en tevens een drijfveer voor het organisatorisch verankeren van de kennis omtrent gegevenstransformatieprocessen. Gegevensexploitatie als geheel krijgt zo een structureel karakter.

### ORGANISATIE VAN GEGEVENSEXPLOITATIE

Blijft over de vraag op welke manier gegevensexploitatie het best georganiseerd kan worden. Het doel is een constructie te vinden waarin enerzijds de kennis en ervaring op het gebied van gegevenstransformatieprocessen gebundeld worden, en anderzijds de besturing de juiste input krijgt voor het treffen van een breed scala aan stuurmaatregelen. We moeten derhalve de rechterpijl van figuur 1 onder de loep nemen.

Dat gegevens vanuit de bedrijfsvoering niet direct de juiste vorm hebben om besluitvorming op te baseren is het bestaansrecht voor de hele BI-industrie, kortom: dit is een erg grote en moeilijke stap. Figuur 2 somt de problemen op die zich daarbij "standaard" voordoen. Dit geheel aan problemen kan al zo complex zijn dat het gewenste eindresultaat nooit bereikt wordt. Deze stap bevat dan ook een dusdanige veelheid aan aspecten dat het eigenlijk niet als één stap gezien kan worden.

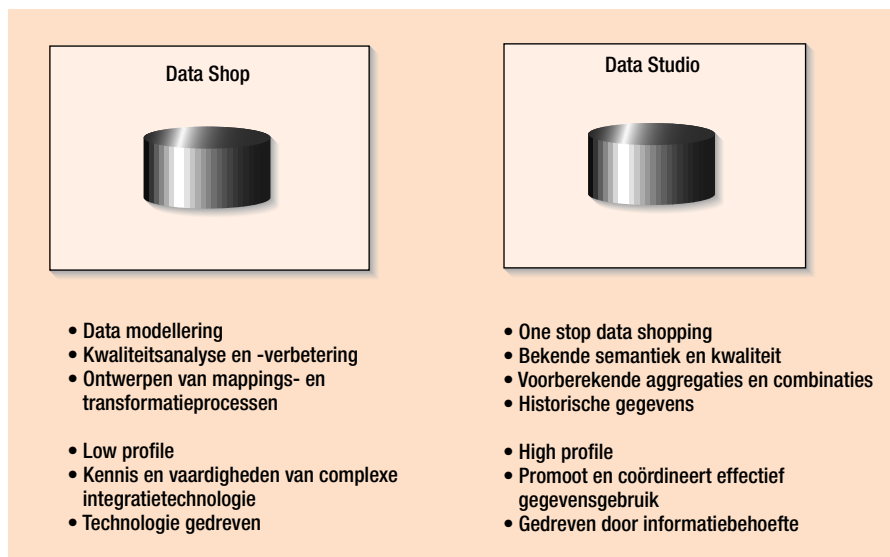
Het grote probleem is dat de gegevensspecifieke, technologische, functionele en organisatorische aspecten die hierbij aan



FIGUUR 3: DE ORGANISATIE VAN GEGEVENSEXPLOITATIE.

bod komen leiden tot een complexe mix. Dit is een van de redenen waarom bijvoorbeeld Data Warehouse projecten vaak zo moeilijk te beheersen zijn, en daardoor soms mislukken.

Het is daarom zaak om de complexiteit van deze stap te reduceren. Een goede manier om dit te doen is een aloud principe uit de informatica toe te passen, namelijk *separation of concerns*. Door onder-



FIGUUR 4: DATA STUDIO EN DATA SHOP.

scheid te maken tussen het "wat" en het "hoe" wordt de complexe mix van gegevens-, technische- en organisatorische aspecten die een project zo lastig maakt voorkomen. Figuur 3 laat de organisatie van de cyclus zien nadat dit onderscheid is aangebracht.

Hiermee zijn twee nieuwe concepten geïntroduceerd, die de structurele organisatie van gegevensexploitatie belichamen: de Data Studio en de Data Shop.

In de Data Studio worden processen vervaardigd om vanuit ruwe gegevens van diverse systemen te komen tot geïntegreerde, geüniformeerde en consistente gegevensverzamelingen. Daarnaast houdt de Data Studio zich bezig met tools en technologieën om deze veelal complexe processen op een efficiënte en onderhoudbare manier in te richten. De Data Studio is het werkterrein van ICT-specialisten, wier vaardigheden liggen op het gebied van gegevensmodellering, -kwaliteitsanalyse, -matching, -integratie en -transformatie. De Data Studio verzorgt het "hoe", werkt 'achter de schermen' in opdracht van de Data Shop, en is daarom minder zichtbaar voor de organisatie.

