

VAN DER LANS

Metadata, kind van de rekening?



Het laatste jaar van mijn studie (ongeveer dertig jaar geleden!) bestond voornamelijk uit een lange stage en het schrijven van een scriptie. Om enig idee te hebben van wat de bedoeling van de scriptie was, toonde de docent ons enkele voorbeelden van reeds afgestudeerde studenten. Bij een van die scripties stond hij lang stil. Volgens hem was dit de perfecte scriptie. Het was geschreven door twee studenten die het jaar ervoor afgestudeerd waren. Deze scriptie ging over het belang van Data Dictionary/Directory Systemen. Het ging dus over metadata. In dat stagejaar werd mij duidelijk dat metadata ten eerste op dat moment een hip onderwerp was en ten tweede een buitengewoon belangrijk onderwerp. Iedereen in de IT-sector zag het belang ervan in. En nog steeds is metadata essentieel. Weinigen in de IT-sector zullen dat ontkennen.

Drie jaar geleden ontwierp ik voor Database Magazine een grote poster van ongeveer een vierkante meter met de titel *Het Business Intelligence Framework*. In deze poster werd enigszins op schematische wijze weergegeven hoe data van de bronsystemen via veel producten en technologieën terecht komen in diverse soorten rapporten en voor analyse beschikbaar komen. Tevens werden de verbanden tussen verschillende termen aangegeven.

Deze zomer ben ik begonnen met een vernieuwde versie van deze poster. Na enkele weken stoeien, besloot ik de in te wijzigen. In de nieuwe poster worden nu alleen maar technologieën en producten opgenomen die mogelijk door leveranciers geboden zouden kunnen worden om een datawarehouse-omgeving op te bouwen. Dus zaken als methodes en aanpakken komen er niet meer in voor. Vandaar dat de naam enigszins aangepast is: *Het Business Intelligence Technology Framework*. Het voordeel van deze gewijzigde aanpak is dat het eenvoudiger wordt om de BI-stack van een leverancier in kaart te brengen, ofwel het staat toe om de volledigheid van een leverancier's BI-stack te controleren. Tevens kunnen we de BI-stacks van leveranciers hiermee vergelijken.

Ondertussen is voor enkele leveranciers gekeken in hoeverre ze alle producten in het BITF ondersteunen. Wat dan direct in het oog springt is dat leveranciers als IBM, Microsoft en Oracle, die zeer complete BI-stacks aanbieden, wat betreft het geïntegreerd beheren van metadata matig scoren. Uiteraard slaan al hun producten, dus hun ETL-producten, rapportageproducten, database servers en datakwaliteitproducten, ergens metadata

op. Echter, er is geen product of module waarmee al die metadata geïntegreerd getoond en geïntegreerd bestudeerd kunnen worden.

Willen wij dus een geïntegreerd beeld krijgen van alle metadata, dan zullen we zelf metadata bij elkaar moeten brengen. In feite betekent dit dat we eenzelfde exercitie moeten ondernemen als die we voor onze data moeten uitvoeren. Naast integratie van data krijgen we dan ook integratie van metadata. Doen we dit niet en kunnen we dit niet, dan worden impactanalyse en lineage erg lastig.

Let wel, het geldt niet voor alle leveranciers. Als we bijvoorbeeld de BI-stacks van SAS of Information Builders bestuderen, zien we dat die veel meer de metadata van al hun producten centraal opslaan, met alle voordelen van dien. Uiteraard hebben SAS en Information Builders het voordeel dat ze bijna alle software zelf gebouwd hebben. Terwijl zeker IBM en Oracle erg veel modules gekocht hebben en het kost een leverancier uiteraard tijd om die serieus te integreren met de reeds bestaande producten.

Het blijft vreemd dat we nu al minimaal dertig jaar weten hoe essentieel metadata zijn en dat het belangrijk is dat deze ook centraal beheerd worden, dat we nu al dertig jaar bezig zijn om metadata goed georganiseerd te krijgen, maar toch moeten we concluderen dat het zelfs bij relatief nieuwe producten nog steeds niet ideaal geïmplementeerd is. Waarom blijft het onderwerp metadata toch altijd het kind van de rekening? Door de jaren heen zijn er al legio standaarden beschikbaar gekomen om metadata te registreren en uit te wisselen, maar ook dat heeft niet erg geholpen.

Voor veel klanten zou het beter zijn als leveranciers hun aandacht tijdelijk zouden verschuiven. In plaats van hun portfolio uit te breiden met zelfontwikkelde producten en met producten van andere leveranciers, zouden ze zich moeten richten op het ontwikkelen van een geïntegreerde metadata-oplossing. Hierbij zouden alle metadata geïntegreerd bestudeerd en geanalyseerd moeten kunnen worden. Hoe ze dat doen is van ondergeschikt belang, als ze het maar doen. Het zou voor gebruikers het leven een stuk eenvoudiger maken.

Rick van der Lans is zelfstandig IT-consultant.