



Bij implementatie datawarehouse noodzaak metadata het sterkst

# Metadata-management staat nog in de kinderschoenen

Daan van Beek

**Met de opkomst van datawarehouses en Business Intelligence, waar redundantie van data en ook metadata een 'vervelend' neveneffect is, ontstond er gelijktijdig behoefte aan centralisatie en coördinatie van metadata binnen organisaties. Daan van Beek bespreekt de rol die metadata speelt in Business Intelligence en datamanagement, wat het betekent, welke voordelen het voor organisaties heeft, welke soorten metadata er zijn en welke architecturen worden onderkend.**

Metadata is data over data, het beschrijft allerlei technische en organisatorische aspecten van de data die bestaat, stroomt of teloor gaat, zie ook het kader Idee. Metadata bestaat slechts in het geval van gestructureerde data. Data zonder structuur staat volledig op zichzelf, kent geen context en kan als zodanig ook niet begrepen worden. Immers, zonder omgeving zijn betekenissen en beschrijvingen hol en is het nut nihil. Ongestructureerde gegevens hebben dus geen metadata.

Om profijt te trekken van metadata binnen organisaties zullen deze het dienen te integreren in de bedrijfsvoering en in het bijzonder ICT. Organisaties staken de afgelopen jaren veel tijd in

het ontwikkelen van allerlei informatiesystemen die de bedrijfsvoering ondersteunen; nu is de tijd gekomen om gereedschappen en informatiesystemen te ontwikkelen die ICT zelf helpen om efficiënter de bedrijfsprocessen te ondersteunen. "We build systems to manage every aspect of our business, except one to manage the systems themselves" (Marco, 2003). Het is als een hovenier die een saaie tegeltuin heeft, een schoenmaker die op afgetrapte schoenen loopt en een dominee die doordeweeks de bloemetjes buiten zet.

Metadata-management, een niet onbelangrijk onderdeel van data-management, is het geheel aan processen dat een onderneming activeert en uitvoert om de in het bedrijf aanwezige data te beschrijven en te communiceren aan mensen én systemen, om het rendement op de data zelf te maximaliseren. Metadata-management bedient zich vaak van een metadata-repository die alle metadata van een onderneming opslaat en ontsluit.

## Idee

Plato was een van de eersten die onderscheid maakten tussen 'het idee' en 'het ding'. Hij noemde het de schaduwen (ideeën) van de werkelijkheid en bedacht dat de dingen die bestaan in het dagelijks leven een voorafgaand idee hebben. Zo heeft een paard een idee paard; een computer een idee computer; een bril een idee bril. Het idee zoals Plato dat indertijd introduceerde, komt overeen met metadata, ook wel data over data genoemd.

Ieder afzonderlijk paard 'stroomt', het groeit, eet en sterft; het idee paard is echter onveranderlijk en eeuwig. Het begrip metadata grijpt dus terug op een oud filosofisch principe, hoewel er toch een verschil is want data op zich bestaat ook niet; het begrip is slechts een verwijzing naar de werkelijkheid, bijvoorbeeld een artikel, een afnemer of een order.

Metadata is dus een beschrijving van een beschrijving van reële zaken. Zelfs die zijn niet altijd reëel, omdat data kan verwijzen naar een virtueel, niet tastbaar aspect van het dagelijks leven, bijvoorbeeld een gesprek of een ontmoeting. Tot zover het idee metadata.

## Toegevoegde waarde

Metadata kent diverse toepassingen en een effectievere manier om informatiesystemen te ontwikkelen en te beheren. Hierdoor zullen ICT-kosten in relatieve zin dalen en zal het rendement van een onderneming op ICT toenemen. Wanneer een organisatie bijvoorbeeld meer afnemers krijgt en in de begintijd gekozen is voor een nummer met een lengte van vier posities en het aantal afnemers stijgt binnenkort boven de 10000, dan helpt een metadata-voorziening dit soort wijzigingen in informatiesystemen snel te implementeren.

Allereerst wil men binnen een organisatie de impact weten van de noodzakelijke wijziging, op hoeveel plaatsen en in hoeveel programma's wordt dit attribuut gebruikt. Hierna zal men graag met één druk op de knop de wijziging willen doorvoeren in zowel de (fysieke) datamodellen en de programma's. Dit zal nog voor

veel organisaties als toekomstmuziek in de oren klinken. Toch zullen zij, vanwege de toenemende complexiteit van de ontwikkelde systemen én de druk om snel te reageren op veranderende marktomstandigheden, metadata-management op een of andere manier gaan omarmen. Snel kunnen reageren is immers in deze turbulente tijden een must.

