

Informatica creëert meta-repository met hulp van PowerCenter en PowerAnalyzer

SuperGlue maakt zakelijk hulpmiddel van metadata

Robbert Hoeffnagel

'Data over data' wordt tot nu toe voornamelijk gebruikt om de interne operatie van de ICT-afdeling te kunnen beoordelen en aan te sturen. Nu overheden echter steeds meer en bovendien steeds strakkere eisen stellen aan, bijvoorbeeld, financiële rapportages neemt het belang van metadata voor zakelijke doeleinden snel toe. Daarom introduceerde Informatica onlangs het product 'SuperGlue'. Met dit metadata-managementsysteem kan metadata uit meerdere repository's bij elkaar worden gebracht en geanalyseerd.

Het belang van metadata kan nauwelijks overschat worden. Informatica citeerde bij de introductie van SuperGlue echter voor de zekerheid toch nog maar even Fred Smith, de oprichter van pakkettenvervoerder Federal Express. Deze zei al in 1970: "The information about a package is as important as the delivery of the package itself." Vrij vertaald maakt Informatica daarvan: informatie over een gegeven is even belangrijk als het gegeven zelf.

Het is voor bedrijven steeds belangrijker dat zij precies kunnen documenteren hoe zij aan bepaalde cijfers en gegevens komen

Metadata wordt traditiegetrouw vooral voor de interne ICT-operatie vastgelegd. De hiervoor gebruikte repository's worden veelal per applicatie gevuld met definities en aanvullende gegevens over data. Het doel van die vastlegging is vrijwel altijd het kunnen beoordelen en sturen van interne ICT-processen. De wereld is de afgelopen jaren echter ingrijpend veranderd. Kijk naar de eisen die Amerikaanse wetgeving als de Sarbanes-Oxley Act of – meer internationaal georiënteerd – Basel II stellen aan financiële rapportages.

Wetgevers stellen in toenemende mate eisen aan de manier waarop gegevens worden aangeleverd aan bijvoorbeeld een belastingdienst of andere overheidsinstellingen. Vragen die steeds vaker gesteld worden, zijn bijvoorbeeld: Uit welke bronsystemen zijn de ingeleverde cijfers afkomstig? Wat waren de oorspronkelijke definities aan de hand waarvan de cijfers werden vastgelegd? Welke 'business rules' speelden hierbij een rol? Hoe zijn de cijfers geconsolideerd? Welke bewerkingen hebben ze daartoe ondergaan? Al deze vragen kunnen aan de hand van metadata worden beantwoord.

Betere informatie

Er speelt echter meer. Het management van veel organisaties heeft betere informatie nodig om tot goede beslissingen te kunnen komen. Dat is onder andere een gevolg van de sterk toegenomen concurrentie. Een vraag als 'wat zijn onze vijf meest winstgevendste klanten?' lijkt in eerste instantie eenvoudig te beantwoorden, maar dat wil in de praktijk nog wel eens flink tegenvallen. Want hoe definiëren we precies een kreet als 'winstgevend'? En wat is een klant? En wanneer is een klant eigenlijk een klant en niet langer een prospect? Vanaf het moment dat de order mondeling binnen is of pas op het moment dat de opdracht formeel en schriftelijk is bevestigd is? Of kijken we naar het moment van betalen? Maar hoe zit het dan met bedrijven die in termijnen een factuur voldoen?

Deze informatie kan in de vorm van metadata in een repository worden opgeslagen. Daarmee kan een vaste en eensluidende definitie van alle belangrijke gegevens worden vastgelegd. Wie precies wil weten wat een klant nu precies is, hoeft dus alleen maar de definitie op te halen uit de bij de applicatie behorende repository.

Dat is op zich natuurlijk niets nieuws. Voor zakelijke toepassingen werkt deze aanpak echter maar ten dele. In veel gevallen zal het voor een vraag als hiervoor gesteld is, namelijk nodig zijn om metadata op te halen uit meerdere repository's. Dat wil dus zeggen dat gebruik gemaakt zal moeten worden van datadefinities en andere aanvullende gegevens, die in meerdere repository's vastliggen. Hoewel het iedere keer zal gaan om data met b etrekking op – bijvoorbeeld – 'klant', zal al snel blijken dat er verschillen optreden, die worden veroorzaakt doordat er geen centraal (met andere woorden: over meerdere applicaties en

repository's heen) vastgelegde definitie van het fenomeen 'klant' bestaat. In ieder systeem wordt een eigen omschrijving gehanteerd. Op deze manier kan het dus heel goed gebeuren dat de definitie van 'klant' in een Siebel-systeem niet geheel overeenkomt met de omschrijving die in SAP of Peoplesoft wordt gehanteerd.

