

Op weg naar het felbegeerde totaaloverzicht van de gegevenshuishouding

“Metabase” als aanjager en kenniscentrum

Sandra Gerrits

Zou het niet geweldig zijn als elk geautomatiseerd systeem als vanzelfsprekend aansluit bij een goed doorzicht en goed bijgehouden metadatasysteem? Dat wij ons geen zorgen hoeven te maken over de beheerbaarheid? Dat alles zorgvuldig is vastgelegd van functionaliteit tot techniek? Helaas is het hebben van een “metabase” als ‘motor en kennisbase’ van informatiesystemen vaak nog een utopie.

Een *metabase* als ‘motor en kennisbase’ van informatiesystemen? Waar praten we over? Ontwikkelaars, beheerders en ontwerpers zijn al tevreden als er goed bijgehouden functionele en technische documentatie over een systeem bestaat! Nieuwe medewerkers staan echter nog geregeld voor de noodzaak programmacode te ontcijferen op zoek naar de informatie die zij nodig hebben. Voor databases hebben de Case-tools niet voor niets reverse engineering ingebouwd. Hierdoor is het in ieder geval mogelijk het meest actuele databasemodel inzichtelijk te maken. Overige informatie over

systemen, data en metadata ligt her en der in de organisatie verspreid. Hoewel ervaren functioneel beheerders en ‘super-users’ op de werkvloer veel waard zijn, beperkt hun kennis zich meestal tot de ins-and-outs van slechts één of enkele systemen.

Een totaaloverzicht maken van de gegevenshuishouding binnen een organisatie is moeilijk en complex. Bovendien beperkt de problematiek zich niet tot de gegevens die een systeem gebruikt en bewerkt.

Vergeleken met enkele jaren geleden worden andere eisen gesteld aan nieuw te ontwikkelen programmatuur. De kosten moeten omlaag; het gaat om return on investment. Dat betekent dat de flexibiliteit groter moet worden. Ook verwacht men snel resultaat en daarmee korte ontwikkeltijden. Het liefst koopt men een standaardpakket met eventueel wat modificaties. De wandelende metabase, de medewerker die ‘vergroeid’ is met het systeem, is bijna uitgestorven.

Gebruik van metadata kan hierop het antwoord zijn, mits de opzet van de metadata-oplossing voldoet aan een aantal voorwaarden. Een van de belangrijkste is dat de metagegevens toegankelijk zijn voor zowel functionele als technische gebruikers en systeemgebruikers. In principe geldt echter dat wat voor de één data zijn, metadata kunnen zijn voor de ander. Een pittig uitgangspunt voor de opzet en implementatie van een metabase.

Metadatatoepassingen vandaag

In een vorig artikel van Sandra Gerrits (DB/M 5, september 2001) is het concept van metadata behandeld aan de hand van een balanced scorecard-implementatie. In dit artikel gaat Gerrits in op de inrichting van een *metabase*. Aan de hand van de implementatie van een datawarehouse geeft zij een voorbeeld van de huidige beschikbaarheid van metadatatoepassingen. Metamodellering blijft in dit artikel buiten beschouwing. In het eerste deel kijkt de auteur naar de voorwaarden waaraan metadata moeten voldoen. Daarna worden de architectuurmogelijkheden van een metabase behandeld. Vervolgens wordt een voorbeeld beschreven van een metadata-oplossing zoals men die nu vaak implementeert bij de opzet van een datawarehouse. Het artikel eindigt met veel voorkomende valkuilen voor het mislukken van een metadatatoepassing en de uitdagingen waarvoor een projectteam zich gesteld ziet.

VOORWAARDEN METADATA

Voor metadata gelden dezelfde regels als voor overige data. De gegevens moeten identificeerbaar, definieerbaar, traceerbaar, toegankelijk en consistent zijn.