## De meer technische entiteit-relatie-diagrammen laten zich goed vertalen in business rules

Tevens zal het bouwen van Business Intelligence toepassingen velen malen sneller gaan wanneer metadata van de bronsystemen beschikbaar en goed toegankelijk is. De eerste stap is het in kaart brengen van de key performance indicators, die belangrijk zijn voor het sturen van de onderneming maar hierna zal op zoek moeten worden gegaan naar de data zelf. Waar is de data opgeslagen? Wat betekent die? Welke relaties heeft de data? Wat is de kwaliteit ervan? Enzovoort. Een metadata-voorziening beantwoordt al deze vragen en kan daarom enorm veel tijdswinst opleveren bij het ontwikkelen van een managementinformatie-voorziening.

### Soorten metadata

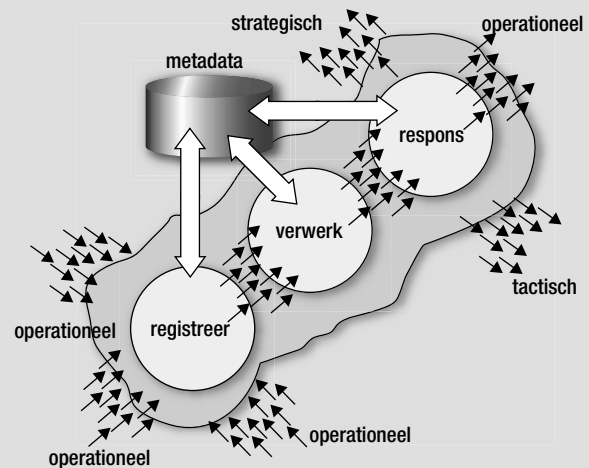
Wat moeten we ons voorstellen bij metadata, welke soorten zijn er en wat is de zin ervan voor een organisatie en haar generieke basisprocessen? Metadata valt grofweg uiteen in een aantal hoofdgroepen, die feitelijk beschrijven wat metadata kan doen voor een organisatie:

- Het beschrijft de betekenis die de data heeft voor een organisatie en haar deelnemers zoals medewerkers, afnemers en leveranciers;
- Het beschermt tegen vervuiling door af te bakenen wat wel of niet is toegestaan (mogen er meer afnemers zijn die dezelfde postcode en huisnummer hebben; wie mag de data inzien of muteren);
- Het bepaalt de relaties die de gegevens met elkaar (mogen) hebben (type relaties 1:1, 1:n, n:m);
- Het verdrijft de geschiedenis van data (wanneer en hoe moet er opgeschoond worden);
- Het verhaalt van de reis die data aflegt door een organisatie (de duur van de reis, query en ETL-statistieken, en het begin en einde van de reis, de mappings);
- Het vertaalt de betekenis in technische aangelegenheden en andersom.

Metadata valt verder onder te verdelen in technische en bedrijfsmatige metadata (Marco, 1999). Dit onderscheid is echter op conceptueel niveau niet echt relevant omdat het technische vrijwel altijd vertaald kan worden in het bedrijfsmatige en andersom. De meer technische entiteit-relatie-diagrammen laten zich in de

## Drie basisprocessen

De drie generieke basisprocessen van een organisatie, registreer, verwerk en respons, vormen het uitgangspunt van metadata-management. In afbeelding 1 is dat weergegeven. Metadata voedt de drie processen met beschrijvingen van de signalen die een organisatie registreert en verwerkt tot een adequate respons. Ieder generiek proces kent zijn eigen metadata-aspecten en -uitdagingen. Binnen het eerste proces zijn *constraints* van groot belang; dit zijn beschrijvingen die verhoeden dat er onjuiste data wordt geregistreerd. In het laatste proces is vooral behoefte aan een beschrijving en betekenis van de data. Alle metadata, in welk proces het dan ook voorkomt, is te zien als kennis van de bedrijfsvoering, van de bedrijfsprocessen en de markten waarin een organisatie opereert. In die zin geeft metadata-management min of meer impliciet handen en voeten aan kennismanagementprocessen binnen organisaties.



**Afbeelding 1:** Metadata beschrijft en stuurt de drie generieke basisprocessen van een organisatie.

Metadata is de lijm die alle bedrijfsprocessen die worden ondersteund door allerlei informatiesystemen bij elkaar houdt. Om de bedrijfsprocessen te optimaliseren kan een organisatie de uitwisseling van metadata binnen en tussen bedrijfsprocessen automatiseren. Dit zal een sterk kostenbesparend effect hebben.

praktijk goed vertalen in business rules, waarmee eindgebruikers uit de voeten kunnen. Het onderscheid is in de praktijk wel zinvol, want eindgebruikers opscheppen met implementatie-aspecten van data is niet altijd zinvol. Tenslotte is technische metadata altijd een afgeleide van de bedrijfsvoering en -processen.