Softwarebedrijf Informatica denkt dat het goed mogelijk is om metadata op een zodanige manier te ontsluiten dat deze gegevens ook voor zakelijke doeleinden kunnen worden gebruikt. Daarvoor is het echter wel nodig dat de gegevens die in de diverse in gebruik zijnde repository's vastliggen bij elkaar kunnen worden gebracht. Deze gegevens dienen verzameld te worden in wat het bedrijf een meta-repository noemt. Om iets met de aldus verzamelde data te kunnen doen, zijn bovendien analyse- en bewerkingsfaciliteiten nodig.

ETL plus BI

Wie een stap terug zet, ziet al snel dat het in feite om een vrij klassiek probleem gaat: net als bij een datawarehouse of een datamart is een combinatie nodig van een ETL-operatie (extraction, translation and loading) en Business Intelligence. Kennelijk heeft men ook bij Informatica die stap terug gedaan en zag men dat men met een combinatie van twee bestaande

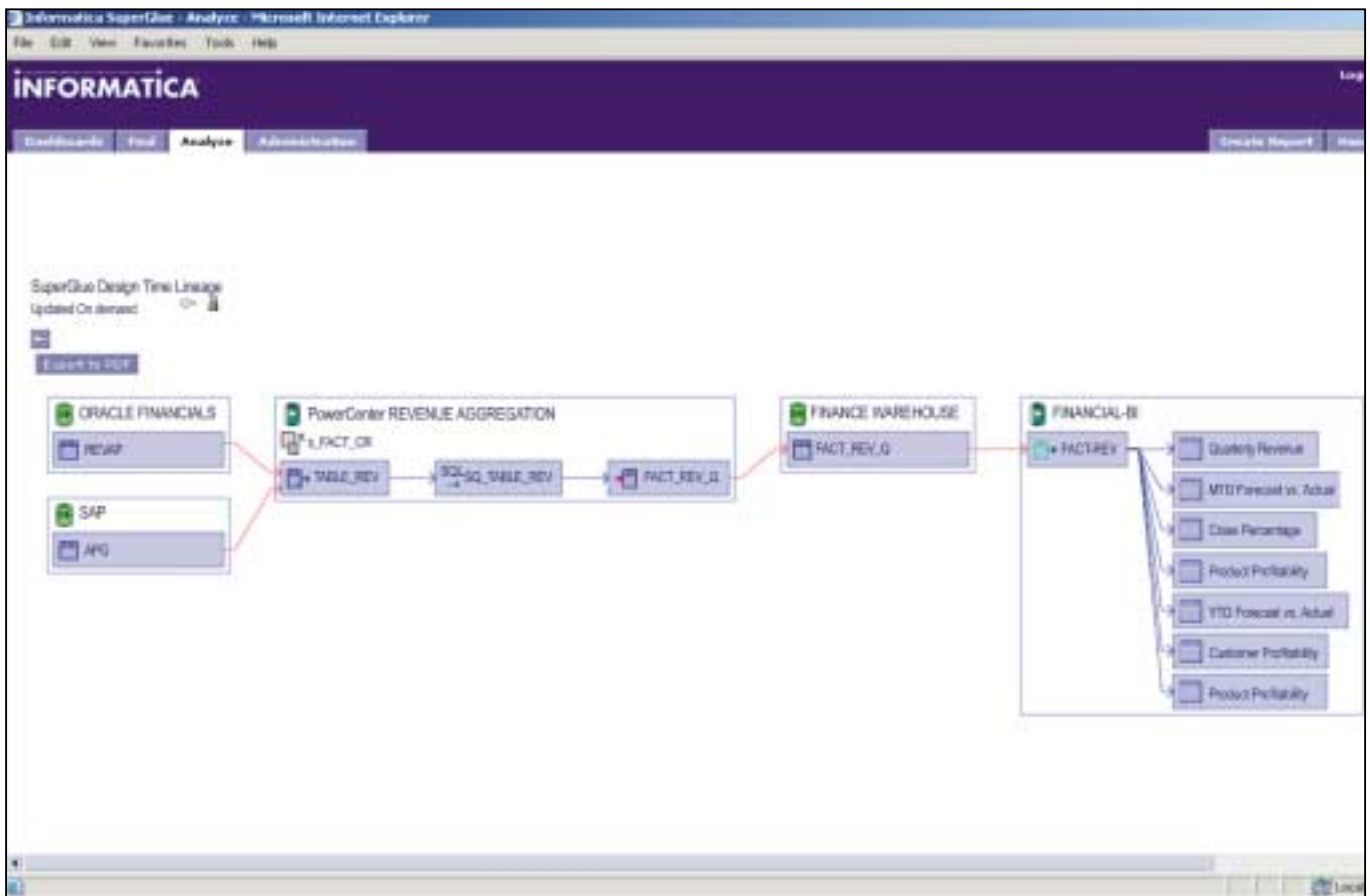
producten al een flink deel van de benodigde functionaliteit in huis had. Voor de ETL-operatie beschikt het bedrijf immers reeds over PowerCenter, terwijl men voor Business Intelligence de PowerAnalyzer-software kent.

