

Op weg naar betrouwbaar Metadata Management

Datakwaliteit of Informatiekwaliteit

Peter Alons

Recent zijn veel publicaties verschenen over datakwaliteit. Vaak betreffen deze artikelen manieren om langs technische weg de kwaliteit van bedrijfsgegevens, en vaak gaat het dan met name over klantgegevens, te verbeteren. Maar wat datakwaliteit in algemene zin is, is een belangrijke vraag om te stellen.

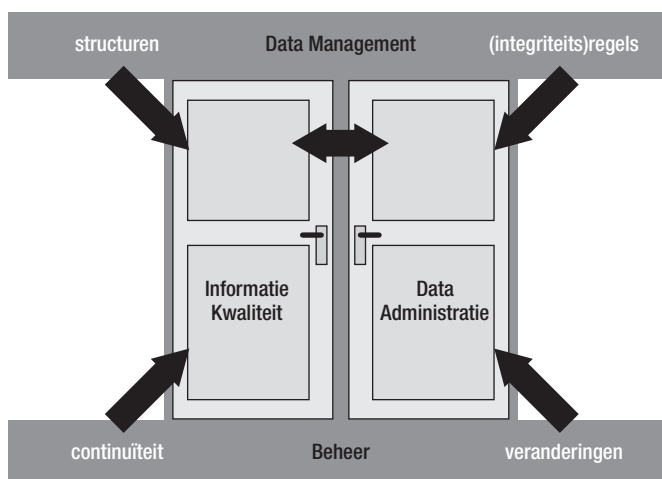
Ik praat niet graag meer over datakwaliteit. Het blijkt bij nadere beschouwing, althans in algemene zin, een wollig begrip te worden. Met collega adviseurs heb ik, in het kader van Leergangen en Master classes over datawarehousing, al een aantal jaren datakwaliteit als onderdeel van ons programma besproken. En daarbij blijkt datakwaliteit in zijn algemeenheid moeilijk vast te pinnen. Er is een flink aantal zinnige criteria te hanteren: volledigheid, correctheid, geldigheid, precisie, enkelvoudigheid (= non-duplicatie), consistentie, tijdigheid van toegang, gangbaarheid voor kennis, objectiviteit bij presentatie, rechtvaardiging (naar klanten), conformiteit, consolidatie, integratie, standaardisatie, om maar eens wat te noemen. Welke eisen stel je in een gegeven praktijk aan je gegevens? Welke prioriteiten stel je daarbij en waarom? En krijg je met deze vragen gemakkelijk een toegesneden aanpak? In de praktijk worden door verschillende IT-specialisten verschillende dingen verstaan onder datakwaliteit. De antwoorden op bovenstaande

vragen zijn daardoor niet eenduidig te geven. De laatste tijd praat ik daarom liever over informatiekwaliteit. Van dat begrip valt door eindgebruikers, zonder terug te vallen op wiskundige definities, gemakkelijk een helder beeld te geven: informatie *deugt* of *deugt niet*. En het verschil daartussen valt in iedere praktische situatie gemakkelijk aan te geven, al is het maar simpel in termen van de gevolgen. Daarmee blijkt achter die woordverschuiving meteen een ontzagwekkend verschil schuil te gaan. Kwaliteit is geen puur technische zaak, het is in eerste instantie een zaak van mensen. Laten we daar eens in wat detail naar kijken.

In mijn verhalen over het Metadata Frame, waarover ik in DB/M een aantal artikelen heb gepubliceerd [1-3], begin ik altijd met twee wel te onderscheiden deuren naar functies binnen een bedrijf: een deur met Informatie Kwaliteit en een deur met Data Administratie (zie afbeelding 1). Larry English, auteur van het boek 'Improving Data Warehouse and Business Information Quality' [4] en daarmee een van de meest prominente en erkende goeroe's op het gebied van Informatiekwaliteit, noemt deze twee functies: Informatiekwaliteit-management en Informatie-management. Informatie-management omschrijft hij als: "de toepassing van gezonde management-principes, processen en praktijken op informatie als een strategische resource van de onderneming". Informatiekwaliteit-management omschrijft hij als: "de toepassing van bewezen kwaliteitsmanagement-principes, processen en praktijken op informatie *als een product van de business-processen*". En het verschil hiertussen uit zich onder andere in de principes die er achter schuil gaan.

