

Ontrafeling van de Babylonische spraakverwarring

# Het Metadata Frame in actie (1)

Peter Alons

**Zeven jaar geleden publiceerde ik twee artikelen over metadata management in DB/M [1, 2]. Daarin beschreef ik een methode genaamd het Atos Origin Metadata Frame. Deze is gericht op volledige benutting van het conceptuele niveau van het informatieaspect. Sindsdien is deze buitengewoon heilzaam gebleken in verschillende branches. In dit artikel wordt beschreven hoe de methode in gebruik is bij het Erasmus Medisch Centrum in Rotterdam.**

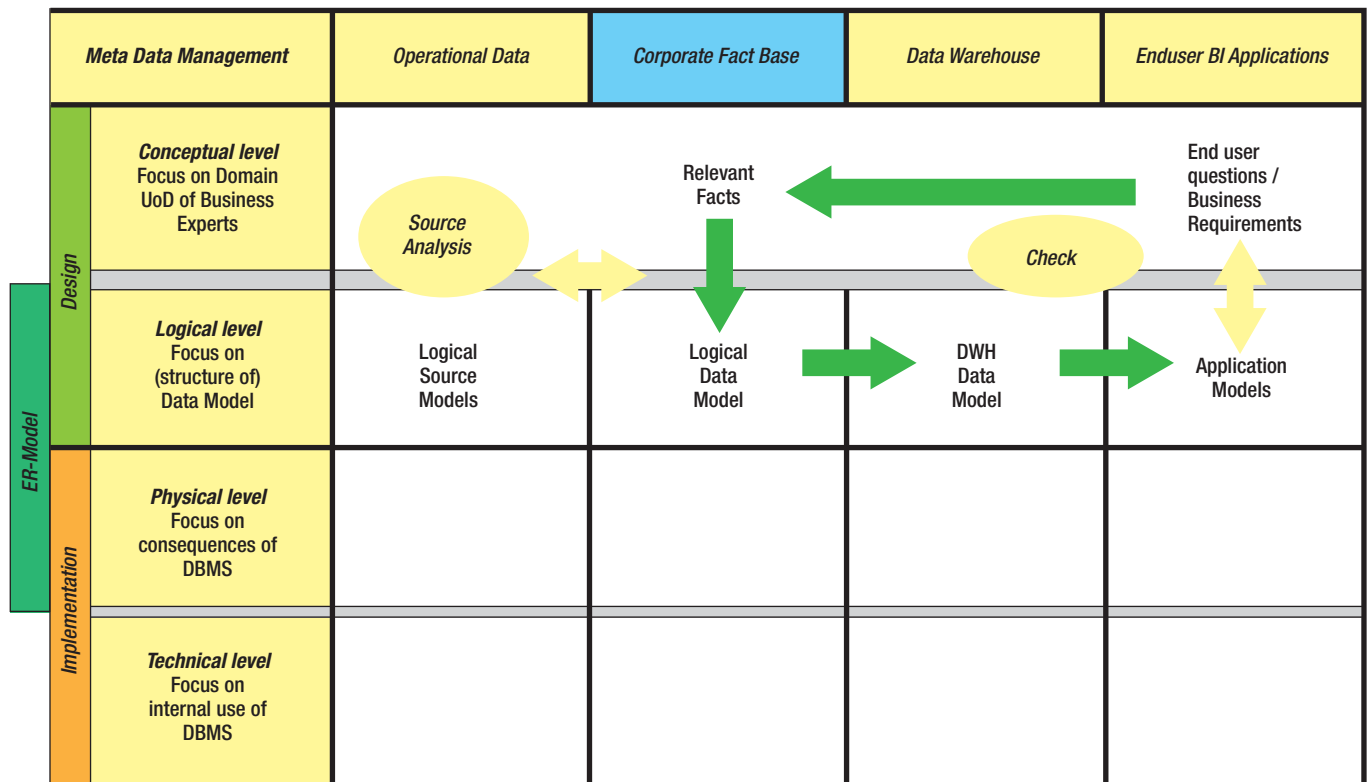
Bij het Erasmus Medisch Centrum in Rotterdam stond de IT-afdeling voor de grote uitdaging om waardevolle informatie te leveren aan de medische staf, in het bijzonder over de zorg in hun Intensive Care afdelingen (IC's). Deze informatie is niet alleen vereist voor rapportage aan de Zorginspectie, maar is ook hard nodig voor hun eigen onderzoek gericht op het verbeteren