Identificatie

Identificeerbaar houdt in dat een gebruiker weet dat de metadata bestaan, maar nog belangrijker: dat hij weet hoe hij die gegevens kan benaderen en die informatie zal vinden die hij zoekt. Daarmee trappen we een open deur in, maar in de praktijk gaat het zelden goed. Deze voorwaarde heeft verstrekkende gevolgen voor de implementatie. Bekend moet zijn:

- wie de gebruikers van de metadata zijn. Behalve menselijke gebruikers, op technisch en functioneel gebied, zijn dat de

applicaties, databasesystemen en tools;

- wat de gebruikersgroepen aan metadata nodig hebben. Uitwerking van deze analyses zal leiden tot gebruikersgroepen die elk hun specifieke aandachtsgebied hebben, technisch of functioneel. Door meerdere gebruikersgroepen gebruikte metadata zullen anders worden behandeld dan die welke voor slechts één gebruikersgroep van belang zijn, zoals systeem-eigen metadata, die geen relatie hebben met en niet gebruikt worden door anderssoortige systemen.

Definitie

Dat metadata te definiëren zijn, houdt in dat de betekenis van een term of waarde bekend en algemeen aanvaard is in de organisatie.

Neem applicatie A. Daarin wordt gebruik gemaakt van de variabele T_SAL_KST (totale salariskosten). De applicatie krijgt haar gegevens uit bron X, waarin dezelfde waarde wordt aangeduid met SKST. Bij de bouw van A, enkele jaren geleden, is nagegaan of T_SAL_KST en SKST dezelfde zijn. Nu wordt bron X vervangen, omdat het operationele systeem dat hem vulde wordt vervangen door een standaardpakket. Maar daarin worden salariskos-

“Similar to the way we can automatically back-up and load, we should be able to locate and clarify” - Adrienne Tannenbaum

ten verder verdeeld; er is onderscheid tussen bruto salariskosten, vergoedingen en kosten voor de verschillende secundaire arbeidsvoorwaarden. Tevens zijn kosten voor externe medewerkers, zoals consultants en uitzendkrachten, apart opgenomen.

Is een optelling van al deze waarden gelijk aan T_SAL_KST? Voor het antwoord op deze vraag moet vaak gezocht worden in het oorspronkelijke systeem waarmee SKST werd gevuld. De uitkomst blijkt niet altijd te zijn wat men had verwacht.

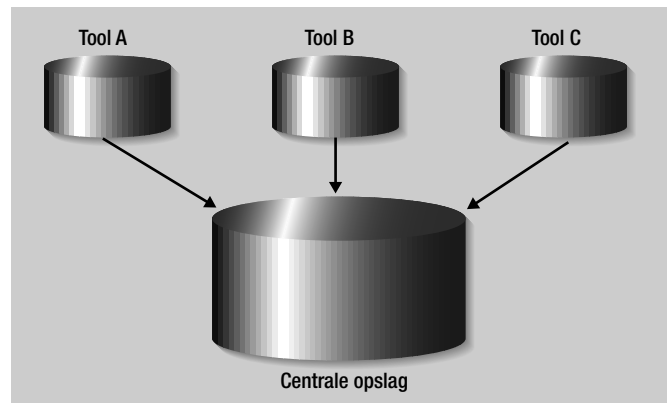
De algemene definities moeten door de organisatie worden gedragen. Dat geldt ook voor het beheer, technisch én functioneel. Inzichtelijk maken van deze definities is vaak een traject op zichzelf.

Tracering

Onder traceerbaarheid wordt verstaan dat te achterhalen is waar de gegevens vandaan komen en waarvoor ze gebruikt worden. Niet altijd is transparant welke systemen met elkaar te maken hebben. Afhankelijkheden zijn zelden volledig in kaart gebracht. Soms gaan data door drie of vier systemen voordat ze de eindgebruiker bereiken. Bovendien worden de gegevens tijdens dit transport bewerkt. Onder traceerbaarheid vallen tevens zaken als laatste mutatedatum en de kwaliteit van de data.

