



Op weg naar betrouwbaar Metadata Management

Over Dingen en Feiten

Peter Alons

Als bestrijding van de vele kwalen rond datamodellering en metadata-beheer is aanpak ontwikkeld die met succes wordt toegepast: het Metadata Frame. Deze is gericht op volledige benutting van het conceptuele niveau van het informatieaspect met speciaal daarop toegesneden case-tools.

Deze aanpak is eerder beschreven in DB/M (zie noten 1 en 2) in twee artikelen over Metadata Management. De strekking daarvan was de bestrijding van enkele hardnekkige kwalen rond data-

modellering en metadata-beheer. Deze waren in het bijzonder:

1. onvoldoende gebruikersparticipatie, leidend tot slechte validatie van gemaakte informatiemodellen;
2. gebruik van niet-conceptuele aanpakken, leidend tot zwakke ontwerpen;
3. ontoereikende documentatie, leidend tot serieuze onderhoudsproblemen;
4. onvoldoende kennis van of gebruik van moderne case tool-ondersteuning, leidend tot tijdverslinding bij het modelleren, ontwerpen, prototyperen en ontwikkelen van applicaties;
5. metadata opgeslagen op verschillende, niet samenhangende locaties, leidend tot redundante opslag met fragmentatie en update-anomalieën;
6. metadata opgeslagen in handmatig onderhouden documenten, leidend tot niet meer op te brengen discipline om deze te onderhouden;
7. geen of matig gebruik gemaakt van intranet- en web-faciliteiten, leidend tot ontoereikende toegang tot metadata voor eindgebruikers;
8. geen gebruik van repository-transformaties, leidend tot onnodig verlies van informatie in de communicatie van metadata.

‘Ding-georiënteerd’

Onlangs kwam ik op Internet een artikel tegen van Marc Reugebrink genaamd ‘Marc worden’ uit De Groene Amsterdammer van 14 februari 1996. Hij beschrijft daarin zijn eigen pogingen door de tijd heen om de gedichten van Paul van Ostaijen op hun merites te waarderen, en begint met de beschrijving van zijn instinctieve afkeer als twaalfjarige jongen tegen het op school voorgedragen gedicht *Marc groet 's morgens*

de dingen, waarvan ik hier alleen de eerste regels citeer:

*Dag ventje met de fiets op de vaas met de bloem ploem ploem
dag stoel naast de tafel
dag brood op de tafel
dag visserke-vis met de pijp
en
dag visserke-vis met de pet*

Bovenstaand gedicht is volgens Marc Reugebrink een voorbeeld van Paul van Ostaijens “kinderlijk naïeve omlijning van een ding-geworden geluk”. Dit suggereert daarmee sterk, dat een kind in zijn nog naïeve kindertijd een sterk ‘ding-georiënteerd’ denken heeft. Waarom zou dit ooit omslaan in iets anders? Dit bracht mij terug bij de nog steeds in de informatiewereld voortwoekerende strijd tussen het objectgeoriënteerde en communicatiegeoriënteerde denken. Ik ben zelf een fervent voorstander van de laatste richting en leg u graag uit, hoe dat zo is gekomen.

Ooit droomde ik in mijn begintijd als IT-professional, dat mijn oudste zoon van toentertijd twaalf jaar mij vroeg een kleine database te maken voor zijn imposante knikkerverzameling. Dat moest dan maar, en omdat ik niet te veel tijd ervoor kon vrijmaken, greep ik voor de modellering meteen naar de aloude techniek van ERM. Ik had al veel van de discussies over knikkeren opgevangen, was al regelmatig op pad gestuurd om voor hem essentiële missende exemplaren te kopen, en dacht het klusje dus snel te kunnen klaren. Ik bestuurde de inhoud van zijn knikkerdoos en trof daarin allerlei soorten knikkers, zoals ‘Eentjes’, ‘Stuiters’, ‘Lodedetten’, ‘Bonken’, en ‘Superbonken’ aan. Deze kwamen alle in diverse types, zoals ‘vlammetjes’, ‘sterretjes’, ‘snotters’, ‘olietankers’, ‘piraatjes’ enzovoort en dan nog in allerlei fabriekskleuren. Ik kwam er ook een tegen die ik via een oud boek herkende als

een 'Mootjesknar', en prees mijzelf gelukkig, dat ik ook die in de nieuwe applicatie had weten in te vangen.