PowerCenter is 'van huis uit' echter niet ontwikkeld met het oog op het aansluiten op repository's. De in dit product opgenomen integratie-server kan echter wel worden toegepast, mits deze via een interface kan worden gekoppeld aan een repository.

Deze nieuw ontwikkelde interfaces noemt men Xconnects

Bij Informatica noemt men deze nieuw ontwikkelde interfaces Xconnects. Deze zijn bedoeld om enerzijds een link met een repository te leggen en anderzijds op een gecontroleerde manier de metadata uit de repository op te halen en over te brengen naar de meta-repository.

Informatica ontwikkelt zelf een reeks Xconnects, waardoor gegevens kunnen worden opgehaald uit bekende enterprise-applicaties, maar bijvoorbeeld – naast het eigen PowerAnalyzer –



Afbeelding 1: Uit welke basisdata is een gegeven opgebouwd?

ook Business Intelligence-producten van bijvoorbeeld Business Objects. Daarnaast is een reeks van databases bereikbaar, evenals 'flat files', XML en op het CWM/XMI-model van de Object Management Group gebaseerde bronnen (CWM staat voor Common Warehouse Metamodel, XMI betekent XML Metadata Interchange).

SuperGlue kan uit de voeten met zowel technische als zakelijk georiënteerde metadata

Ook metadata uit Office-bestanden kan worden geïmporteerd. Ook heeft het bedrijf een 'software development kit' gepubliceerd, waarmee ICT-afdelingen zelf adapters kunnen ontwikkelen voor maatwerkomgevingen, terwijl kleinere softwarebedrijven dit voor de eigen programmatuur kunnen doen.