De kwaliteit van de ruwe gegevens heeft zijn impact op elke vorm van gegevensintegratie. Eén van de belangrijkste taken van de Data Studio is kennis opbouwen over de ruwe gegevens teneinde uitspraken te kunnen doen over de kwaliteit ervan. In het geval van gesloten, slecht gedocumenteerde legacy systemen is deze kennis soms al jarenlang geleden verloren gegaan. Zonder deze kennis loopt elke poging tot het integreren en uniformeren van gegevens het risico om voor zeer onaangename verrassingen te zorgen, bijvoorbeeld: datumfouten, niet-unieke sleutels, niet bestaande gegevens, dubbele registraties binnen één bron, coderingen die andere dan de "toegestane" waarden bevatten, inconsistenties in relaties en invoerfouten. Het is nagenoeg onmogelijk om een gegevenstransformatieproces volledig te specificeren (en te implementeren) zonder deze kennis te hebben. De Data Studio dient deze kennis daarom op te bouwen. Enerzijds wordt deze kennis voor

elke oplossing met betrekking tot gegevensintegratie in de Data Studio zelf (her)gebruikt. Anderzijds kan deze kennis leiden tot gerichte verbeteracties ten aanzien van de ruwe gegevens, en is het dus kennis die in de Data Shop ter exploitatie wordt aangeboden.

De Data Shop is een zeer prominent aanwezig bedrijfs onderdeel dat het effectieve gebruik van gegevens promoot en coördineert. In de Data Shop zijn consistente en geüniformeerde gegevenssets

### *De Data Studio is het loket voor "one-stop-data-shopping"*

beschikbaar met een bekende structuur, semantiek en kwaliteit, die afgenomen kunnen worden als basis voor analyse of gegevensuitwisseling. De Data Shop is de adviseur en balie van gegevens-afnemers, die daar ook hun vragen kunnen uitzetten voor gegevens die nog niet voorhanden zijn. De Data Shop houdt zich dus bezig met het "wat".

De aard van de gegevens die in de Data Shop aangeboden worden is divers, zowel op het gebied van aggregatieniveau als structuur. De Data Studio is het loket voor "one-stop-data-shopping", en voorziet de organisatie op alle niveaus van de benodigde gegevens.

Naast gegevens stelt de Data Shop ook kennis over de gegevens beschikbaar, die in de Data Studio is opgebouwd. Dit gaat, zoals reeds beschreven, verder dan statische, beschrijvende metadata (structuren, semantiek).

Figuur 4 somt de karakteristieken van beide bedrijfs onderdelen op. Samen vattend bereiken we met deze organisatorische indeling de volgende voordelen:

- Complexiteitsreductie door scheiden van "wat" en "hoe"
- Efficiëntie door hergebruik van kennis en specialisme in de Data Studio voor verschillende doeleinden

- Eén plek (Data Shop) waar semantiek, consistentie en kwaliteit gedefinieerd worden
- Eén loket voor afname van geïntegreerde gegevens en kennis over gegevens, waar de taal van de klant gesproken wordt
- Een zeer brede basis voor gegevensexploitatie

## HELDERE TAAKVERDELING

De focus bij het realiseren van een BI oplossing ligt terecht bij het doel: ontsluiten van moeilijk toegankelijke of verborgen informatie ten behoeve van strategische besluitvorming. Hierbij wordt echter vaak over het hoofd gezien dat de opgedane kennis en vaardigheden tijdens de realisatie van de oplossing minstens zoveel waarde hebben als de informatie die de oplossing beoogt op te leveren. Door te redeneren vanuit deze kennis en vaardigheden wordt een veel bredere visie op het uitnutten van gegevens zichtbaar: gegevensexploitatie. Hierdoor kan op alle niveaus in een organisatie in de gegevensbehoefte voorzien worden.

Gegevensexploitatie heeft een meer structureel karakter, hetgeen een goed concept voor de organisatie ervan noodzakelijk maakt. In dit artikel is zo'n concept voorgesteld. De cyclus van ruwe gegevens naar stuurmaatregelen krijgt hiermee een stabiele vorm, waardoor opgedane kennis behouden en optimaal benut wordt, en waarin oplossingen kunnen evolueren. Daarnaast levert de scheiding van het "wat" (Data Shop) en het "hoe" (Data Studio) een heldere taakverdeling op. Hiermee wordt de vermenging van technische en organisatorische aspecten, die het realiseren van effectieve gegevensexploitatie zo complex maakt, voorkomen. ●

Ir. Gerton Kuis is senior consultant bij IT Consultancy Group.