De principes van Informatie-management kunnen worden samengevat als:

- Gegevens zijn een waardevol bezit, die verzameld en beheerd worden ten voordele van de onderneming en haar afdelingen langs vastgelegde standaarden en richtlijnen;
- Gegevens zijn de enige waardevolle bezittingen van een onderneming die volledig herbruikbaar zijn, niet uitgeput raken door gebruik; daarmee zijn ze de laatste die een hoge redundantie vereisen;
- Gegevens hebben één enkele administrateur, helder gedefinieerde locaties, zijn toegankelijk voor ieder die ze nodig heeft, en worden in zo hoog mogelijke mate gedeeld binnen de gegeven eisen van beveiliging en vertrouwelijkheid;



Afbeelding 1: Twee deuren: dual task force.

- Records in elektronisch formaat worden bewaard en beheerd, en blijven toegankelijk voor hun aangegeven periode van retentie.

Als principes van Informatiekwaliteit-management zijn aan te merken:

- Informatie en gegevens zijn strategische resources van een onderneming;
- Kwaliteitsinformatie leidt tot concurrentievoordelen en business-effectiviteit;
- Informatiekwaliteit is geen geïsoleerde functie, maar een inherent en integraal onderdeel van business management;
- *Iedereen* in de organisatie heeft een rentmeesterschapsrol voor informatiekwaliteit;
- Zonder informatie zal een onderneming failliet gaan; zonder kwaliteitsinformatie zal een onderneming gesuboptimaliseerd raken en mogelijk als gevolg daarvan failliet gaan;
- De kosten van slechte kwaliteit in informatie zijn hoog: het veroorzaakt procesverstoringen, en sloop- en herstelwerkzaamheden, die menskracht, geld, materiaal, en facilititeit-resources verspillen, en bovendien zowel medewerkers als klanten frustreren en vervreemden.

Alleen al een scherpe blik op het verschil in bovenstaande principes helpt het ontzagwekkend verschil tussen datakwaliteit en informatiekwaliteit te ontmaskeren. Dit verschil kan nader worden geïllustreerd aan voorbeelden van de hoge kosten van non-kwaliteit in informatie, zoals die enerzijds zijn te vinden in ref. [4] en aangevuld met recente voorbeelden uit het nieuws in Nederland.

1. Uit *The Tennessean*, 12 juli, 1998 [4]: Een arts in Florida amputeerde het verkeerde been van een patiënt. De oorspronkelijke instructie welk been het betrof was veranderd, maar de arts volgde de oude opdracht. De verpleegster die wist dat de opdracht was gewijzigd, had de operatiekamer net voor de operatie verlaten, maar had aangenomen dat de arts ook op de hoogte was van de wijziging. Drie jaar later zag dezelfde arts kans na te laten de naam op de polsband van een patiënt te controleren en als gevolg een riskante operatie uit te voeren op de kamergenoot van de betreffende patiënt. Hierna werd zijn licentie ingetrokken.
2. Een petroleummaatschappij die een nieuwe bron aanboorde in de Noordzee, boorde door de schacht van een aangrenzende bron, omdat door gebrekkige gegevens de precieze locatie van deze schacht verkeerd was ingeschat. De bijbehorende bron gaf gelukkig geen olie meer. Anders zou de druk van de olie in de beschadigde schacht omhoog zijn gegutst in de nieuwe schacht en daarbij een investering van vijfhonderd miljoen euro aan stukken hebben gereten, waarschijnlijk met dodelijke gevolgen voor de bemanning [4].
3. Uit *The Wall Street Journal*, 24 december, 1997 [4]: Oxford Health Plans Inc. maakte bekend dat door computerfouten bij het overgaan op een nieuw computersysteem en daaruit

