



Integratie voor een flink deel nog tool- en technology-driven

Metadata: The Final Frontier

Martin Misseyer en Karel Auwerda

Wat op papier één van de makkelijkste onderwerpen in de IT-wereld zou moeten zijn, blijkt in de praktijk doorgaans één van de moeilijkste: metadata. Dit artikel is een aanzet om volop aandacht te geven aan metadata, en beschrijft enkele belangrijke ervaringen en ontwikkelingen.

Vanaf de geboorte van de eerste computer-programma's, databases en informatiesystemen zouden data en metadata hand in hand moeten gaan. Het onderwerp metadata zou moeten schitteren door aanwezigheid, het zou de rol moeten vervullen van baken, niet de rol van duistere kracht. In de praktijk worstelen helaas veel organisaties nog steeds met metadata.

Tussen de oren

De moraal van het verhaal luidt, dat metadata (management) 'tussen de oren moet zitten', en niet kan worden afgekocht met 'tooling' of kan worden weggestopt in een 'repository'. Het onderwerp metadata begint bij het 'begrip' en de 'definitie', en eindigt bij het 'gebruik' en de 'toepassing'. Metadata management houdt onder andere in dat administratie van metadata eenvoudig en eenduidig is en, bij voorkeur, op één plaats gebeurt. Metadata management houdt ook in dat niet alleen syntax wordt vastgelegd, maar ook semantiek.

Na beschouwing van allerlei recente metadata-gerelateerde ontwikkelingen (van SOX en IRFS aan de business-zijde tot en met SOA, ontologieën en semantic web aan de IT-zijde) concluderen wij dat veel organisaties feitelijk opnieuw moeten leren omgaan met metadata.

Babylon

Systeemontwikkeling heeft geleid tot generalisatie (per type modellering gebruik maken van eenzelfde hulpmiddel) en tegelijkertijd specialisatie (meerdere typen modellering onderscheiden en koppelen aan verschillende, gescheiden functies). Waar in het verleden informatiesystemen werden ontwikkeld met een 'monolithisch' product (CASE-tool, workbench), is er steeds meer plaats gemaakt voor modernere ontwikkelstraten die ofwel zijn gebaseerd op technologie van één software-leverancier, ofwel zijn samengesteld uit hulpmiddelen afkomstig van meerdere

software-leveranciers. Naarmate er meer hulpmiddelen worden gebruikt neemt de rol van metadata-integratie toe. Immers, verschillende hulpmiddelen delen meer of minder metadata. Hierbij is het niet de bedoeling om deze meerdere keren te administreren. Delen betekent in deze context: zo efficiënt en effectief mogelijk. Hoewel men zou denken dat door het aantal verschillende te gebruiken hulpmiddelen er sprake kan zijn van een Babylonische spraakverwarring (dat wil zeggen meerdere hulpmiddelen, hulpmiddelspecifieke formaten, beperkte interfaces tussen de hulpmiddelen), is dit in de praktijk meegevallen. Ten eerste komt dit doordat er een gezonde weerstand bestaat om maximaal te specialiseren (niet alle hulpmiddelen gebruiken, maar een beperkt aantal). Ten tweede komt dit doordat er de laatste jaren bijzonder veel inspanning is geweest op het gebied van metadata-integratie (onder andere CWM, MOF) en -uitwisseling (zoals XMI, XML, UML). Het voorgaande betreft in eerste instantie de integratie van de technologie of 'onderzijde'. Hoewel metadata via deze weg zijn te integreren, is de vraag in welke mate dit functioneel is. Ten derde zijn er de afgelopen jaren ook specifieke metadata-oplossingen op de markt gekomen, die verregaande metadata-integratie faciliteren.