van de kwaliteit van hun zorg. En hoewel ze over uitstekende operationele informatiesystemen beschikten, bleef hun vermogen om daar informatie uit te halen jammerlijk achter. Of zoals de geneesheer-directeur van de IC voor volwassenen, Prof. Dr. Jan Bakker, het uitdrukte: "Ik kom om in de gegevens, maar ik heb *geen informatie!*"

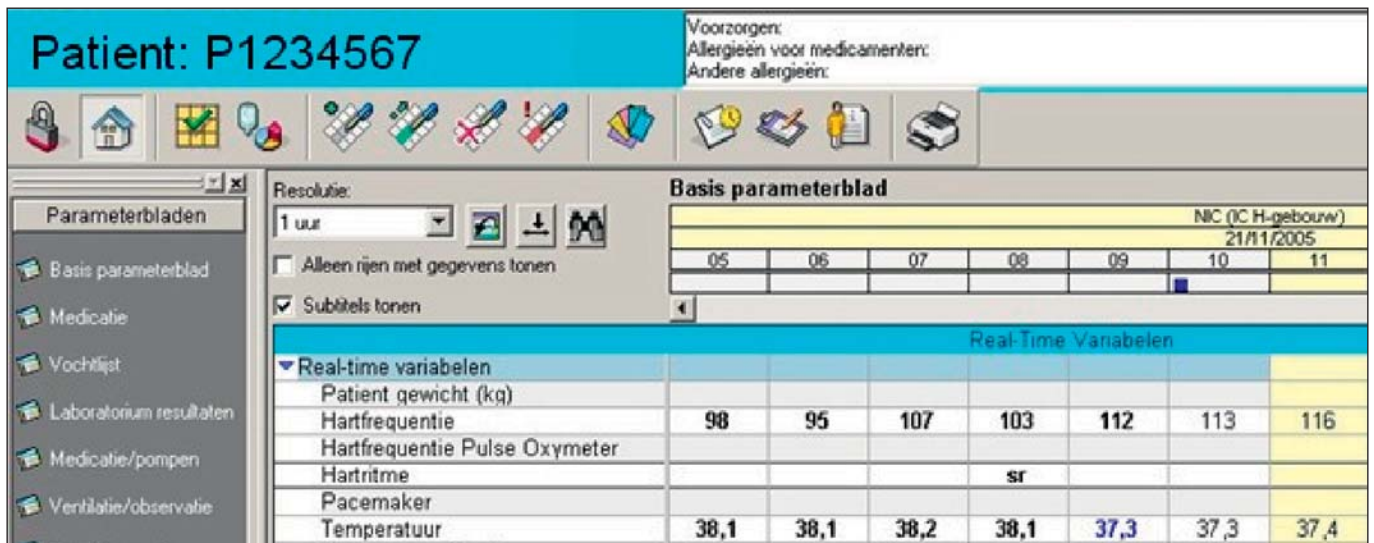
### Informatiebehoefte

Uit interviews met IC-artsen, verpleegkundigen en onderzoekers bleek dat zij adequate informatie nodig hadden om diverse redenen:

- Onderzoek voor kwaliteitsverbetering na reorganisatie van de IC's;
- Analyse van diagnoses: reden voor opname, diagnose bij opname en bij ontslag, en de uiteindelijke, gedetailleerde diagnose;
- Het met groot gemak beantwoorden van simpele vragen als: hoeveel patiënten zijn in een gegeven periode behandeld door



Afbeelding 1: Het Metadata Frame en het gebruik daarvan in het ontwerpproces.