Toegankelijkheid

Metadata zijn toegankelijk als een gebruiker ze makkelijk kan benaderen. Catalogisering en het kunnen zoeken op vrije tekst, naam en waarde kunnen helpen de gewenste informatie te achterhalen. Tegenwoordig worden vaak portals gebruikt om metadata



FIGUUR 1: CENTRALE OPSLAG VAN METADATA.

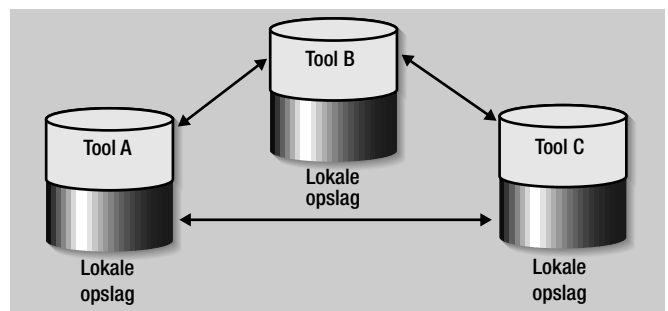
beschikbaar te stellen. Webtechnologie kan toegankelijkheid tot metadatagegevens sterk stimuleren. Het intranet van een organisatie leent zich uitstekend voor laagdrempelige toegang tot de metadata-toepassing; gebruikers zijn vertrouwd met format en zoekmogelijkheden.

Legacy-termen en -naamgevingen kunnen opgenomen worden als een gebruiker op zoek is naar informatie. Zodra een gebruiker op naam gaat zoeken, is automatische terugkoppeling gewenst. Op deze manier zijn gangbare synoniemen inzichtelijk te maken. Een naar klantinformatie zoekende gebruiker kent KLNT als standaard benaming voor klant. Daarop zoekt hij als hij wil weten in welke bronsystemen klantinformatie is opgeslagen. Het systeem koppelt aan hem terug of hij ook wil zoeken op KL_NAAM, CUST en KLANT. Deze oude naamgevingen worden nog in enkele systemen gebruikt. Zonder deze terugkoppeling zou deze gebruiker een gedeelte van de voor hem belangrijke informatie missen. Dit samenbrengen van gestandaardiseerde en bestaande naamgevingen moet niet worden onderschat. Het kan een belangrijke voorwaarde zijn van acceptatie van de opgeleverde metadata-toepassing.

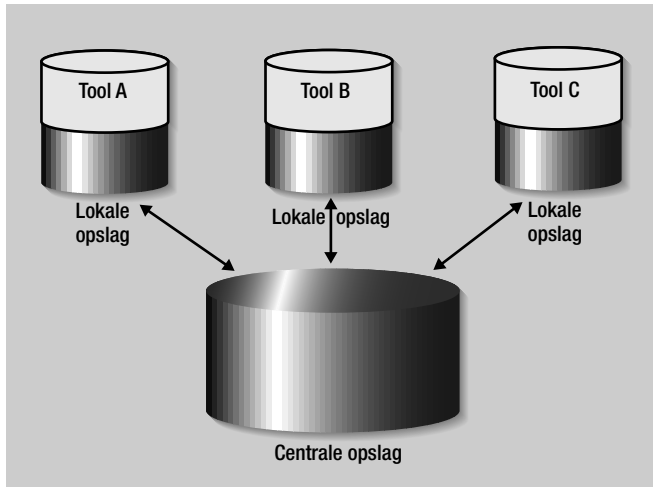
Toegankelijkheid heeft ook invloed op de eventuele autorisaties van de metadata. Bepaald moet worden wie toegang heeft tot welke meta-informatie.

Consistentie

Onder consistentie wordt verstaan dat gegevens elkaar niet mogen tegenspreken. In een callcenter wordt klanten bij inbellen gevraagd hun klantnummer in te toetsen. Zodra een klant wordt doorverbonden, staan zijn gegevens van op het scherm van de



FIGUUR 2: GEDISTRIBUEERDE OPSLAG VAN METADATA.