Ik zal u de details van de modellering besparen en u meteen de afloop vertellen. Ik liet mijn zoon de nieuwe applicatie testen en was bepaald geschokt, toen hij na enige tijd terugkeerde met een bekommerde blik op zijn gezicht. "Het ziet er best leuk uit, pap, maar eerlijk gezegd: er klopt geen hout van." Ik vroeg hem natuurlijk waarom niet en hij begon. "Allereerst kennen we geen lodedetten meer, die tellen gewoon als stuiters of eentjes, afhankelijk van hun grootte. Verder kennen wij geen mootjesknar, die zal misschien ergens anders wel bestaan, maar niet bij ons. Verder heb je groene snotters vastgelegd en die kennen we ook niet." Ik grabbelde in zijn doos en viste er triomfantelijk een op. "Dat is een blauwe snotter", zei hij, "hij is in de fabriek wat groen uitgevallen, maar dat maakt niet uit, hij telt als blauw."

Enfin, ik moest dus wel geloven aan een grotere betrokkenheid van de eindgebruiker en vroeg hem dus zelf eens de inhoud van zijn knikkerdoos in voor hem relevante termen te beschrijven. Het uiteindelijke resultaat was een behoorlijk ander datamodel en een applicatie, die hem na testen heel erg goed beviel. Om de goede afloop te vieren trokken we op naar de onvermijdelijke McDonald's. Daar kon ik niet vermijden wat weg te zakken in gemijmer, hoe het aanvankelijk zo fout kon zijn gegaan. Maar ook hier bleek een oude waarheid: kinderen en dronkaards spreken

de waarheid. Mijn zoon keek mij plotseling over zijn hamburger diepzinnig aan en zei: "Ik begrijp best waarom je het eerst fout had. Het ging mij om het spel en jou om de knikkers."

Feitgeoriënteerd denken

Sindsdien heb ik het geluk gehad bij diverse projecten – zowel grote als kleine – betrokken te zijn geweest, die aanvankelijk geheel dreigden te mislukken, maar met name door het gebruik van FCO-IM [3] – de meest moderne vorm van volledig communicatiegeoriënteerde informatiemodellering – weer vlot zijn getrokken. Al die voorbeelden zette mij definitief in de voetsporen van Ludwig Wittgenstein, die in zijn 'Tractatus Logico-Philosophicus' uit 1921 [4] het belang van *feitgeoriënteerd* denken benadrukt.

"Die Welt ist alles, was der Fall ist.

Die Welt ist die Gesamtheit der Tatsachen, nicht der Dinge.

Die Welt ist durch die Tatsachen bestimmt und dadurch, dass es alle Tatsachen sind.

Denn, die Gesamtheit der Tatsachen bestimmt, was der Fall ist und auch, was alles nicht der Fall ist.

Die Tatsachen im logischen Raum sind die Welt.

Die Welt zerfällt in Tatsachen.

....

Wovon man nicht sprechen kann, darüber muß man schweigen."

Niettemin blijft er ook een krachtige stroming van object-georiënteerde denkers en uit hun rijen wordt wel eens het volgende bezwaar tegen het feit- of communicatiegeoriënteerde denken ingebracht. Zij stellen, dat er situaties zijn waarin communicatie niet mogelijk is, omdat objecten niet geïdentificeerd kunnen worden, en wel omdat ze niet te onderscheiden zijn.

Een voorbeeld daarvan is volgens hen het damspel, waarbij zich aanvankelijk in de doos twintig witte en twintig zwarte damschijven bevinden. Bij programmeren van het damspel interesseert het de programmeur niet wat die schijven onderscheidbaar maakt: hij moet gewoon twintig witte en twintig zwarte 'instantiaties' van één object, een damschijf, hebben. In een objectgeoriënteerde model is dat geen enkel probleem. Je maakt gewoon twintig instantiaties per kleur aan van de objectklasse 'damschijf'.

De identificatie daarvan gebeurt in 'OO-land' onder water, en daarmee is het probleem dan opgelost. Of toch niet?

Wel, ik wil niet beweren dat de geboden oplossing voor een OO-programma geen goede aanpak is, maar wel dat het een oplossing is voor een probleem dat niet bestaat! Want wat wil je nu eigenlijk vastleggen van al die damschijven? Dat er twintig

De identificatiegeschiedenis van een stuk kent nog een kleine complicatie

witte en twintig zwarte in een damdoos zitten? Maar dat maakt het aantal damschijven per kleur een *feit* dat iets zegt over een *dam* doos, niet over de schijven zelf. Zodra je iets kunt zeggen over één specifieke schijf, bijvoorbeeld: "Ton Sijbrands slaat de zwarte schijf met het hapje uit de rand" dan is de betreffende schijf kennelijk identificeerbaar doordat er een hap uit de rand is. Maar doorgaans vinden we in het kader van een dampartij dat soort feiten niet zo relevant.