De keerzijde is dat de metadata-integratie voor een flink deel tool- of technology-driven is. Ondanks standaarden en generieke interfaces maakt dit een en ander complex en ondoorzichtig, en zijn zowel de initiële- als de onderhoudskosten aanzienlijk (zowel voor de hulpmiddelen als voor de inspanning). Vandaar dat metadata-technologieën nog niet op grote schaal in het bedrijfsleven worden toegepast, maar dat – nog steeds – Word en Excel fungeren als 'prima' metadata documentatie-tools, hetgeen de weg naar Babylon nog steeds open houdt. 'Echte' metadata-integratie zou aan de 'bovenkant' moeten plaatsvinden. Data over business-definities, -begrippen en logische

Niveau	Hulpmiddel	Ontwikkeling	Gebruik en Beheer
Programmeren 'in code'	Programmeertaal	In beperkte mate worden de functioneel benodigde metadata handmatig vastgelegd.	Er worden nauwelijks gebruik- en beheergerichte metadata vastgelegd. Autorisatiemodel zit niet in het informatiesysteem.
Genereren 'in repository en in runtime'	Generatiehulpmiddel 'CASE-tool'	Alle functioneel benodigde metadata worden in een ontwikkelomgeving gespecificeerd. Deze is met name datagericht en beperkt procesgericht.	Hoewel er meer aandacht kan worden besteed aan logging, is de vastlegging van gebruik- en beheergerichte metadata afhankelijk van het toegepaste product en de ondersteunde én geactiveerde functionaliteit. Autorisatiemodel zit niet in het informatiesysteem.
Virtualiseren 'in (metadata) repository'	Engine 'Business Rules' 'Transformation'	Alle functioneel benodigde metadata worden in het hulpmiddel gespecificeerd. Metadata is nu zowel data- als procesgericht.	Gebruik- en beheergerichte metadata worden in belangrijke mate meegenomen. Dit omdat het doorgaans een 'all-in-one' product is, waarbij er impliciet allerlei zaken worden gelogd, hergebruikt, e.d. Autorisatiemodel zit in het informatie systeem.
Creëren 'in (kennis-) repository'	Expertsysteem (shell)	Alle functioneel benodigde metadata worden in het hulpmiddel gespecificeerd. Metadata zijn zowel data- als procesgericht. Nota bene: met procesgericht wordt bedoeld dat er bijvoorbeeld kennisregels worden ingebracht die het proces behelzen en/of kunnen beïnvloeden (gewicht, voorwaarden, e.d.)	Idem als het voorgaande, met dien verstande dat er meer aandacht kan zijn voor het 'creatieve' oftewel kennisgedeelte – het inferentiemechanisme – dan voor de rest. Autorisatiemodel zit in het informatiesysteem.

Afbeelding 1: Soorten primaire hulpmiddelen voor ontwikkeling informatiesystemen en metadata.

modellen (informatie, proces, data) dienen te worden beheerd en onderhouden door een afdeling 'die er verstand van heeft'. Oftewel informatie- en procesdeskundigen zouden metadata management als zelfstandige (deel)activiteit hebben en moeten er op toe kunnen zien dat metadata op de juiste plaats worden geadministreerd, verwerkt en geraadpleegd. Hulpmiddelen zijn enablers, maar het op de juiste plaats belegde metadata management-proces is leidend.

Metadata Warehousing

Ten opzichte van primaire informatiesystemen is de ontwikkeling van datawarehouses extra bijzonder, omdat bij de ontwikkeling van datawarehouses nagenoeg altijd meerdere hulpmiddelen worden toegepast. Bovendien speelt hier nog mee dat we te maken hebben met de vertaling van functie- of bronspecifieke metadata naar een generiek model, welke op haar beurt weer is te gebruiken door de (andere?) specifieke functies of het (overall) management. Tenslotte wordt menig datawarehouse-implementatie hard geconfronteerd met de imperfecte reële wereld, waarin een bron de nodige inconsistenties ten toon kan spreiden. De integratie van verschillende (bron)data gaat daarom zelden zonder slag of stoot. Gegeven de hoeveelheden en soorten metadata die bijeen worden gebracht, kunnen we in de praktijk spreken van een *metadata warehouse*. Hiermee bedoelen we dat in het ideale geval alle relevante metadata aan elkaar kunnen worden geknoopt, zodat er een 'schaduw-datawarehouse' wordt gevormd. Ten eerste wordt