FIGUUR 3: GEDEELDE OPSLAG VAN METADATA.

medewerker: Jansen, 100134. De in een CRM-systeem opgeslagen data moeten consistent zijn met klantgegevens in de overige productiesystemen. Bij het metamodelleren zullen de modellen en definities op elkaar moeten worden afgestemd. Dit geldt ook voor metabasesystemen, vooral die met een gedistribueerd karakter.

ARCHITECTUUR

Alle voorwaarden voor een metadatatoe passing komen bij elkaar in de architectuur. Een goede architectuur bevat onder meer de

manier van opslag, maar ook applicaties voor beheer en ontsluiting. Daarnaast vormen procedures voor metadatabeheer een belangrijk onderdeel van de totale oplossing. Natuurlijk is het onbegonnen werk meteen alle bronnen van metadata fysiek te ontsluiten. Bovendien kennen veel metadata in de praktijk nog geen opslagvorm: een *company data dictionary*, met standaard termen voor de organisatie en de bijbehorende definities en synoniemen, is veelal niet beschikbaar. Dergelijke 'nieuwe' en vaak niet technische metagegevens zullen een plaats moeten krijgen in de architectuur, die dan ook open moet zijn, met ruimte voor toekomstige uitbreidingen. Verder is een inventarisatie van alle metadatatagerelateerde systemen en de gebruikers van de metadata daaruit van groot belang. Flexibiliteit van de opslagstructuur en de metadata-modellen is essentieel. Organisaties veranderen, data veranderen; dus ook metadata. Naast de opslag zal men faciliteiten moeten creëren voor het beheer van de metadata.

Opslag van metadata in een architectuur kan op een aantal manieren, elk daarvan met zijn voor- en nadelen: centraal, gedistribueerd of gedeeld (shared).