### Metadata en generieke basisprocessen

Ieder generiek basisproces in een organisatie kent zijn specifieke metadata. In het eerste basisproces is het beschermen en bepalen van belang, immers de juistheid en nauwkeurigheid waarmee een organisatie gegevens registreert bepaalt mede het rendement

# Thema Metadata

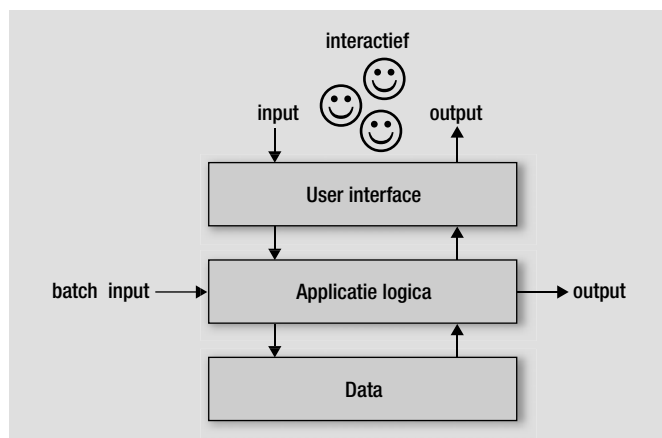
erop. Dit is voornamelijk het domein van de ontwikkelaars van informatiesystemen.

Het verhalen van de reis en het verdrijven van de geschiedenis van de gegevens is typisch voor het verwerkproces. Dit is in het bijzonder van belang voor het efficiënt IT-beheer van de informatiesystemen. Wat zijn de responstijden van query's en de laadtijden van een datawarehouse.

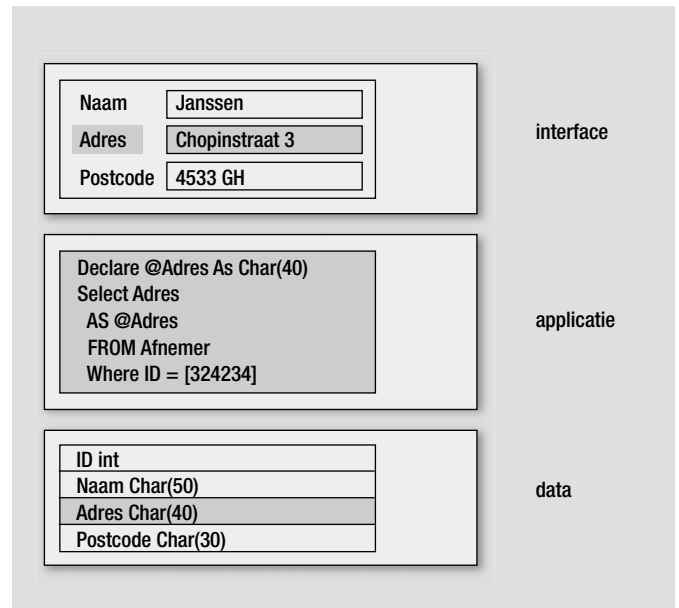
Een beschrijving van de betekenis is essentieel voor het genereren van een adequate respons, immers zonder goede beschrijving en betekenis zullen organisaties fouten maken in het operationele proces (bijvoorbeeld het verwarren van bezoek- en huisadres) en nemen zij foute beslissingen (bijvoorbeeld wanneer de netto marge niet echt de netto marge blijkt, doordat er nog extra kosten in een ander systeem worden bijgehouden). Het bepalende aspect is evenzeer van belang in het responsproces. Gebruikers van managementinformatiesystemen, die voor een belangrijk deel de respons bepalen op tactisch en strategisch niveau, moeten weten wat ze met wat mogen combineren; een productdimensie combineren met een meetwaarde orderkosten is immers helemaal niet zinvol. Orderkosten liggen over het algemeen niet per product vast maar op een hoger niveau. Hier doet zich een fenomeen voor dat typisch is bij het ontwerpen van managementinformatiesystemen: wanneer een meetwaarde a zijn oorsprong heeft in entiteit x en een dimensie b vindt zijn oorsprong in entiteit y, en x heeft een one-to-many-relatie met y, dan kan de meetwaarde a niet (zonder meer) gedimensioneerd worden door de dimensie b (Van Beek, 1999).