hiermee consistentie afgedwongen en is het mogelijk om data-integratie te bewerkstelligen. Ten tweede refereert het begrip metadata warehouse aan het feit dat metadata in het datawarehouse evenzo kunnen worden gebruikt om te rapporteren en te analyseren. Hierbij kan worden gedacht aan 'data lineage' of herleidbaarheid van bijvoorbeeld rapportage teruggeleid tot brondata. Bijvoorbeeld om de gebruiker te laten controleren waar hij/zij de data vandaan haalt. Daarnaast kan op basis van deze 'data lineage' worden bekeken wat de impact van wijzigingen is van het datawarehouse. Ten derde is het handig om metadata over de datakwaliteit van de opgeslagen data vast te leggen. Immers, het overgrote deel van de datawarehouse-implementaties heeft te maken met vervuilde en incomplete data.

Bijzondere aandacht vragen de plaatsen waar onder meer business-definities (van termen en begrippen) kunnen worden vastgelegd. Immers: we hebben aan de ene kant allerlei bronsystemen waarin definities kunnen zijn vastgelegd, als ook de hulpmiddelen waarmee deze bronsystemen zijn ontwikkeld en worden beheerd; aan de andere kant zijn er ook de hulpmiddelen waarmee het datawarehouse is ontwikkeld. De strekking van het verhaal is dat er per metadata-soort één plaats moet worden aangewezen waar deze metadata worden geadministreerd, bewaard en beheerd. Deze 'Single-Point-of-Definition' (SpoD), is essentieel om de kwaliteit van de metadata te behouden en te garanderen. Overigens betekent toepassen van een SpoD niet dat bepaalde

metadata maar op één plaats voorkomen; ze is bedoeld om aan te geven dat voor de betreffende metadata één bron is aan te duiden en dat de synchronisatie/distributie van daaruit wordt geregeld. Een SPoD wordt niet zomaar bereikt; een bepaalde afdeling of een informatiesysteem wil niet graag afhankelijk zijn van een andere afdeling wanneer het om het verstrekken van metadata gaat. Toch is het belangrijk dat metadata management een centrale plaats in de organisatie krijgt toebedeeld. Dit is de enige manier om metadata management 'tussen de oren te krijgen'. Tenslotte, metadata management, hoewel ze historisch gezien in een adem met datawarehousing wordt genoemd, is feitelijk iets dat niet per se thuis hoort in een datawarehouse-omgeving. Dit kan net zo goed in combinatie met de data management-functie in een organisatie. De vervolgvraag die we ons stellen is, welke ontwikkelingen kunnen een bijdrage leveren in het bereiken van metadata management en de gevraagde 'Single-Point-of-Definition'?

Ontwikkelingen

Wanneer we de laatste twee decennia beschouwen, dan zijn feitelijk drie gebieden te onderscheiden waarin belangrijke, metadata-gerelateerde ontwikkelingen hebben plaatsgevonden. Deze worden afzonderlijk behandeld.

Standaardisatie en hulpmiddelen.

Zoals eerder aangegeven, heeft de markt zelf gezorgd voor initiatieven waarin werd gestreefd naar metadata-integratie. De hieruit voortvloeiende (OMG) standaarden hebben met zich mee gebracht dat er steeds meer leveranciers hun hulpmiddelen hebben 'geopend'. Hiermee wordt het dus mogelijk dat ten eerste specifieke hulpmiddelen onderling elkaars metadata kunnen gebruiken, en ten tweede dat er nieuwe hulpmiddelen zijn ontstaan die een overkoepelende blik kunnen tonen over de verschillende soorten metadata en metadata-hulpmiddelen heen. Met dit laatste worden de metadata repository-producten bedoeld, die de laatste jaren aan belangrijkheid toenemen. Hoewel de metadata repository-producten steeds intelligenter worden, zijn deze producten wel goed in staat om metadata te verzamelen uit allerlei hulpmiddelen en informatiesystemen van een organisatie, maar bi-directioneel (publiceren van metadata) is vaak nog

tamelijk beperkt (bijvoorbeeld alleen binnen het datawarehouse) of tamelijk specifiek (bijvoorbeeld terug naar slechts één applicatie/informatiesysteem).