**Afbeelding 2:** Voorbeelden van de meting van real-time variabelen uit het medische bronsysteem.

welke arts; welke artsen hebben welke medicatie toegepast in een gegeven periode; wat is het percentage van bedgebruik?;

- En het krijgen van de juiste antwoorden op moeilijker vragen als: laten we de juiste patiënten toe; wat is het aantal doses versus dagen bij medicaties voor patiënten die X doses of meer ontvingen; wat is de doorstroming van Eerste hulp, Operatiekamers, IC?

Net als veel bedrijven, probeerde de klant in dit soort informatiebehoefte te voorzien door het opzetten van een datawarehouse. En daarbij stuiten zij op dezelfde uitdagingen waarmee veel van onze klanten geconfronteerd worden, als zij een datawarehouse bouwen om waardevolle informatie te leveren aan hun eindklanten: hoe kom je aan ondubbelzinnige gegevensdefinities; hoe zet je metadata management effectief op; hoe krijg je de eindgebruikers afdoende betrokken bij het ontwerp van een datawarehouse? En hoe draagt een methodische aanpak bij aan het beantwoorden van deze drie vragen? Dat brengt ons terug bij onze Metadata Frame methode, die we vooral ontwikkeld hebben om deze vragen effectief te beantwoorden.

Als we een datawarehouse bouwen stellen we ons hoge doelen met hoge verwachtingen, en hebben behoefte aan adequate kennis en documentatie, gepaard aan een evolutionaire realisatie in een complexe en veranderlijke omgeving. Vaak is die omgeving gefragmenteerd, met verschillende codes voor dezelfde dingen en meer dan een operationeel systeem voor hetzelfde doel. Geen wonder dat we in zulke projecten serieus worden bedreigd met een Babylonische spraakverwarring. De standaard IT-aanpak om deze problemen het hoofd te bieden is een poging om een bedrijfsbreed 'business model' op te zetten in termen van een Entity-Relationship model. Helaas leidt deze aanpak niet tot ontrafeling van de Babylonische spraakverwarring.

## Ontwikkeling van informatiemodellen

Om een datawarehouse te bouwen hebben we toegesneden data-modellen nodig. Entity-Relationship Modelling (ERM) is om een

aantal redenen geen goed middel om het creatieproces van de vereiste datamodellen aan te sturen. Allereerst zijn de meeste bedrijfsmensen buiten het werkgebied van de IT niet vertrouwd met ERM. Daarom kunnen we niet verwachten dat communicatie met de business experts in termen van ER-diagrammen doeltreffend kan zijn. Ook is het niet ethisch om met iemand een gesprek op te zetten, waarvan de uitkomst voor hem van vitaal belang is, in een taal die hij niet begrijpt.

Op zijn best is deze aanpak doelmatig, maar dan alleen omdat we helemaal niet communiceren met de business experts. Maar dan moeten we niet verbaasd zijn, als de uiteindelijke uitkomst van de bouw van het datawarehouse inefficiënt en ineffectief blijkt te zijn.

Afbeelding 1 toont het Metadata Frame zoals het functioneel door ons wordt gebruikt. Dit frame richt zich op het informatieperspectief van informatiesystemen, en verdeelt dit perspectief in vier niveaus:

- Het Conceptuele niveau, waar de focus op het domeingebied ligt: het 'Universe of Discourse' van de business experts. Dit is feitelijk het 'eindgebruikerniveau', of iets fraaiër: 'eindklant niveau'.
- Het Logische niveau, waar de focus op het datamodel ligt: de structuur van en de relaties tussen alle entiteiten en gegevens-elementen. Dit is het 'ER niveau'.<sup>1</sup>
- Het Fysieke niveau, waar de focus op het fysieke platform ligt: de consequenties voor het relationele model van de gegevens-elementen. Dit omvat het toevoegen van afleidbare informatie, het benoemen van de 'foreign key' elementen, en naamsveranderingen die worden afgedwongen door het DBMS dat wordt gebruikt. Dit is het 'RM niveau' (Relationeel Model).
- Het Technische niveau, waar de focus op het internal gebruik van het gekozen DBMS ligt. Dit omvat alle indexes, 'views', en 'partition-keys' die de DBA als geschikt beschouwt. Dit is het 'DBMS niveau'.