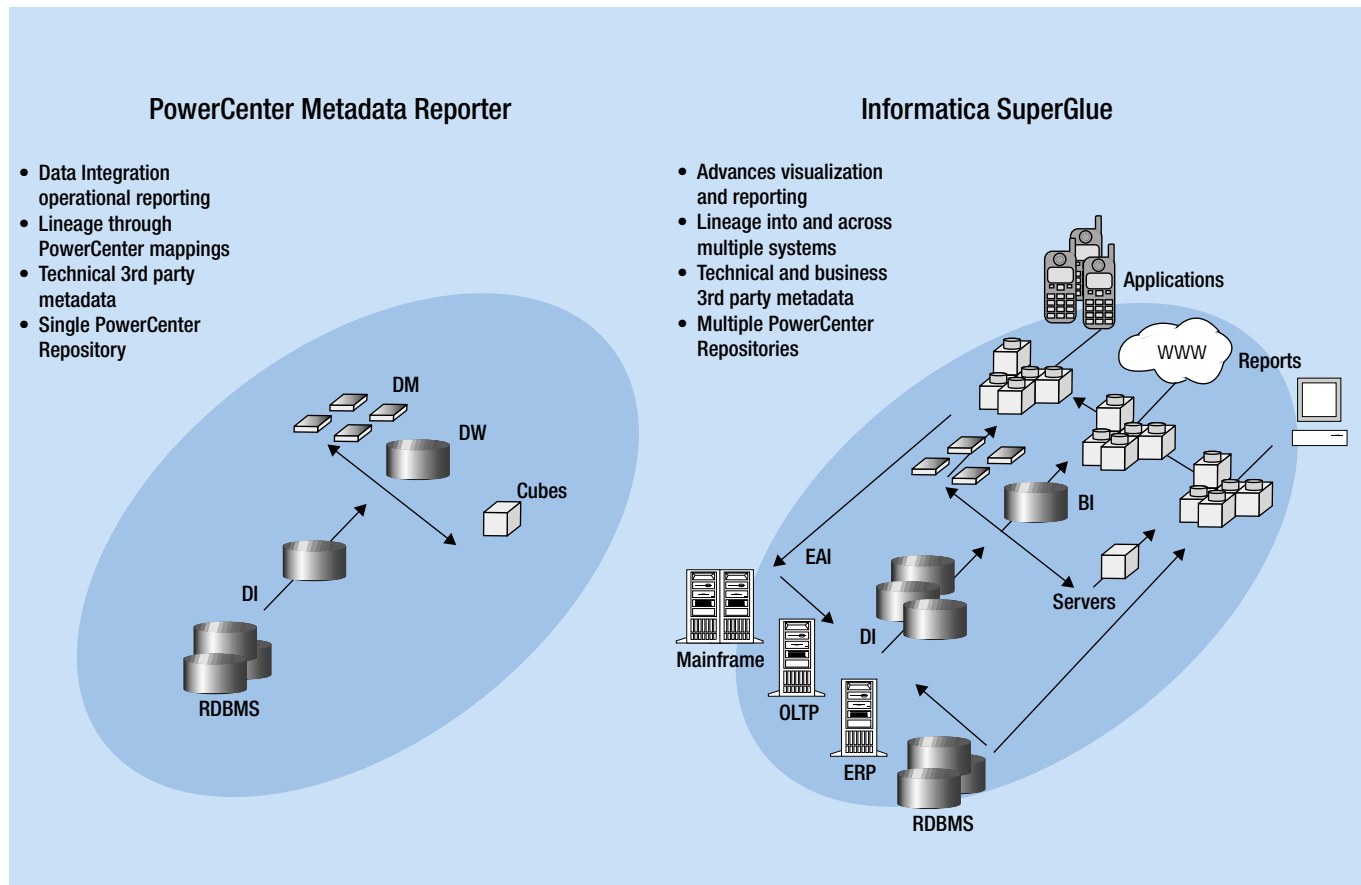
Adapters bouwen

Vooralsnog door de strakker wordende regelgeving is het voor veel bedrijven steeds belangrijker dat zij precies kunnen documenteren hoe zij aan bepaalde cijfers en gegevens komen, waar deze vandaan komen en dergelijke. Vooral de Xconnect-adapters zijn

hierbij belangrijk. Het zijn immers deze interfaces die een hoofdrol spelen bij het uit individuele repository's ophalen van metadata. Een Xconnect mag dan uiteraard geen 'black box'-achtig stukje software zijn, maar zou juist goed gedocumenteerd en openbaar moeten zijn. Alleen dan kunnen overheden zich immers een beeld vormen van de manier waarop cijfers en gegevens tot stand komen. Dit zou bijvoorbeeld kunnen door deze adapters te laten keuren en eventueel te laten certificeren door onafhankelijke EDP-auditors of andere partijen. Maar zover zijn de ontwikkelingen binnen Informatica echter nog niet.

Hoe werkt SuperGlue nu in de praktijk en wat levert het gebruik ervan op? Uit tal van systemen, van ERP tot BI, wordt metadata opgehaald uit de onderliggende repository's en bijeengebracht in de meta-repository. Dit kan een Oracle-, DB2- of bijvoorbeeld SQL Server-database zijn. Hier kan deze informatie worden bewerkt, bijvoorbeeld om definitieverschillen op te lossen.

Helaas is het vooralsnog niet mogelijk de aangepaste metadata vervolgens weer terug te sluizen naar de oorspronkelijke repository. In dat geval zou van een echte vorm van metadata-management sprake zijn. Dit wordt waarschijnlijk wel mogelijk als de ontwikkelaars van de onderliggende systemen eveneens besluiten om de XMI-standaard te ondersteunen. In dat geval wordt in principe tweewegverkeer tussen beide metadata-omgevingen mogelijk.



Afbeelding 2: Vergelijking van de Metadata Reporter uit PowerCenter en SuperGlue.