Dit houdt in dat de metadata management-functie, analoog SPoD, wel kan worden uitgevoerd, maar hiervoor is dan één tool set (zijnde het metadata repository-product) niet voldoende. Tenslotte dient te worden opgemerkt dat, zoals eerder betoogd, men niet zozeer naar de mogelijkheden en beperkingen van metadata-hulpmiddelen moet kijken. Metadata en metadata management vergen vooral eerst een *visie* (wat is de kijk op metadata), vervolgens een *strategie* (hoe gaan we dit bereiken) en tenslotte een *implementatieplan* (procedures, hulpmiddelen, resources, en dergelijke). Overigens bieden diverse software-leveranciers nog steeds eigen adapters, bridges en connectors om data van de ene tool in de andere te krijgen.

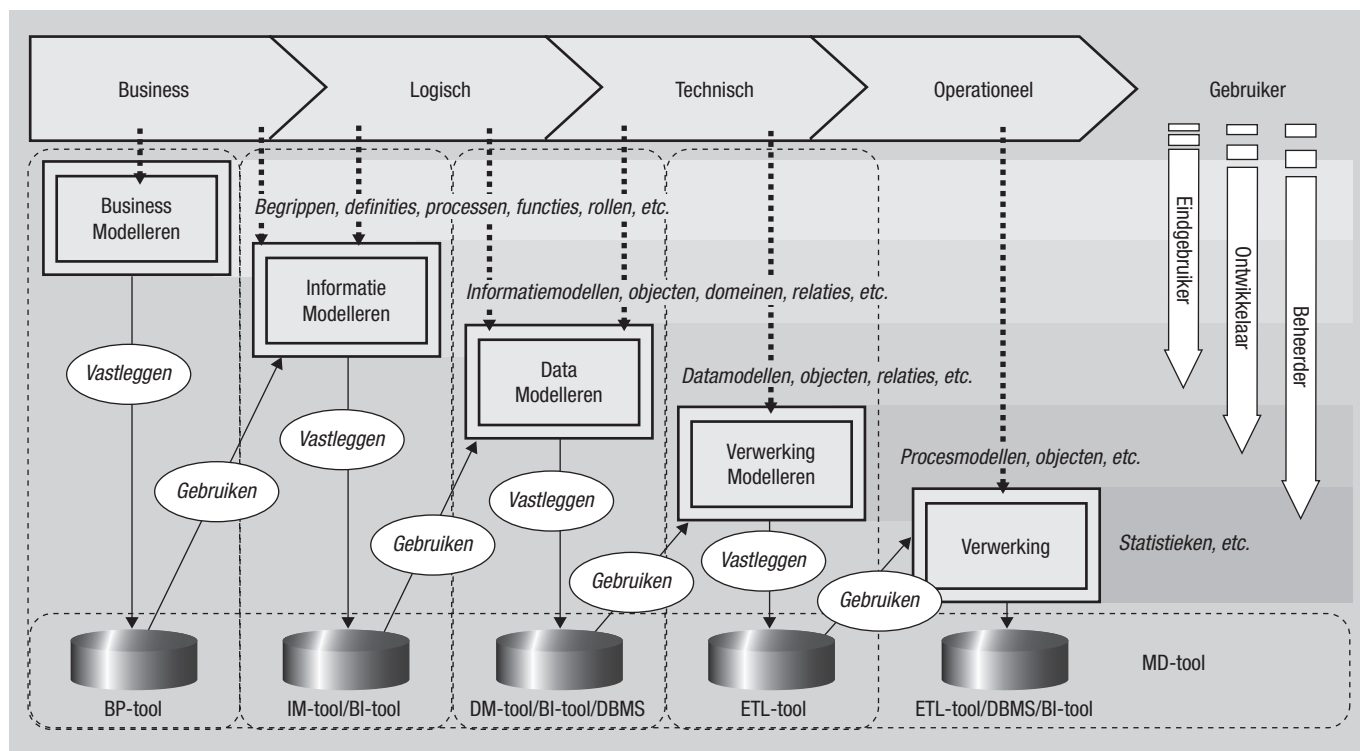
De integratie van verschillende (bron)data gaat zelden zonder slag of stoot

Verregaande verplichtingen en verantwoording van het bedrijfsleven: IFRS, Basel II, SOX en XBRL.