Wel interessant is het verloop van de partij. Daarin begint Zwart met damschijven op veld 1 tot 20 en Wit met schijven op veld 31 tot 50. En dat maakt elke schijf vanaf dat moment identificeerbaar en traceerbaar, totdat hij weer naast het bord belandt. Steeds kan tijdens de partij op een gegeven veld maar één schijf staan, en dat identificeert op dat moment de betreffende schijf. Tijdens de partij blijven de schijven dus identificeerbaar door de plaats waar ze staan, maar hun identifier verandert tijdens de partij, dat wil zeggen in de tijd. Niettemin is hiermee de geschiedenis van elke afzonderlijke schijf vanaf zijn beginpositie door de partij heen te volgen via het notatieformulier. Daarmee is altijd te herleiden hoe het spel verlopen is.

Als een schijf eenmaal geslagen is en weer naast het bord staat,

heeft hij zijn identificatiemogelijkheid die hij tijdens de partij had, weer verloren. Feitelijk geldt de beginsituatie weer. Je kunt niet meer eenduidig van een schijf in of naast de doos zeggen op welk veld het begonnen was. Maar ook dat vinden we niet echt relevant om aan iemand mee te delen.

Hetzelfde geldt natuurlijk voor schaakpartijen. Bij aanvang zitten er één koning, één dame, twee torens, twee lopers, twee paarden en acht pionnen van zowel wit als zwart in één *schaakdoos*. De schaakpartij begint met één gegeven witte toren op het veld a1, enzovoort. Het notatieformulier volgt de partij en laat zien, dat het witte paard van g1 achtereenvolgens naar f3, g5, h3, f2 en e4 ging, aldaar geslagen werd en om verder functieloos naast het bord te staan.

Daarnaast kent bij het schaken de identificatiegeschiedenis van een stuk nog een kleine complicatie, omdat een stuk naar keuze kan terugkeren op het bord door de promotie van een pion. Maar dat brengt ons feitelijk precies terug bij het voorbeeld van het knikkeren. Het gaat ons hierbij om communicatie over het *spel*, en niet over de *dingen*. in dat spel. En zo zijn we ook terug bij Wittgenstein. Alles wat we over een partij willen melden, is in FCO-IM ondubbelzinnig 'verwoordbaar'. En: "*Wovon man nicht sprechen kann, darüber muß man schweigen.*"

En hoe zit het dan met de kleine Marc uit het gedicht van Paul van Ostaïen? Welnu, die richt zich wellicht ogenschijnlijk op de dingen, maar feitelijk communiceert hij aan ons zijn 'dinggeworden geluk' in termen van heldere identificatoren. Het gaat hem niet om alle ventjes, stoelen, broden enzovoort in de wereld, maar specifiek om dat ene *ventje met de fiets op de vaas met de bloem* (als eenvoudig IT'er zal ik niet proberen het toegevoegde ploem ploem nader te duiden), en om de *stoel naast de tafel* en het *brood op de tafel*.

Zo kom ik tenslotte tot deze conclusie. Als beginnend IT'er zul je wellicht nog de neiging hebben om je te richten op de dingen om je heen, zoals Chen dat ooit voor ERM aanbeval. Maar als je door je werk tenslotte wijs bent geworden, weet je dat je steeds weer – soms op bijna lotsbepalende manier – wordt achterhaald door de feiten ... en de noodzaak van het vinden van ondubbelzinnige identificatoren daarbij.

Noten

1. *Database Magazine*, jaargang 11, no. 8 (december 2000).
2. *Database Magazine*, jaargang 12, no. 1 (januari 2001).
3. *Volledig Communicatiegeoriënteerde Informatiemodellering*, G. Bakema, J.P. Zwart, H. van der Lek, Kluwer BedrijfsInformatie, 1996
4. *Tractatus logico-philosophicus*, Ludwig Wittgenstein, Atheneum-Polak & Van Gennep, 1976

Peter Alons

Dr. Peter W.F. Alons (Peter.Alons@AtosOrigin.Com) is senior consultant bij Atos Origin/BI-CRM en betrokken geweest bij een groot aantal Business Intelligence- en Data Warehouse-projecten bij diverse bedrijven.