De bovenste twee niveaus van het frame vormen de *ontwerplaa*g van het datamodelleringsproces, de onderste twee de *implemen-*  
*tatielaa*g. Het logische en fysieke niveau vormen samen de ER-laag, zoals ook is te zien in alle ER-tools zelf. Essentieel in het gebruik van het Metadata Frame is dat we de vier niveaus strikt gescheiden houden: we lossen de problemen op een gegeven niveau op tijdens werk op dat niveau zelf, niet als we aan het werk zijn op een ander niveau.

De kolommen in het Metadata Frame worden gevormd door de diverse 'zuilen' van toepassing. In afbeelding 1 zijn dit zuilen die geschikt zijn voor het bouwen van een datawarehouse. Als het frame in een andere context wordt gebruikt, kunnen de zuilen naar behoeven worden aangepast. Gemeenschappelijk voor alle toepassingen van het Metadata Frame is de zuil 'Corporate fact base', die essentieel is voor integratie van alle relevante informatie binnen het bedrijf. Hier verwachten we een bedrijfsbreed gegevensmodel dat vrij is van dubbelzinnigheden en redundanties over alle bedrijfsprocessen heen. In afbeelding 1 zijn er op het conceptuele niveau geen scheidingslijnen tussen de zuilen. Op dit niveau wordt alle relevante informatie immers opgesplitst in elementaire feiten, uitgedrukt in zinnen die geheel vrij zijn van niet-conceptuele informatie of constructie. Deze zijn daarom precies hetzelfde in iedere zuil.

## Praktijk

Laten we kijken hoe alles in zijn werk gaat in de praktijk. We beginnen met fase 1 van de Metadata Frame methode, de voorbereidende fase. Hierin vormen we een beeld van het totale gebruikersdomein, splitsen dit op in gebieden die gescheiden kunnen worden afgehandeld in een evolutionaire aanpak, en voorzien deze van een prioriteit om te bepalen welk gebied we het eerst afhandelen. Het resultaat is een spreadsheet met vijf lijsten:

1. De betrokken organisatie-eenheden;
2. De betrokken processen;
3. De vereiste en gewenste rapporten;
4. Alle relevante performance en key-performance indicatoren (PI's en KPI's);

5. Globaal vastgestelde onderwerpen die in verband kunnen worden gebracht met de (K)PI's, en kandidaat-dimensies zijn in een dimensioneel ontwerp.

De spreadsheet bevat verder een aantal matrices, die aangeven welke organisatie-eenheden, rapporten, (K)PI's, en onderwerpen betrokken zijn bij welke processen, en matrices die het verband aangeven van de (K)PI's met de rapporten en de kandidaat-dimensies. Dit geeft op voorhand een redelijk beeld van de Data Marts die naar voren zullen komen uit het datamodelleringsproces. In feite zijn deze matrices het equivalent van een matrix die Ralph Kimball de 'Datawarehouse Bus Architecture matrix' noemt [3]: een matrix van processen tegen kandidaat-dimensies. Daarom noemen we deze spreadsheet vaak de 'Bus matrices'. Fase 1 van de Metadata Frame methode lijkt op een klassieke informatieanalyse, maar het resultaat is geen formeel rapport, maar een helder beeld van alles wat we moeten modelleren in fase 2 van de methode.