Elk decennium worden organisaties 'wakker' geschut en geconfronteerd met weer nieuwe ontwikkelingen, of dit nu nieuwe inzichten, kansen of bedreigingen zijn is om het even. Zo heeft men in de accounting-wereld recentelijk de handen op elkaar gekregen voor een nieuwe financiële rapportagestandaard: IFRS. In de financiële wereld is een nieuw kredietrisico-systeem ingevoerd: Basel II. En eveneens in de accounting-wereld zijn er als gevolg van talloze schandalen nieuwe systemen ingevoerd ter controle, bijvoorbeeld de Sarbanes-Oxley wet (SOX). Wat heeft dit met metadata te maken? Zeer veel: ten eerste heeft invoering van IFRS en Basel II geleid tot het aanpassen van allerlei boekhoud- en financiële rapportagesystemen, inclusief het datawarehouse. Voor Basel II is het zelfs zo dat sommige

Hulpmiddel categorie	Hulpmiddel	Voorbeeld
Business Process-modellering	Bedoeld om bedrijfsprocessen mee te modelleren. Actie en interactie, halffabrikaten en eindresultaten hierin zijn zeer belangrijk. Maar ook definitie van rollen en functiescheiding!	UML
Informatie-modellering	Bedoeld om informatiehuishouding/landschap mee te modelleren. Informatiebehoefte, -generatie, -interactie, en -presentatie worden op deze wijze gedocumenteerd. Belangrijk is ook: waar wordt de informatie gebruikt, gecreëerd, en dergelijke.	NIAM, FCO-IM
Data-modellering	Bedoeld om logische en technische datamodellen te definiëren, waarmee efficiëntie kan worden gemodelleerd in termen van functionaliteit, opslag en implementatie.	ERM
Systeem-modellering	Bedoeld om logische en technische informatiesysteem-componenten te definiëren en te modelleren.	UML

Afbeelding 2: De belangrijkste categorieën secundaire hulpmiddelen voor de ontwikkeling van informatie-systemen.



Afbeelding 3: Samenhang van hulpmiddelen.

organisaties zelfs hun hele financiële (kredieten) informatiehuishouding hebben moeten aanpassen als gevolg van het feit dat kredieten meer en fijnmazig(er) online gescord dienen te worden. Bovendien dienen nu van front-office tot en met back-office, én datawarehouse-omgeving, alle data aan elkaar gekoppeld te worden, met maximale betrouwbaarheid. Zowel IFRS en Basel II, als SOX hebben er ook toe geleid dat concepten als logging en audit trail, essentieel onderdeel zijn gaan vormen van de oplossingen. Sinds IFRS, Basel II en SOX een rol zijn gaan spelen liggen metadata en metadata management steeds meer onder de

aandacht van organisaties en hun management. De belangrijkste problemen die organisaties hebben ervaren ten aanzien van IFRS, Basel II en SOX zijn: datakwaliteit; ontbreken van metadata en documentatie van legacy-systemen waardoor analyse, ontwerp en realisatie erg moeilijk zijn; het afwezig zijn van hulpmiddelen voor het verzamelen, administreren en beheren van metadata; het ontbreken van een information dictionary en modelleerhulpmiddelen. Ontwikkelingen als IFRS, Basel II en SOX hebben duidelijk gemaakt dat een goed (meta)data management een belangrijke voorwaarde is voor succes. Alles valt of staat met duidelijke/eenduidige definities (metadata).

Drie manieren

Er bestaan grofweg drie manieren om metadata te ontsluiten, te weten: 1. met de individuele hulpmiddelen met hun eigen repository's en onderlinge metadata interfaces (de eerdergenoemde modelleerhulpmiddelen); 2. met een virtuele metadata-schil, welke de onderliggende individuele metadata-hulpmiddelen en bijbehorende repository's transparant maakt; 3. met een centrale metadata-repository, welke als fysieke integrator fungeert op de onderliggende individuele metadata-hulpmiddelen en bijbehorende repository's. De virtuele metadata-schil van 2. kent een beperkt eigen repository voor 'overige metadata' en bijvoorbeeld gebruik(ers)statistieken en is doorgaans (near) real-time, omdat ze doorprijkt naar de onderliggende repository's. De centrale metadata repository van 3. werkt doorgaans op basis van snapshots van de onderliggende repository's van de metadata-hulpmiddelen.