voortvloeiende gegevensfouten inkomsten overschat en medische kosten onderschat waren. Bijkomende kwaliteitsproblemen hadden geleid tot te hoge facturen aan klanten. Het gevolg was een verlies van 69 miljoen dollar, waardoor hun aandelen 62 procent in waarde daalden. De staat New York gaf Oxford ook nog eens een boete van drie miljoen dollar voor het schenden van verzekeringswetgeving en droeg Oxford op 500.000 dollar aan klanten terug te betalen. In 1998 waren de aandelen van Oxford nog maar 8 dollar waard – 9 procent van hun hoogste waarde van 89 dollar – en hing hun toekomst aan een zijden draad.

4. Barbra Streisand heeft eind vorige eeuw haar gehele investeringsportefeuille van een investeringsbank teruggetrokken, omdat deze haar naam foutief als Barbara had gespeld [4].
5. [Teletekst, 2002]: In 2002 werd in een ziekenhuis in Amerika de bloedgroep bij een donorhart foutief geregistreerd als A+. Kort daarna werd het hart geïmplantemd bij een jonge vrouw. Toen bleek dat de bloedtypering niet klopte was het te laat; een nieuwe operatie was niet mogelijk en de vrouw overleed kort daarna.
6. [Teletekst, 2005]: Als gevolg van fouten in een nieuw programma voor apotheken, worden de te hanteren doses voor sommige zware medicijnen van patiënten met onder andere hartklachten met een factor tien overschat. Apothekers worden door de overheid opgedragen alle computergegevens handmatig na te rekenen, alvorens een recept uit te reiken (sic!).

In de praktijk worden verschillende dingen verstaan onder datakwaliteit

Wat al deze voorbeelden vooral aantonen is, dat kwaliteit en dus ook informatiekwaliteit in de eerste plaats een zaak van *mensen* is en hun gedrag. Zonder mensen heeft kwaliteit geen zinnige inhoud. Voor zover wij weten zal een boom zich niet bekommeren om informatiekwaliteit, een rots wel zeker niet. In het eerste nieuwsvoorbeeld is eerst sprake van herstel van een fout in een medisch dossier, die vervolgens niet goed wordt doorgegeven: het informatieproces deugt niet. Drie jaar later wordt op zich (hopelijk) correcte informatie door dezelfde arts geheel genegeerd. De betreffende investeringsbank van Barbra Streisand zou wellicht het hele voorval kunnen afdoen met de opmerking, dat mevrouw Streisand een 'bitch' is, maar schelden doet geen pijn. Ze hadden moeten beseffen, dat zij – aldus gekwalificeerd – een 'very rich bitch' is, en dat verschil deed wel pijn. En hoe zou de fout in het vijfde voorbeeld zijn ontstaan? Een data-entry proces, dat om 16.55 uur niet kon worden afgesloten zonder eerst het verplichte veld bloedgroep in te vullen, terwijl dat niet goed op het etiket te lezen was? En wat doet een apotheker in het laatste voorbeeld, als hij bij het narekenen van een computerresultaat een

andere uitkomst krijgt. Automatisch aannemen, dat de computer het weer fout heeft en zijn eigen berekening goed is? Wat is de waarde van vermeend ondeugdelijke informatie. Wat doet dat met vertrouwen? Hoe zijn inmiddels de kosten en waarde van dit computersysteem te typeren?