Meta Dimensions

Wie zich in Nederland bezighoudt met metadata komt al snel de naam 'Meta Dimensions' tegen. Hoewel het bedrijf nog altijd bestaat, houdt het zich vooralsnog niet meer bezig met metadata. "Informatica heeft de intellectuele eigendommen van Meta Dimensions voor een niet nader bekend bedrag overgenomen," vertelt Bert Oosterhof, tot voor kort directeur van Informatica Nederland, maar inmiddels doorgeschoven naar het Europees management. Hij was zelf als één van de investeerders bij Meta Dimensions betrokken. "Het is eigenlijk een beetje jammer dat Informatica al redelijk ver gevorderd was met het ontwikkelen van SuperGlue. Hierdoor was het niet meer mogelijk om de aanpak van Meta Dimensions in dit product mee te nemen. Ik weet echter wel dat de ideeën die bij dit bedrijf leefden zeker hun weg binnen de Informatica-organisatie zullen weten te vinden."

De metadata die vastligt in de meta-repository, kan met PowerAnalyzer worden doorzocht en geanalyseerd. Aangezien SuperGlue geheel op het CWM-model is gebaseerd, is het in principe ook mogelijk om deze analyseslag uit te voeren met niet door Informatica ontwikkelde producten, mits deze eveneens

conform het CWM-model functioneren. Ook de visualisatiemogelijkheden van PowerAnalyzer kunnen worden toegepast, men kan denken aan door de gebruiker zelf te configureren dashboards, met daarin de bekende grafische hulpmiddelen om snel de essentie van informatie te kunnen doorgronden.

Herkomst zichtbaar maken

Al deze grafieken en dergelijke zijn in principe doorklikbaar, zodat de onderliggende informatie beschikbaar komt. De gebruiker komt daarmee in de zogeheten 'metadata directory'. Hierin is de hiërarchie van alle geïmporteerde repository's zichtbaar en kan bijvoorbeeld de classificatie van de metadata worden bekeken. Wie dieper door wil dringen in een bepaalde class krijgt bijvoorbeeld te zien wat in afbeelding 1 is weergegeven.

Dit is een voorbeeld van wat Informatica 'intelligent lineage' noemt. Te zien is hoe een bepaald gegeven tot stand is gekomen door het combineren van gegevens uit Oracle Financials en SAP. Doordat standaard een exportfaciliteit naar PDF-formaat is toegevoegd, is het nu eenvoudig om een niet meer door derden te veranderen bestand te maken waarin dit soort onderbouwende gegevens zijn vastgelegd.

'Lineage' is één van de mogelijkheden die SuperGlue in de visie van Informatica aanbiedt. Deze faciliteit is in feite niet veel anders

dan het in de vorm van een stroomschema visualiseren hoe een bepaald gegeven uit andere gegevens is opgebouwd. Daarnaast kan met PowerAnalyzer de meta-repository geanalyseerd worden en kunnen hiervan rapporten worden gemaakt – op papier of via

Via een gepersonaliseerde metadata-directory is het mogelijk te zoeken over systemen en projecten heen

een dashboard, waarmee een gepersonaliseerde kijk op de meta-repository mogelijk wordt. Via een gepersonaliseerde metadata-directory is het bovendien mogelijk te zoeken over systemen en projecten heen. Ook kunnen alarmeringen en andere typische BI-functies worden toegepast.

Gebruikers van Informatica-software zullen zich wellicht afvragen wat nu precies het verschil is tussen SuperGlue en de reeds langer in PowerCenter beschikbare Metadata Reporter. Het verschil is te zien in afbeelding 2. De Metadata Reporter

biedt operationele rapportages op basis van data-integratie, terwijl 'lineage' alleen via mappings in PowerCenter mogelijk is. Bij SuperGlue heeft de 'lineage' bovendien betrekking op meerdere systemen. SuperGlue kan daarnaast uit de voeten met zowel technische als zakelijk georiënteerde metadata, waar Metadata Reporter alleen met technische metadata uit de voeten kan. Bovendien werkt de Metadata Reporter met slechts één PowerCenter Repository, terwijl SuperGlue met meerdere PowerCenter Repository's overweg kan.

Conclusie

Metadata kan met de hulp van metadata-managementsystemen als SuperGlue uitgroeien van een puur technisch onderwerp tot een zakelijk hulpmiddel, waarmee het management betere beslissingen kan nemen en bovendien kan voldoen aan de steeds strengere eisen die overheden stellen. Belangrijk hierbij is echter wel, dat in principe alle repository's van belangrijke bedrijfs-informatiesystemen kunnen worden ontsloten. Gebeurt dat niet, dan is de kans groot dat beslissingen genomen worden op basis van incomplete gegevens.

Robbert Hoeffnagel is freelance journalist.