CENTRALE OPSLAG

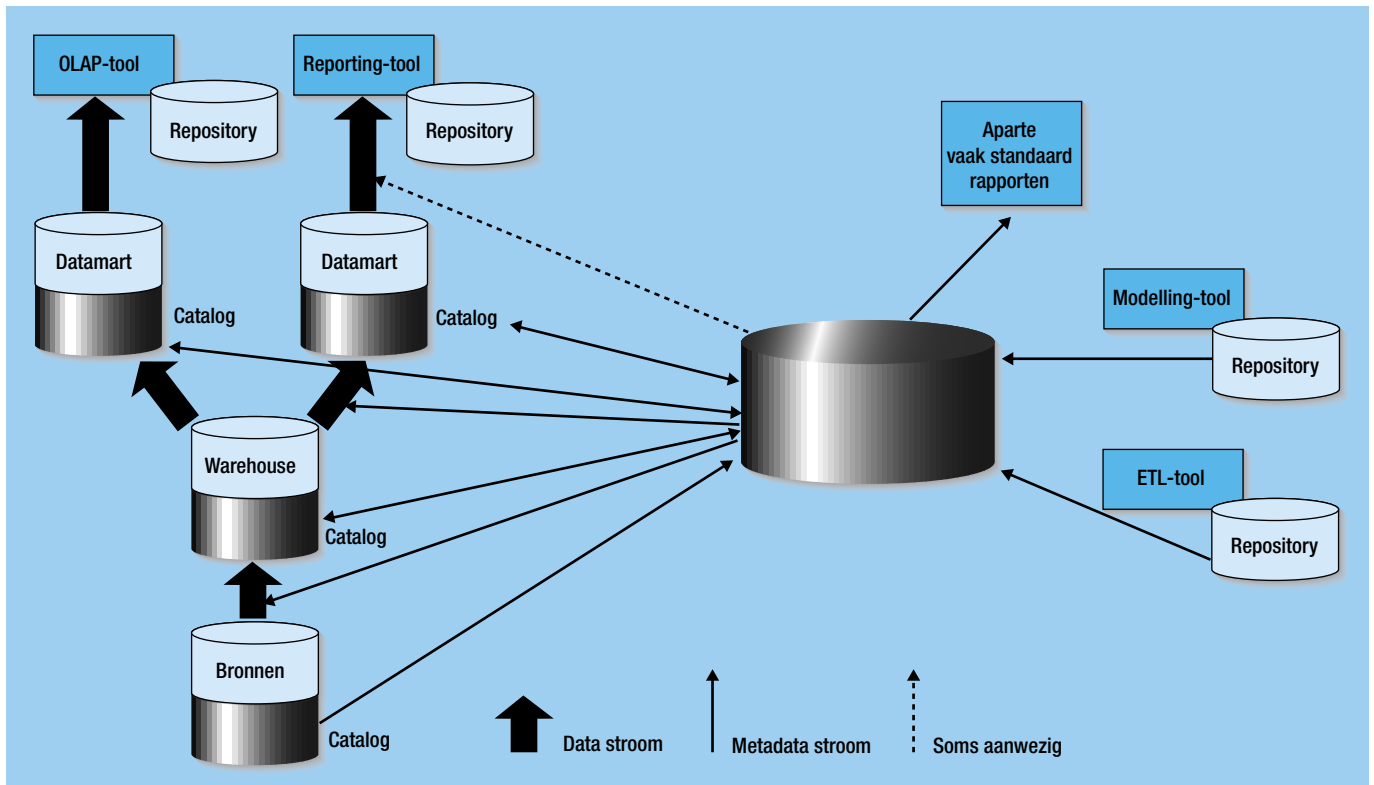
Bij centrale opslag van de metadata (zie figuur 1) gebruiken alle databases, warehouses, applicaties, tools en eindgebruikers deze repository of dbms met metadata. Lokale opslag van metadata vindt niet plaats. Ook de systeemspecifieke metadata worden centraal opgeslagen en beheerd. Metadata worden niet onderling uitgewisseld. Hierdoor is er geen metadata-replicatie, wat de consistentie bevordert. Er hoeft geen mechanisme ingericht te worden om de data te repliceren. Ook is dit ene punt van referentie gemakkelijk aan een eindgebruiker te communiceren.

Daartegenover staat dat onderhoud van systeemspecifieke metadata niet gemakkelijk is. Hiervoor zullen aparte procedures en mechanismen moeten worden opgezet. Ook worden alle systemen aangestuurd vanuit deze ene database of repository. Is deze opslag niet beschikbaar, dan zullen veel functionaliteiten of zelfs complete systemen niet werken. Dit betekent dat de betrokken medewerkers overweg moeten kunnen met aansturing vanuit een centrale repository. Performance is een belangrijk aandachtspunt, omdat metadata voor de systemen zelf lastig bereikbaar zijn.

Voor een organisatie met een klein aantal systemen kan centrale metadata-opslag een goede keuze zijn. Gebruik van veel verschillende systemen maakt deze opzet onbeheerbaar.

GEDISTRIBUEERD

Bij gedistribueerde metadata-opslag (zie figuur 2) worden de onderlinge repository's en catalogs met elkaar verbonden, zodat zij gebruik kunnen maken van elkaars metadata. Deze vorm heeft zo zijn voordelen. De verschillende tools, applicaties en dbms'en blijven op deze manier maximaal onafhankelijk. Hierdoor is de flexibiliteit binnen één tool optimaal. Of dit op lange termijn ook voor-



FIGUUR 4: METADATA OPGEBOUWD NAAST EEN DATAWAREHOUSE.

delig is, hangt helemaal af van de uiteindelijke doelstelling van de metadatatoeëpassing. Ook zijn de lokale metadata sneller te benaderen dan centrale. Dit kan van belang zijn voor de performance van deze afzonderlijke systemen.

