



Demystifying metadata & metamodel

Plan van Aanpak

Radboud Universiteit Nijmegen

Naam: Kim Cratsz
Studentnummer: 0419028
Faculteit: Faculteit der Natuurwetenschappen, Wiskunde en Informatica
Instituut: Nijmeegs Instituut voor Informatica en Informatiekunde
Datum: 10 augustus 2006
Versie: Versie 1.0
Afstudeernummer: 31IK

Inhoudsopgave

1. Inleiding	2
1.1 Achtergrond.....	2
1.2 Doelstelling.....	2
1.2.1 Probleemdefinitie en Vraagstelling	3
1.2.3 Onderzoeksvragen	3
1.4 Aanpak onderzoekstraject	3
2. Globale Planning.....	4
3. Referenties	5

1. Inleiding

In dit hoofdstuk zullen de probleemstelling (waartoe), doelstelling (waarom) en de vraagstelling (wat) aan bod komen.

1.1 Achtergrond

We leven in een tijd waar de technologie niet stil staat, maar technologieën voor het managen van de informatie is vaak het probleem en niet de oplossing. Waaraan ligt dit precies? We zijn in staat om wolkenkrabbers van meer dan 500 meter hoog en vliegtuigen te maken. Maar toch kunnen we niet eens zo iets simpels als informatie managen. De beschikbare informatie wordt niet ten volle benut. Goede informatie genereren en omzetten tot kennis is niet een eenvoudige taak en het informatieaanbod zal de komende jaren alleen maar blijven toenemen.

Met behulp van modelleringstechnieken wordt er geprobeerd om de processen en informatiestromen in organisaties beter in kaart te brengen en op deze manier software applicaties te ontwikkelen die geschikt zijn voor het genereren van de gewenste informatie. In de literatuur wordt er veel aandacht besteed aan modelleertalen en modelleertechnieken, maar hoe deze talen en technieken samen gebruikt kunnen worden om het probleem op te lossen is niet te vinden in de literatuur. Vooral nu er een teveel is aan technieken en talen is het nodig om een oplossing te vinden. Om de gewenste informatie te kunnen genereren zijn de data, de modelleertalen en de technieken van belang. Deze drie moeten niet afzonderlijk van elkaar bekeken worden, maar samen in de vorm van metadata (data) en metamodelen (techniek en taal).

Tegenwoordig worden de termen metadata en metamodel vaak gebruikt in de IT wereld. Het is bijna een hype geworden, maar deze termen zijn niet nieuw. Metamodelen en metadata zijn samen de fundamentele voor data integratie. Maar vaak worden deze twee termen verkeerd gebruikt of verkeerd geïnterpreteerd. Hierdoor weten bijvoorbeeld organisaties niet wat de voordelen zijn die met deze termen behaald kunnen worden.

1.2 Doelstelling

Door gebruik te maken van metadata en metamodelen kan de interoperabiliteit en integratie van (legacy) systemen vergemakkelijkt worden. Informatie is heel erg belangrijk binnen organisaties. Zonder informatie (gegevens) kan een organisatie niet bestaan. Met behulp van metamodelen en metadata oplossingen kunnen relevante informatie binnen en buiten organisaties (World Wide Web), makkelijker ontdekt worden en hierdoor beter gebruikt worden door organisaties. De uitwisseling van modellen (metadata) en van metamodelen (meta-metadata) is mogelijk tussen producten die dezelfde meta-metamodelen hebben.

Het doel van het onderzoek is om de termen metadata en metamodelen dichterbij elkaar te brengen zodat de samenhang tussen de termen als basis kan dienen voor toekomstige ontwerpen met bijvoorbeeld interoperabiliteit en integratie problemen.

1.2.1 Probleemdefinitie en Vraagstelling

Het is bekend dat organisaties de beschikbare informatie niet ten volle benutten en het genereren van goede informatie is niet een eenvoudige taak. Hoe kunnen metadata en metamodelen hierbij helpen? In de literatuur is van beide termen wel informatie en onderzoeken te vinden, maar vaak worden deze termen verkeerd geïnterpreteerd en ook afzonderlijk van elkaar behandeld. Het is nog niet duidelijk wat men hiermee kan doen.

Hoe verwacht zijn de termen data, informatie, metadata, modellen, metamodelen en kennis? Weten gebruikers, administrators, modellers, managers enz. wat deze termen precies betekenen en hoe ze met elkaar in verband staan? Het is hierdoor belangrijk om te weten wat de voordelen zijn die behaald kunnen worden door *goed* gebruik te maken / toepassen van deze termen. Maar om deze termen goed toe te kunnen passen is het eerst nog belangrijker om te weten wat deze termen betekenen.

