

Bilateraal

Datakwalitijd

Paul van der Linden

Wie verantwoordelijk wordt gemaakt voor datakwaliteit weet het zeker: je baas heeft een ontzettende hekel aan je! Natuurlijk wordt het positief gebracht maar op een bepaalde manier houdt je er geen blij gevoel aan over. Het is meer een misselijk makend gevoel in je buik, voorbehouden aan zwangere vrouwen. Wie datamanager, of nog erger datakwaliteitsmanager wordt, staat het huilen nader dan het lachen.

Niemand zal het belang van data en datakwaliteit ontkennen. Data is immers 'the life blood' van elke organisatie. Het is de grondstof voor informatie. De tijdige beschikbaarheid van correcte en eenduidige informatie aangeleverd in het juiste formaat wordt terecht gezien als een van de basisvoorwaarden voor bedrijfssucces.

Datakwaliteit is een omvangrijk en complex onderwerp. Omvangrijk, omdat het niet alleen gaat om het ontbreken van data of het hebben van verkeerde data. Het gaat ook om dataconsistentie (in, maar ook over systemen heen), om data-logistiek (het tijdig beschikken over data) en om de relevantie van de beschikbare data. Het is een complex probleem omdat we het zowel hebben over personen (datamanagers, data-kwaliteitsmanagers, data owners, data stewards), als over processen (data life cycle: van dataproductie naar informatie-consumptie), uitgangspunten (data management principles), metadata, datastatistieken (data metrics) en de ondersteuning daarvan met softwareproducten (technologie). Datakwaliteit gaat de hele organisatie aan en heeft betrekking op alle organisatorische lagen en geledingen. Ga er maar aan staan. Organisaties geven aan datakwaliteit van het hoogste belang te vinden. Echter, uit onderzoeken blijkt dat het met de data-kwaliteit slecht is gesteld. Volgens Gartner is tot 2007 zo'n 25 procent van de kritieke data van Fortune 1000-organisaties onjuist. The Data Warehouse Institute stelt dat in een typisch datawarehouse de datakwaliteit maandelijks met 3 procent afneemt. Andere onderzoeken laten vergelijkbare resultaten zien. Uit het vorig jaar gehouden Nationale Datawarehouse Onderzoek (DWO 2005) blijkt dat 82 procent van de organisaties *data-eigenaren* heeft aangesteld. Echter, aan de invulling van die rol scheelt nog het een of ander. Wie doorvraagt naar hetgeen gebeurt indien de datakwaliteit in het betreffende proces achterblijft, stuit op een beklemmende stilte. En dat is de regel – niet de uitzondering.

Er is echter goed nieuws. Compliance-regels zoals Sarbanes-Oxley, Basel II en de Patriot Act schrijven voor dat informatie traceerbaar en reproduceerbaar moet zijn. Sterker: chief executive officers en chief information officers moeten een dure eed zweren op de juistheid van de informatie die ze

rapporteren. Voor wie door de mand valt wacht de cel (ga niet door Start) of dreigt een aanzienlijke persoonlijke boete (geen optie om te kiezen voor een Kanskaart).

Waar organisaties in de afgelopen decennia de datakwaliteit niet op een acceptabel peil hebben kunnen krijgen, dwingen deze Compliance-regels om nu eens eindelijk schoon schip te maken. O.k., het gaat in eerste instantie alleen om financiële en persoonlijke data – maar het is een goede start. De ingevoerde procedures en processen, maar ook de opgedane kennis en ervaring, kunnen zonder veel moeite ook op de rest van de data worden toegepast. In de Verenigde Staten hebben organisaties zich in eerste instantie beklaagd over de extra kosten en tijd die met zo'n verbeteractie samenhangen. Inmiddels is de focus echter verschoven naar het uitbaten van deze kwalitatief betere data en zijn de klachten verstomd.

Twee andere recente ontwikkelingen vereisen zelfs een verbeterde datakwaliteit. De eerste is Performance Management: de vertaling en implementatie van bedrijfsstrategie naar alle bedrijfsgeledingen waarbij key performance indicators feedback geven over uitgevoerde acties. Handig als die feedback ook hout snijdt en er geanalyseerd kan worden wat wel en wat niet werkt. Dat Business Intelligence nu ook afdaalt naar de rest van de organisatie ('operational BI') is zelfs nog belangrijker. Voorheen beperkte de toepassing van BI zich tot managers en een handjevol analisten. Als dezen met data werken die onvolledig, onjuist, niet aanwezig, te laat of multi-interpretabel zijn, levert dat al behoorlijke hoofdpijn op. Verspreid je dit naar de rest van de organisatie, dan heb je met een heuse tsunami te maken. Eerst dus datakwaliteit verbeteren, dan pas breder verspreiden. Softwareleveranciers als Business Objects en Informatica kopen momenteel datakwaliteitsbedrijven als FirstLogic en Similarity Systems op. Ook IBM's overname van Ascential is bedoeld om de data-integratie- en dus ook de datakwaliteitscomponent van WebSphere te versterken. Datakwaliteit is goede business! Voor *pure play* leveranciers van datakwaliteitssoftware zoals Trillium, Human Inference en DataFlux (SAS) is dit uiteraard niets nieuws. Wat wel nieuw is, is dat datakwaliteit sexy dreigt te worden. Wie nu de job van data quality officer krijgt aangeboden hoeft niet meteen te denken dat het de 'kiss of death' is. Het zou je carrière zelfs verder kunnen helpen ... De tijd om datakwaliteit nu eens serieus op te pakken lijkt eindelijk aangebroken.

Paul van der Linden (Paul.PFH.vanderLinden@AtosOrigin.com) is senior consultant Data Warehousing/BI bij Atos Origin en geeft leiding aan Data Warehousing Cost & Lifecycle Management (CLM).