Nadelig in deze opzet zijn de vele connecties tussen de verschillende repository's, die alle moeten worden onderhouden. Bewaking van de synchronisatie en de beheerbaarheid van de

tussen de systemen en de centrale opslag blijven nodig, maar in een behoorlijk kleiner aantal dan bij een gedistribueerde oplossing. In de centrale opslag kunnen ook de company data dictionary en overige meer functioneel gerichte metadata opgenomen worden.

Een metadatatoeëpassing is een datawarehouse op zichzelf, met alle bouweisen die dat met zich meebrengt

metadata kunnen hierdoor voor problemen zorgen. Bovendien zijn extra goede voorzieningen noodzakelijk om de zoektocht naar informatie voor een eindgebruiker te faciliteren.

GEDEELD

Bij de gedeelde opslag van metadata (zie figuur 3) komt neer op een samenvoëging van de voordelen van beide andere oplossingen. Hierdoor is het mogelijk de algemene voor gebruikers(groepen) belangrijke metadata centraal op te slaan en toch de performance van tools op peil te houden. Ook zijn de toolspecifieke repository's en dbms-catalogs onafhankelijk, met alle voordelen van dien: performance, flexibiliteit binnen het tool en meer. Connecties

PRAKTIJKVOORBEELD

Hoe gaan bedrijven op dit moment met metadata om? Hoever zijn zij met de implementatie? En wordt hierbij voldaan aan de voorwaarden die we noemden? Hoe ziet de architectuur eruit?

Tegenwoordig hebben veel bedrijven al één of zelfs meer datawarehouses. Met de opbouw daarvan proberen zij ook de metadata een plaats te geven. Vaak is dat een plek naast het warehouse (zie figuur 4). Modellen, rekenregels en batchprocessen voor het laden van het datawarehouse en eventuele datamarts zijn de eerste onderdelen van het nieuwe metadatasysteem. In een aantal gevallen zijn ook de metadata uit de bronsystemen erin betrokken. Deze situatie is tamelijk gangbaar.

Een metadatagestuurde frontend blijkt al lastiger. Maar in een volledige metadata-architectuur horen ook de applicaties en systemen van de frontend thuis. Onduidelijk is verder vaak de opzet van de opslag van de metadata: centraal, gedistribueerd, complete redundantie of een combinatie van deze mogelijkheden.

De technische metadata van deze toëpassing blijven doorgaans het meest actueel, vooral doordat de laadprocessen metadatagesturd zijn. Dit maakt dat ook de modellen en rekenregels frequent worden bijgewerkt. De inhoud van de metabase is nog voornamelijk beperkt tot de technische metadata, de gebruikers zijn met

Mislukkingen

Datawarehouses hebben metadata sterk in de belangstelling gebracht. Bij de bouw zelf wordt vaak gezocht naar metadata. Door die eindelijk eens goed vast te leggen, blijft de technische en functionele gebruikers in de toekomst een frustrerende zoektocht bespaard. Er wordt veel aan gedaan om metadata voor de verschillende eindgebruikers zowel duidelijk als onderhoudbaar beschikbaar te maken. Toch komen veel van dit soort initiatieven niet van de grond, doordat:

- de gebruikersgroepen niet of onvoldoende zijn geïdentificeerd. Daardoor sluit de oplossing niet of niet goed aan bij de behoeften van de gebruiker;
- er geen overkoepelende opzet en architectuur voor de metadatatoepassing bestaat. Men modelleert wel voor de toepassing die samenhangt met het datawarehouse, maar niet met alle overige metadata in de organisatie, zoals andere warehouses,

name de technische gebruikers.

Rapportages over de metadata staan vaak nog los van de functionele gebruiker. Begrijpelijk, omdat het metadatasysteem voor deze gebruiker nog weinig informatie bevat. Verder sluit de informatie die opgevraagd kan worden uit de metabase veelal niet aan bij de belevingswereld van de gebruikers.