Vaak is de omgeving gefragmenteerd met verschillende codes voor dezelfde dingen

Aan het eind van fase 1 toonde de lijst van rapporten voor het Erasmus MC de vereiste rapporten voor de Zorginspectie, rapporten over patiëntenlogistiek per afdeling en per specialisme, en ad hoc rapporten voor de medische informatieafdeling, de financiële afdeling en onderzoek. De lijst van PI's en KPI's was indrukwekkend lang, en bevatte standaard items voor patiëntenlogistiek, zoals aantal ligdagen, opnamen, ontslagen, overplaatsingen van en naar andere afdelingen, overnames van en door andere specialismen, polikliniekbezoeken, geannuleerde bezoeken, en niet-komers. Voor de IC's bevatte de lijst items zoals het aantal positieve bloedkweken, beademingsdagen en -uren, IC-opnamen, IC-ontslagen, sterfgevallen, afgezegde

Patient: P1234567		Voorzorgen: Allergieën voor medicamenten: Andere allergieën:						
Laboverzicht								
Alle	Alle		27/11/2005	27/11/2005	27/11/2005	27/11/2005	28/11/2005	28/11/2005
	Bepalingen	Referentiebereik	11:10	13:44	13:46	19:17	00:38	06:15
Algemeen	Opmerkingen	-	---					
Arteriële/veneuze bloedgas	pH (bloed)	7.35 - 7.45	---	7.34	7.36	7.36	7.36	7.40
Chemie algemeen	PaCO <sub>2</sub> (bloed)	4.7 - 6.4	kPa	4.9	4.6	4.6	4.7	4.3
Electrolyten	PaO <sub>2</sub> (bloed)	10.0 - 13.3	kPa	16.0	10.4	10.7	11.8	11.2
Hart enzymen	Actuele bicarbonaat	21.0 - 27.0	mmol/l	19.1	19.0	19.1	19.4	19.4
Hematologie diff	Base excess (bloed)	-3 - 3	mmol/l	-6		-5	-5	-4
Hematologie	Standaard bicarbonaat	21.0 - 27.0	mmol/l	19.8		20.1	20.3	20.8
Hemostase	Saturatie (bloed)	0.950 - 0.990	mol/mol	0.985		0.963	0.971	0.973
Leverfuncties								

Afbeelding 3: Voorbeelden van de resultaten van laboratoriummetingen uit het medische bronsysteem.

IC-datum	Uur (0-23)	Device-type	Beademings- uur starttijd	Beademings- uur eindtijd	Patient	Afdeling	Beademings startdatum	Beademings starttijd	Aantal uren	Start beademings- periode	Einde beademings- periode
18/11/2005	23:00	BA1	23:35	00:00	1111111	ICP1	18/11/2005	23:35	0.42	1	0
19/11/2005	0:00	BA1	00:00	01:00	1111111	ICP1	18/11/2005	23:35	1	0	0
19/11/2005	1:00	BA1	01:00	02:00	1111111	ICP1	18/11/2005	23:35	1	0	0
19/11/2005	2:00	BA1	02:00	02:22	1111111	ICP1	18/11/2005	23:35	0.37	0	0

**Afbeelding 4:** Ontworpen concrete voorbeelden van performance indicatoren voor beademing.

operaties, eerste hulp bypass-uren, bloedtransfusies, medicaties, en scores zoals de APACHE score<sup>2</sup> en TISS-score<sup>3</sup>. Als kandidaat-dimensies kwamen naar voren: datum, tijd, patiënt, afdeling, specialisme, specialist, behandelend arts, diagnose, type apparaat, medische staf, medicatie, etcetera. Tenslotte maakten we ook een lijst van attributen die voorkwamen op de vereiste rapporten voor de Zorginspectie.

Fase 2 van onze methode omvat de echte gegevensmodellering met behulp van FCO-IM [4]. Dit vereist het verzamelen en ontwerpen van concrete voorbeelden van alle relevante informatie die in fase 1 is geïdentificeerd. Afbeelding 2, 3 en 4 tonen voorbeelden daarvan voor de meting van real-time variabelen, laboratoriummetingen, en beademing. De eerste twee komen direct uit het medische bronsysteem. De derde is een voorbeeld dat speciaal voor het modelleren werd ontworpen op basis van analyse van de in fase 1 gevonden PI's en KPI's voor beademingsuren. Ook hier leverden onze medische experts sterke en beslissende bijdragen aan de analyse. Voor patiëntenlogistiek en de IC's samen verzamelden we 15 verschillende sheets met concrete voorbeelden uit de medische bronsystemen en ontwierpen we nog eens 28 sheets om alle relevante informatie in de spreadsheet van fase 1 af te dekken.