Volgens Larry English dient informatiekwaliteit bekeken te worden in het licht van de waardeketen: Informatie leidt – door haar betekenis in groter verband – tot Kennis, Kennis leidt (door zinnige acties) tot Wijsheid. Maar vervang het basisingrediënt informatie door non-kwaliteit informatie en het resultaat wordt Dwaasheid. In dit licht is het ook goed eens te kijken naar wat de hoofdproducten zijn van Informatiesystemen. Volgens Shoshana Zuboff [5] zijn dit: *getransformeerd werk om waardetoevoegende activiteiten te vermeerderen en verspilling te verminderen, en 'geïnformeerde' kenniswerkers*. Als serviceproducten zijn *gecontroleerde processen en gecontroleerde informatie* aan te merken. Alle overige producten van informatiesystemen, zoals IT-infrastructuur, gegevensmodellen, databases, datawarehouses, object bases, web bases, kennisbanken, repository-infrastructuur en andere componenten zijn aan te merken als *bijproducten of hulpmiddelen*.

Informatiekwaliteit is in de eerste plaats een zaak van mensen en hun gedrag

Wat zegt dat over de twee deuren in afbeelding 1? Wel, als IT'ers leggen wij al snel de nadruk bij de rechterdeur. Daar zijn we thuis en weten we wat te doen. Maar achter de linkerdeur wordt in elk bedrijf veel geld verdiend ... of verloren! De rechterdeur staat in dienst van de linker. En het geheel gaat voornamelijk over mensen, en dan pas over techniek en technische hoogstandjes. Natuurlijk mogen we achter de rechterdeur best zoeken naar middelen die onze klanten werkelijk effectief en efficiënt ondersteunen. Dat is het doel van betrouwbaar Metadata management. Maar niet zonder ons in de eerste plaats serieus te bekommeren om wat de zaken achter de andere deur goed laat gaan. En dat zijn vooral zinnige procedures en het stimuleren van juist gedrag. Vanuit de rechterdeur moeten we vooral *die* zaken ondersteunen, bijvoorbeeld door consequent gebruik te maken van communicatie. Dat is ook een van de uitgangspunten van ons Metadata Frame. En daarover vallen in dit kader best nog heel interessante dingen te melden.

Ik praat niet graag meer over datakwaliteit. Het blijkt bij nadere beschouwing, althans in algemene zin, een wollig begrip te worden. Er worden door verschillende IT-specialisten verschillende dingen onder verstaan. Het begrip komt daarmee in de praktijk feitelijk achter de rechterdeur vandaan. Ik praat daarom liever

over informatiekwaliteit. Dat is een begrip waarvan door eindgebruikers, zonder terug te vallen op wiskundige principes, altijd een helder beeld valt te geven. En achter die simpele woordverschuiving gaat een ontzagwekkend verschil schuil. Alleen als we – als IT'ers samen met onze klanten – in staat en bereid zijn dat verschil werkelijk aan te voelen, hebben we een kans er in de praktijk met succes iets aan te doen.

Literatuur

1. Database Magazine, jaargang 11, no. 8 (december 2000).
2. Database Magazine, jaargang 12, no. 1 (januari 2001).
3. Database Magazine, jaargang 16, no. 2 (maart 2005).
4. Improving Data Warehouse and Business Information Quality, Larry P. English, John Wiley & Sons Inc., 1999
5. Information Quality Management: The Next Frontier, DAMA/Benelux/IAIDQ Conference, Leuven, België, maart 2005

Peter Alons

Dr. Peter W.F. Alons (Peter.Alons@AtosOrigin.Com) is senior consultant bij Atos Origin/BI-CRM en betrokken geweest bij een groot aantal BI- en datawarehouse-projecten bij diverse bedrijven.

Online archief Database Magazine

Database Magazine-lezer opgelet! Artikelen over onderwerpen als Datawarehousing, SQL, ETL, Business Intelligence, Relationale databases, modellering en nog veel meer vindt u in het Online Archief van Array Publications. Vaktijdschriften als Storage Magazine, Database Magazine, IT Service Magazine, Java Magazine en ons Oracle vakblad Optimize hebben hun artikelenarchief online gezet. Met een Google-achtige zoekstructuur vindt u snel wat u zoekt op www.dbm.nl

Adverteerdersindex

Array Seminars	42, 44
BI4U	2
Business Objects	47
Computable	17
LogicaCMG	4
Software AG	48
Solipsis	8