## De drielagen en metadata

Het drie-lagen-model (interface-applicatie-data) is een aanvaard globaal raamwerk voor applicatieontwikkeling en -beheer. Over het algemeen is het goed om de logica te scheiden van de data en de interface, zie afbeelding 2. Dit verhoedt redundantie in programmatuur en data en speelt in op de trend dat er steeds meer communicatiekanalen (interfaces zoals webpagina's) zijn waarmee gegevens (van afnemers) geregistreerd kunnen worden.



**Afbeelding 2:** Metadata en het drie-lagen-model; de pijlen geven de datastromen inclusief de benodigde metadata weer.



**Afbeelding 3:** Metadata is niet alleen redundant over diverse informatiesystemen heen, ook binnen één informatiesysteem is er redundantie; namelijk over de drie lagen heen.

Data komt dan via meerdere interfaces de organisatie binnen maar dient dan verwerkt te worden door één applicatie en opgeslagen in één database.

In de praktijk beperkt metadata-management zich vaak tot de onderste laag, terwijl het van invloed zou moeten zijn voor alle drie de lagen. Immers, de eindgebruiker zal een bepaald attribuut dat ligt opgeslagen in de database willen zien (interface) en kunnen muteren (applicatie). De definitie van een enkel attribuut komt dan voor in drie lagen. Dit is weergegeven in afbeelding 3. Het Millennium-probleem had veel sneller opgelost kunnen worden, als de drie lagen hadden geput uit dezelfde bron van metadata. Dit betekent dat metadata over de drie verschillende lagen heen verspreid zou moeten worden. Dit vereist dan een centrale bak met metadata waaruit de drie lagen kunnen putten. Metadata voorziet niet alleen in centrale opslag en ontsluiting van metadata over verschillende informatiesystemen heen, maar ook over de drie lagen heen. Het is zelfs zo dat metadata deels de interface van een applicatie kan bepalen en genereren (Baar e.a., 1992; Puerta, 1997).

Dit principe, dat metadata wordt gedeeld over de drie lagen heen, is binnen de meeste ontwikkelgereedschappen geen gemeengoed, omdat vaak bij het compileren en 'parsen' van programmatuur bekend dient te zijn van welk type de data is en zijn grootte. Dit in verband met het toewijzen van geheugenruimte. Wanneer deze mogelijkheid wel bestaat, zal dit in de praktijk vaak leiden tot een slechtere performance omdat interpretatie *at runtime* plaats moet vinden. XML-parsers, zoals deze zich bevinden in de huidige generatie internetbrowsers, beschikken standaard over dit principe; de presentatie wordt dan *on-the-fly* bepaald.

## Volwassenheidsniveau van metadata-management

Er bevinden zich op het pad naar volwassenheid op het gebied van metadata-management vijf fasen die een organisatie doorloopt. Niet iedere organisatie zal alle faseringen (in hetzelfde tempo) doorlopen; sommige organisaties hebben te maken met een meer competitieve omgeving die de noodzaak tot veranderen sterker doet gevoelen. Zij zullen bijvoorbeeld sneller besluiten een data-warehouse in gebruik te nemen (fase 2) of om metadata leidend te laten zijn bij de ontwikkeling van nieuwe informatiesystemen (fase 5). Deze faseringen zijn weergegeven in afbeelding 4.

Een organisatie bevindt zich in fase 1, als er in principe nauwelijks behoefte is aan metadata-management, omdat alle applicaties een eigen database hebben en de applicaties min of meer op zichzelf staan. Integraal prestatie-management, een belangrijke aansporing om een geïntegreerde managementinformatievoorziening zoals een datawarehouse te creëren, doet organisaties in fase 2 belanden. Al tijdens de ontwikkeling daarvan is er grote behoefte aan een centrale bak met metadata; immers bij dit soort trajecten is de definitie (welke entiteiten en attributen gebruiken we om iets te berekenen en wat betekent het) van kengetallen en dimensies van groot belang. Vaak bevredigen organisaties echter, uit pragmatische overwegingen, eerst de informatiebehoefte om pas daarna in fase 3 een centrale bak met metadata aan te leggen. Deze centrale bak met metadata vergaart alle metadata binnen een onderneming uit alle operationele informatiesystemen en management-informatiesystemen.