Service Oriented Architecture (SOA), ontologieën en het semantic web.

Als er één categorie de *runner up* prijs verdient op het gebied van metadata, dan is het wel SOA. SOA werkt alleen wanneer metadata in vergaande mate zijn geïntegreerd. Dit wil zeggen dat services worden verleend op basis van componenten die werken op basis van (contextgebaseerde) metadata via een gestandaardiseerde API en bus. Zowel API als bus kunnen gekocht zijn dan wel zelfgemaakt: in het eerste geval heeft men een standaard aantal metadata services ter beschikking; in het tweede geval zal dit zelf moeten worden ontwikkeld. In sommige gevallen bestaat er een combinatie, waarbij de technische API die voor het fysieke datatransport is gekocht is, en hierop zijn dan een functionele API en bus geïmplementeerd voor de feitelijke applicatie-integratie. In dit geval is het dan zo dat er dan additionele metadata services kunnen worden ontwikkeld. Toepassing van SOA impliceert kiezen voor metadata.

Wanneer SOA wordt toegepast, is er sprake van een objecten-model. Een op SOA gebaseerde applicatie (logica, services, en dergelijke) hoeft niet het hele objectenmodel te kennen, alleen zijn context. Gezamenlijk bedekken de SOA-applicaties het objectenmodel. Idealiter zijn niet alleen de primaire informatiesystemen gebaseerd op SOA, ook secundaire informatiesystemen voor bijvoorbeeld de bestuurlijke informatievoorziening zijn geïntegreerd in SOA (nota bene: dit zegt nog niet of ze dezelfde beschikbaarheid hebben, dezelfde actualiteit of verversings-frequentie hebben).

Waar SOA als concept prima is, verloopt de implementatie van SOA doorgaans lastiger. Dit heeft te maken met het feit dat men tamelijk specifiek dient te modelleren wat en hoe zaken moeten worden uitgevoerd. Verschillende leveranciers hebben hun eigen oplossing. Wanneer we naar de web-implementatie kijken, is de laatste jaren steeds gestandaardiseerd op WSDL (de services), UDDI (de catalogus) en SOAP (de toegang). In vergelijking met applicaties die één of meer decennia geleden werden ontwikkeld (op basis van oudere concepten, zie boven) en die aanzienlijk meer business-georiënteerde metadata bevatten, concentreren de op basis van SOA ontwikkelde (web)applicaties zich voornamelijk op technische metadata. Maar er gloort hoop aan de horizon. Een veelbelovende ontwikkeling op dit gebied is de opkomst van ontologieën en het semantic web. Het begrip ontologie houdt in dat er kennisrijke modellen worden ontwikkeld die bestaan uit: een taxonomie (hiërarchische kennisstructuur); basisconcepten als bijvoorbeeld overerving; een verzameling kennisrijk beschreven kenmerken en eigenschappen. Ontologisch of semantisch modelleren is sterk verwant met frame-gebaseerde kennismodellering. Dit wil zeggen dat het vastleggen van betekenis en context bepalend is (semantische modellering). Dit maakt het mogelijk dat er met de kennis kan worden 'geredeneerd'. Het moge duidelijk zijn dat hier metadata (en kennis ervan) een voorname rol spelen.

