



BI moet tot wasdom komen als actiegericht kennisplatform

Weg met beperkende definities

Joost Enzing

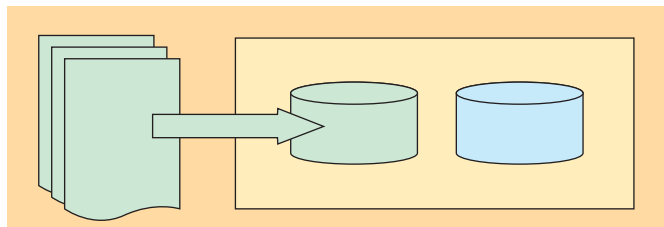
Business Intelligence levert informatie gebaseerd op gestructureerde gegevens, terwijl ongestructureerde gegevens buiten beschouwing worden gelaten. Dit ondanks dat 80 procent van de bedrijfsgegevens ongestructureerd is.

BI beperkt haar doelgroep tot managers en negeert daarmee 90 procent van de medewerkers. BI ondersteunt bij het nemen van beslissingen, maar verzaakt bij ondersteuning van de benodigde acties. Kortom, het is de hoogste tijd om BI minder eng te gaan benaderen en daarmee concurrentievoordeel te behalen. Want alleen wanneer breder gedacht wordt, kan BI tot wasdom komen als organisatiedekkend actiegericht kennisplatform. Weg met beperkende definities als 'het genereren van management-informatie vanuit gestructureerde gegevens'.

Verruiming

Als eerste de beperking 'management informatie'. Microsoft roept al enige tijd *BI for the masses* en heeft het over *the information worker*, Information Builders is trots op haar sterke positie als leverancier van *operational BI*, terwijl weer anderen seminars houden over *pervasive BI* (alomtegenwoordige BI). Was BI ooit synoniem voor management-informatie, momenteel laten de slimmere organisaties zich ermee doordrenken, van strategisch niveau via het tactische tot op de werkvloer.

En dat alles met een goede reden. Hoe meer mensen de juiste informatie in de juiste vorm op het juiste moment krijgen, hoe beter doelstellingen behaald worden. Vaak kan reeds verzamelde informatie zijn waarde met relatief weinig inspanning exponentieel vergroten; voor elke werknemer een toegesneden BI-product.



Afbeelding 1: Ongestructureerde gegevens worden gestructureerd.

En dan hebben we het nog niet eens over klanten, leveranciers en andere betrokkenen. De manager moet zijn 'speelgoed delen' en wordt daarvoor beloond met resultaatverbeteringen.

Technisch zal het opheffen van deze beperking vooral neerkomen op het implementeren van bestaande technologieën, met extra aandacht voor onder andere performance en autorisatie. De grootste uitdaging ligt op organisatorisch vlak.

De database krijgt er een collega bij

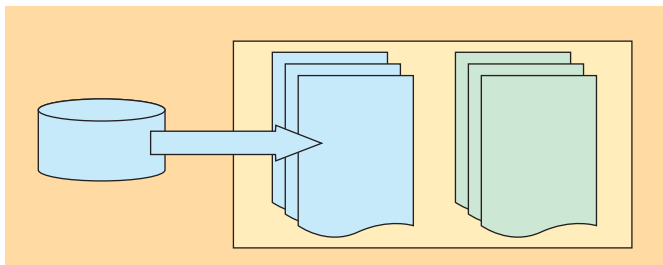
Als tweede de beperking 'vanuit gestructureerde gegevens' (lees: vanuit bron-databases). Tijden lang hadden we onze handen vol aan het ontsluiten van de beschikbare databases. Ondertussen groeide de stapel ongestructureerde gegevens: e-mail, Word files en andere documenten vol informatie. Nu de ontsluiting van gestructureerde gegevens veelal goed op de rails staat, is er geen excuus meer om kennis te laten liggen alleen omdat het niet in een database staat. Vanuit hun verschillende aard leiden ongestructureerde en gestructureerde gegevens gezamenlijk tot evenwichtige beslissingsondersteuning.

Hier is technisch werk aan de winkel. Het integreren van gestructureerde gegevens met ongestructureerde gegevens is met recht een uitdaging te noemen. Inmon benoemt het expliciet in zijn DW2.0, maar een standaard aanpak is er nog niet. Hier valt eer voor technici en concurrentievoordeel voor managers te behalen. Het vervolg van dit artikel concentreert zich op dit vraagstuk.