Impact-analyses, waarmee vragen worden beantwoord over de impact van wijzingen in de datastructuren, kunnen dan met een enkele druk op de knop opgevraagd worden. De wijziging daadwerkelijk doorvoeren in de databases kan pas in fase 4, wanneer er bi-directioneel metadata uitgewisseld wordt tussen de centrale bak en de onderliggende systemen. Fase 4 brengt slechts wijzigingen aan in de onderste laag van het drie-lagen-model, de laag waarin de databases thuishoren. Programmatuur past zich

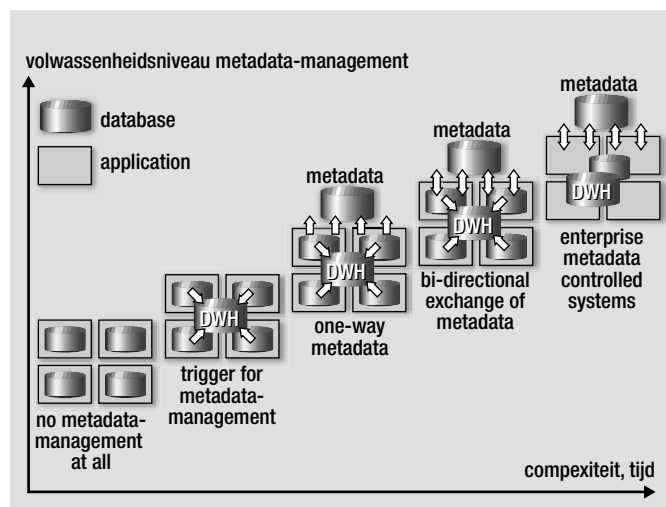
echter nog niet 'automatisch' aan. In fase 5 worden de (management)informatiesystemen gecontroleerd en aangestuurd door de centrale bak met metadata. Wanneer in deze centrale bak een wijziging wordt aangebracht dan wordt deze 'automatisch' doorgevoerd in alle onderliggende systemen en in alle drie de lagen van het drie-lagen-model. Applicatie-ontwikkeling is dan sterk metadata gestuurd en data-uitwisseling tussen de diverse applicaties vindt plaats via de metadata-voorziening.

## Conclusie

In dit artikel, de laatste van de serie over datamanagement en de drie generieke basisprocessen binnen een organisatie, is de betekenis en toegevoegde waarde van metadata aan de orde gekomen. De belangrijkste aansporing voor organisaties om serieus aan de slag te gaan met metadata is de ontwikkeling en implementatie van een datawarehouse. De noodzaak om te beschikken over een centrale bak met metadata doet zich dan het sterkst gevoelen.

Redundante data brengt immers ook redundante metadata met zich mee én het bouwen van Business Intelligence-toepassingen gaat immers velen malen sneller als de metadata van de bronsystemen beschikbaar en goed toegankelijk is. Verder kent ieder generiek basisproces een eigen specifieke toepassing voor metadata en biedt het drie-lagen-model inzicht waar metadata van belang is.

Metadata-management staat nog in de kinderschoenen; velen zijn 'pas' in fase twee of drie. Dit hoeft natuurlijk niet te betekenen dat die organisaties slecht presteren, want de markt waarin een organisatie opereert dicteert vaak het tempo waarin organisaties zich ontwikkelen. Het ontwikkelen van een gedegen visie en beleid ten aanzien van metadata-management is echter van groot belang, want de turbulentie in de samenleving en economie neemt hand over hand toe. Organisaties dienen dan te beschikken over eigenschappen zoals adaptiviteit, lenigheid en wendbaarheid. Metadata-gecontroleerde en -gestuurde organisaties en systemen kunnen zich veel sneller aanpassen aan hun omgeving en daardoor beter overleven en succesvoller zakendoen.



Afbeelding 4: Volwassenheidsniveaus voor metadata-management.

## Daan van Beek

Daan van Beek MSc (daanvanbeek@passionned.nl), is actief als zelfstandig consultant binnen passionned BI, een netwerk van denkers en doeners op het gebied van datamanagement, Business Intelligence, adaptive enterprises en knowledge management.

## Literatuur

- Baar, Dennis J.M.J., James D. Foley, Kevin E. Mullet (1992), *Coupling application design and user interface design*, ACM CHI '92 May 3-7, 1992.
- Marco, David (2003), *Metadata Repository: A System That Manages Our Systems*, TDAN.com Issue 23.0, January 2003.
- Marco, David (2000), *Building and Managing The Metadata Repository*, John Wiley & Sons, New York.
- Puerta, Angel R. (1997) *A Model-Based Interface Development Environment*, IEEE Software, jul/aug 1997.
- Van Beek, D. (1999), *Heuristic User Interface Design for Senior Managers*, Master's Thesis, Utrecht.