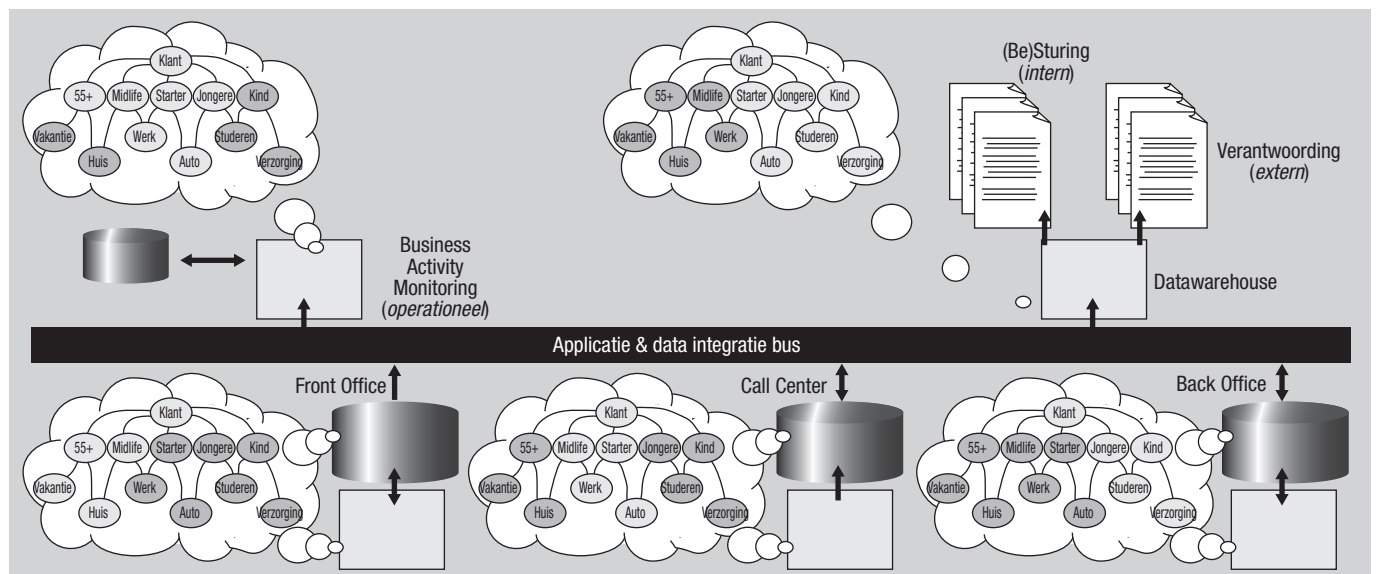
Ontologie staat niet op zichzelf. Ontologisch modelleren wordt onder andere toegepast in het 'semantische web', een ontwikkeling om op een 'intelligente manier' kennis te associëren en vast te leggen. Voor de toepassing van ontologie in een web-context, dus om het 'semantische web' te creëren, zijn er specificatietaal ontstaan als OWL (Web Ontology Language), en haar voorloper RDF (Resource Description Framework). Hoewel we van mening zijn dat er SOA-vriendelijke vormen van een 'semantisch web' in en tussen organisaties zullen ontstaan, waarin metadata transparant zijn en op een zeer eenvoudige wijze te ontsluiten, zal het nog wel enige tijd duren voordat dit gemeengoed is.

Conclusie

We hebben slechts een tipje van de sluier kunnen oplichten ten aanzien van de 'open' ruimte die metadata heet. Hoewel de ontwikkelingen van de laatste decennia laten zien dat er veel bereikt is rondom metadata, wordt het onderwerp metadata in de praktijk vooral technisch aangepakt. Vandaar dat we aangeven dat metadata en vooral metadata management 'tussen de oren' moeten zitten en vanuit een visie en (business-) strategie moeten worden aangevlogen, in plaats van met een specifieke tool. Tegenwoordig zijn de belangrijkste redenen voor gebruik van metadata consistentie, eenduidigheid, kwaliteit, en accuratesse, welke zijn terug te vinden in een op basis van SOA ontwikkeld applicatielandschap en met name het datawarehouse. De belangrijkste remmers op inzet van metadata zijn, nog steeds, de hoge kosten, de complexe technologie, en de nadruk op technische metadata.

Dr. Martin Misseyer (martin.misseyer@ordina.nl) is Profession Leader bij Ordina VisionWorks.

Drs. Karel Auwerda (karel.auwerda@ordina.nl) is Sr. Consultant bij Ordina Finance.



Afbeelding 4: De Service Oriented Architecture.