Oplossingsrichtingen voor integratie

De twee meest voor de hand liggende oplossingsrichtingen om gestructureerde en ongestructureerde gegevens te combineren zijn:

1. *Ongestructureerde gegevens structureren.* Leveranciers als Informatica bieden producten waarmee ongestructureerde documenten worden ingelezen en behandeld, om uiteindelijk in een database te worden opgeslagen. (De behandeling is veelal



Afbeelding 2: Gestructureerde gegevens worden ongestructureerd.

niet standaard en dient te worden ontwikkeld op basis van de specifieke klantbehoefes.) Het resultaat is een gevulde database die als 'normale' database in de BI architectuur kan worden opgenomen (zie afbeelding 1);

2. *Gestructureerde bronnen behandelen als ongestructureerde documenten.* Een andere benadering is BI-eindproducten, zoals rapportages, te laten indexeren als waren ze gewone documenten. Information Builders biedt hiervoor een oplossing mede gebaseerd op Google-technologie. De gebruiker ziet in zijn zoekresultaten links naar gestructureerde en ongestructureerde gegevens gebroederlijk naast elkaar (zie afbeelding 2). Hou op dit gebied de term 'Enterprise search' in de gaten.

Overeenkomst tussen beide benaderingen is dat steeds één van de twee gegevenssoorten zich aanpast aan de andere. Het voordeel hiervan is dat na een dergelijke behandeling bestaande technieken kunnen worden losgelaten op de geüniformeerde gegevens-pool. Zo heeft SAS een mooie case van een organisatie die vrachtbrieven structureert en de database inlaadt. Vervolgens worden datamining-technieken losgelaten en kunnen vrachtbrieven met een hoge fraudekans worden aangewezen. Ook haalden ze het nieuws met hun e-mail leugendetector. Technieken vanuit de gestructureerde hoek worden dus toepasbaar op de ongestructureerde. Een fictief voorbeeld van de als tweede genoemde aanpak is een marketingorganisatie die product managers toegang geeft tot een zoekmechanisme, waarmee zowel productbrochures als actuele rapportages betreffende producten kunnen worden gevonden. In dat geval kunnen dus technieken uit de ongestructureerde hoek hun diensten bewijzen voor gestructureerde gegevens.

Nadeel van de besproken benaderingen is dat bij aanpassing aan de andere soort, voordelen van de eigen soort verloren gaan. Zo raakt de illustratieve kracht van een bepaalde klachtenbrief verloren wanneer hij opgeknipt, gecategoriseerd en geaggregeerd met alle andere brieven wordt. En de flexibiliteit van een datamart blijft deels onbenut wanneer er voornamelijk nog voorgebakken rapporten uit worden ontsloten. Wat nu als we wel de voordelen willen van het combineren van gestructureerd met ongestructureerd, maar niet de genoemde nadelen? Dan lijkt een derde oplossingsrichting nodig:

3. *Gestructureerd en ongestructureerd overkoepelen.* Bij deze benadering passen we de gegevens niet aan, maar zoeken we de oplossing in het gebied rondom de beide soorten gegevens. Door

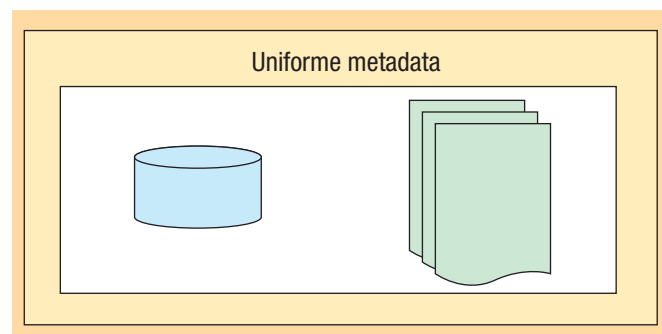
het compromisloze karakter lijkt dit de meest wenselijke oplossing. Hoe deze benadering daadwerkelijk vorm kan krijgen vraagt een nadere uitwerking.

Metadata als lijm

Realiseren van overkoepeling is gemakkelijker gezegd dan gedaan. Hoe weet de ontsluitende software de onderlinge relaties tussen gegevens? Hoe weet het welke gebruikers geautoriseerd zijn voor een gegeven? Hoe weet het de actualiteit, de betrouwbaarheid, de eigenaar, enzovoort. Dit is uiteraard de taak van metadata, oftewel de gegevens die onze verzamelingen gegevens beschrijven. Metadata zullen als lijm fungeren die de twee werelden na het samenbrengen ook daadwerkelijk laat samen-vloeien. Dit kan alleen wanneer de metadata uniform zijn voor beide soorten gegevens. Alleen dan kunnen de twee silo's technisch en functioneel integreren (zie afbeelding 3).