De volgende stap in het proces bestaat uit het verwoorden van alle relevante feittypen die in de concrete voorbeelden te zien zijn. Dit gebeurde samen met als medische experts: de Kinderarts-intensivist en Chief Medical Information Officer, Dr. Jan Hazelzet, een IC-verpleegkundige, Guido Lansbergen, en een promovendus, Marleen de Mul. Voor wat betreft de feiten zichtbaar in afbeelding 2 waren de experts het met elkaar eens, dat deze correct verwoord waren met zinnen als:

- Er is bij Patiënt P1234567 op 21-11-2005 om 9:00:00 uur voor Hartfrequentie een waarde van 112 slagen/min. vastgelegd.
- Er is bij Patiënt P1234567 op 21-11-2005 om 7:00:00 uur voor Temperatuur een waarde van 38.2 °C vastgelegd.

De feiten zichtbaar in afbeelding 3 werden volgens de experts correct weergegeven door zinnen als:

- De labbepaling bij Patiënt P1234567 op 28-11-2005 om 00.38 uur voor Standaard bicarbonaat (bloed) had als resultaat 20.8 mmol/l.
- De referentie ondergrens van Patiënt P1234567 op 28-11-2005 om 00.38 uur voor Standaard bicarbonaat (bloed) was 21.0 mmol/l.

- De referentie bovengrens van Patiënt P1234567 op 28-11-2005 om 00.38 uur voor Standaard bicarbonaat (bloed) was 27.0 mmol/l.

Voor de feiten zichtbaar in afbeelding 4 tenslotte, keurden de experts zinnen als deze goed:

- Het beademingsuurfragment van patiënt 1111111 beginnend op 18-11-2005 om 23:35 uur hoort bij het uur beginnend om 23:00 uur.
- Tijdens het beademingsuurfragment van patiënt 1111111 beginnend op 18-11-2005 om 23:35 uur werd devicetype BA1 gebruikt.
- Tijdens het beademingsuurfragment van patiënt 1111111 beginnend op 18-11-2005 om 23:35 uur werd de patiënt verzorgd op afdeling ICP1.
- Markeert het beademingsuurfragment van patiënt 1111111 beginnend op 18-11-2005 om 23:35 uur de start van een Beademingsperiode? 1.

Zodra de feiten geformuleerd waren, bepaalden we de interne structuur van deze feiten. Ook hierbij konden de medische experts ons uitstekend helpen. De uitwerking daarvan leest u in het tweede deel van dit artikel in Database Magazine 7.

#### Noten

1. Bij gebruik van Class Diagrams kan men hier ook UML-niveau lezen.
2. Acute Physiology And Chronic Health Evaluation: een maat voor de overlevingskans van patiënten, berekend aan de hand van een verzameling fysiologische variabelen verzameld tijdens de eerste 24 uur op de IC.
3. Therapeutic Intervention Scoring System: een getal dat de werkdruk van de IC verpleegkundige staf aangeeft.

#### Literatuur

1. Database Magazine, Jaargang 11, no. 8 (december 2000), P.W.F. Alons.
2. Database Magazine, Jaargang 12, no. 1 (januari 2001), P.W.F. Alons.
3. The Datawarehouse Lifecycle Toolkit, Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite, (Wiley Computer publishing, 2002), ch. 7.
4. Volledig Communicatiegeoriënteerde Informatiemodellering, G. Bakema, J.P. Zwart, H. van der Lek, Kluwer BedrijfsInformatie, 2002.

#### Peter Alons

Dr. P.W.F. Alons (peter.alons@atosorigin.com) is senior consultant bij Atos Origin/BI-CRM en al ruim vijftien jaar betrokken geweest bij een groot aantal BI- en Datawarehouse-projecten bij diverse bedrijven in diverse branches.