Omdat we OLAP- en reportingtool buiten beschouwing laten in dit voorbeeld, zijn de aanwezige gegevens maar voor een deel

Gebruik van metadata kan het antwoord zijn op het uitsterven van de wandelende metabase, mits wordt voldaan aan een aantal voorwaarden

identificeerbaar en traceerbaar. De metadata worden toegankelijk gemaakt door middel van (standaard) rapporten. Uit dit voorbeeld is niet te achterhalen of wordt voldaan aan de eisen van definitie en consistentie.

Bredere implementaties van een metadatasysteem zijn zeldzaam. Maar wat als een tweede datawarehouse wordt neergezet? En wat gebeurt er rond de al bestaande applicaties en eventuele nieuwe? Deze zullen niet allemaal gebruik maken van het datawarehouse. Omdat de focus zo duidelijk alleen daarop is gericht, is de draagkracht van het systeem beperkt. Bouw of uitbreiding van een datawarehouse kan echter een goede katalysator zijn voor het inrichten van een metadata-toepassing.

UITDAGINGEN

Veel uitdagingen houden rechtstreeks verband met de oorzaken waarom metadata-toepassingen geen -of slechts ten dele- toegevoegde waarde voor de organisatie hebben. Niet meteen met de

operational datastores en operationele applicaties. Het overkoepelende beeld is niet aanwezig;

- functionele metadata niet of niet voldoende meegenomen zijn in de opzet. Veel tools en implementaties zijn nog technisch georiënteerd, waardoor de functionele gebruikers niet of weinig gebruik maken van de beschikbare metadata. Hierdoor zijn draagkracht en acceptatie beperkt;
- de oplossing niet of niet goed bekend is bij de gebruikers (technisch en functioneel);
- goede en gemakkelijke toegang tot de metadata ontbreekt;
- geen duidelijke verantwoordelijke voor of eigenaar van de metadata is benoemd;
- opslag en beheer vaak worden gezien als 'bijproduct' van een datawarehouse-implementatie in plaats van aparte implementatie door de organisatie heen. Hierdoor zijn ze vaak sluitstuk op de begroting;
- een duidelijke sponsor ontbreekt.

implementatie en evenmin na een paar jaar (zie kader *Mislukkingen*). In elk geval vragen de volgende zaken de aandacht:

- synchronisatie en standaardisatie van definities, naamgevingen (bijvoorbeeld gebruik van klant_id in nieuwe applicaties in plaats van weer een variant);
- integratie van de 'legacy'-metadata in de gekozen oplossing;
- toegankelijk maken en vooral ook houden van de metadata voor alle gebruikers, dus zowel de technische en functionele gebruikers als de tools en applicaties;
- goed opnemen van historie en versiebeheer;
- opnemen van 'losse', ongestructureerde informatie, zoals Word-documenten en webpagina's;
- bewustwording bij de organisatie dat het implementeren van een metadata-toepassing een apart project is, dat losstaat van de implementatie van een datawarehouse;
- verkrijgen van een sponsor voor het project.

Kortom, een metadata-toepassing is een datawarehouse op zichzelf, met alle bouwweisen die dat met zich meebrengt. Het wordt tijd dat we dit inzien en metadata-implementaties als zodanig gaan behandelen. ●

Referenties:

- Metadata Solutions: Using Metamodels, Repositories, XML, and Enterprise Portals to Generate Information on Demand*, Adrienne Tannenbaum. Uitg.: Addison-Wesley, 2002. ISBN 0-201-71976-2
- On metadata interoperability in datawarehouses*, Hong Hai Do, Erhard Rahm. Technischer Report I-2000, Institut für Informatik, Universität Leipzig, 2000; <http://dol.uni-leipzig.de/pub/2000-13>

Sandra Gerrits (sandra.gerrits@eniklas.com) is consultant bij eNiklas Netherlands, een in business intelligence gespecialiseerde ICT-dienstverlener.