Technieken vanuit de gestructureerde hoek worden toepasbaar op de ongestructureerde

Stel nu dat we de genoemde uniforme metadata hebben en we de gegevenssoorten geïntegreerd kunnen gaan benutten. Wat voor tool zouden we dan inzetten? Business Objects? Cognos? WebFocus? Nee, geheel in de sfeer van verruiming nemen we hiervoor een kijkje in een parallelle wereld. Binnen de IT komt het namelijk veelvuldig voor dat sterk gelijkende vakgebieden langs elkaar heen leven. Ze begonnen ooit wezenlijk verschillend maar ontwikkelden zich beide in dezelfde richting. Uiteindelijk hebben ze meer dan alleen een raakvlak. Ze hebben significante overlap en daarmee veel potentie elkaar te kunnen benutten. Dit is ook waar BI mee te maken heeft, zeker nu ongestructureerde gegevens tot haar domein gaan behoren. In dit licht springt Enterprise Content Management (ECM) naar voren. ECM, een geschikte fusiekandidaat die, eenmaal versmolten met BI, de gehele organisatie van alle relevant informatievormen voorziet (noem het Enterprise Information Management).



Afbeelding 3: Het overkoepelen van beide soorten gegevens.

De AIIM (Association for Information and Image Management) defineert ECM als volgt: *Enterprise Content Management (ECM) is the technologies used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists.* (Zie www.aiim.org/about-ecm.asp).

Concreet vallen onder ECM onder andere Document Management, Web Content Management, Workflow, Collaboration en Archiving. Het is dus een vakgebied waar al een hoop werk verzet is waarvan BI kan profiteren. Wat zijn de voordelen die een dergelijke combinatie kunnen brengen?

Uiteraard de ontsluiting van gestructureerde en ongestructureerde gegevens binnen één systeem, zodat een OLAP-analyse kan doordrillen tot aan documentniveau. De manager kan bijvoorbeeld vanuit een grafiek de offertes openen die een opvallende piek hebben veroorzaakt. Of de voorman krijgt naast een SMS alert ook direct een bestand met de benodigde noodprocedure gemaïld wanneer de BAM-applicatie een grensoverschrijding ziet. Zo zijn er nog vele toepassingen te bedenken.

ECM is gebaseerd op een sterke infrastructuur waar BI kan meeliften. Hierdoor is maar één autorisatielaag en één gebruikers-interface nodig. Voor gebruikers is dit prettig en voor beheer kostenbesparend. Ook zaken als archivering, backup en dergelijke kunnen door BI worden benut.

ECM maakt BI actiegericht (actionable). Wanneer er in een BI-product een 'sein op rood springt', start ECM een workflow op voor de opvolging hiervan. De verantwoordelijken krijgen taken en de voortgang wordt gevolgd. Een ander voorbeeld is het gebruik van collaboration. Een medewerker ziet iets vreemds in zijn rapportage en start een digitale discussiegroep. Op het bewuste rapport kunnen ze virtuele aantekeningen maken, ze kunnen berichten plaatsen, MSN'en en documenten uploaden.

Conclusies

Momenteel worden veel organisaties nog beperkt door een te enge definitie van BI. Ze sluiten gebruikers buiten omdat het geen managers zijn. Ze sluiten kennisbronnen uit omdat ze niet gestructureerd zijn. Ze beperken de functionaliteit tot het bereiken van informatie terwijl veel meer kan worden geboden. Wanneer ze deze beperkingen van zich afschudden en met frisse blik kijken, opent zich een wereld vol nieuwe mogelijkheden. Iedereen is een potentiële BI-gebruiker, alles is een potentiële kennisbron. En BI ondersteunt beslissingen én daaropvolgende acties zodat resultaten worden bereikt.

De organisatie moet aan de slag met doordacht metadata management en selectie en inrichting van de juiste tools.

Joost Enzing

Drs. J.J. Enzing is BI Consultant bij VLC.



Ben jij ook fan van K3?

Kwaliteit, Kennis en Kunde

Join FourPoints !!

Dé specialist in Data Warehousing & Business Intelligence

DATA WAREHOUSING | BUSINESS INTELLIGENCE | ARCHITECTUUR

Interesse? Neem contact op met Steven Tjin | Stjin@FourPoints.nl | T +31 (0)20 - 452 75 05 | www.FourPoints.nl