De centrale vraagstelling voor dit onderzoek luidt: *“Hoe kunnen metadata en metamodelen als basis dienen voor het genereren van de gewenste informatie?”*

Het resultaat van dit onderzoek is een overkoepelend model waar de samenhang tussen deze termen (data, metadata en metamodelen) zal worden weergegeven.

1.2.3 Onderzoeksvragen

Voor de beantwoording van de centrale vraagstelling in deze scriptie is het nodig om de hoofdvraag in kleine onderzoeksvragen onder te verdelen. Hieronder volgen de onderzoeksvragen:

- Wat zijn data, metadata, model en metamodel?;
- Wat is een mogelijke definitie voor de termen metadata en metamodel?;
- Hoe worden deze termen momenteel toegepast in de praktijk?;
- Bestaat er een relatie tussen deze termen?;
- Wat is de relatie tussen deze termen?;
- Hoe kan het verband, de relatie tussen deze termen worden weergegeven?

1.4 Aanpak onderzoekstraject

Het doel van dit onderzoek is om de termen data, metadata, model en metamodel dichterbij elkaar te brengen in de vorm van een overkoepelend model. Om dit te kunnen bereiken zal er gebruik worden gemaakt van het vier lagen model van OMG. Dit vier lagen model dient als basis voor het aangeven van de samenhang. Als eerste zullen de termen afzonderlijk van elkaar behandeld worden, zodat de inhoudelijke betekenis begrepen kan worden. Dit zal als volgt plaats vinden; als eerste zal er van elke term een

algemene inleiding gegeven worden met een definitie. Als tweede zal er van elke term toepassingsvoorbeelden gegeven worden die momenteel in de praktijk gebruikt worden. Op technisch niveau zullen er ook formele onderbouwingen plaats vinden. Vervolgens zullen de termen met behulp van het vier lagen model bij elkaar gebracht worden in één groot model, het overkoepelende model.

2. Globale Planning

Hieronder volgt de globale planning voor dit onderzoekstraject. De planning is verdeeld over 28 weken. De volgenden activiteiten worden onderkend:

- **Literatuurstudie:**
Het raadplegen van gevonden bronnen zoals wetenschappelijke artikelen, boeken en Internet. Deze activiteit houdt in dat de artikelen gesorteerd en gelezen worden. De literatuurstudie vindt plaats gedurende het hele traject.
- **Literatuur verwerken:**
De gevonden en relevante literatuur wordt in de scriptie verwerkt.
- **Tussentijdse voortgang:**
Tijdens het onderzoekstraject vinden een aantal gesprekken plaats met de begeleider om de voortgang te bespreken. In de gesprekken wordt het werk geëvalueerd en indien nodig aangepast.
- **Afronding:**
Afronding van het onderzoek. Oplevering van de scriptie en eventueel andere documenten.

Weeknr \ Activiteit	Feb				Mrt				Apr					Mei					Jun			Jul			Aug		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	22	23	24	25	26	27	28
Literatuur studie																											
Literatuur verwerken																											
Tussentijdse voortgang																											
Afronding																											

3. Referenties

- [1] Bakker, J., G., M., Data Management is Metadata Management Trendrapportage Data Management Rijkswaterstaat, 3 juni 2002, <http://www.wadi.nl/datamanagementismetadatamgt.html>
- [2] Bentley, John., E., First Union National Park, Metadata: Everyone talks about it, But what is it?, Datawarehousing and Solutions.
- [3] Brand, Amy, Daly, Frank en Meyers, Barbara, Metadata Demystified, A guide for publishers, Published by NISO Press, National Information Standards Organisation, 2001, www.niso.org
- [4] Brasethvik, Terje, A Semantic Modeling approach to Metadata, Department of Computer and Information Science, IDI, Norwegian University of Science and Technology, NTNU, April 1998.
- [5] Breton, E., Bézivin, J., Towards an understanding of Model Executability, Ogunquit, Maine, USA, October 2001.
- [6] Clark, T., Evans, A., Sammut, P., Willans, J., Applied Metamodelling: A foundation for language Driven Development Version 0.1, www.xantium.com, 21 Dec 2005
- [7] Coles, C., Cabinet office, e-Government metadata, Metadata & search presentations, Contentmanagement for the public sector, 23 november 2004.
- [8] Date, C., J., An introduction to Database Systems, sixth edition. SOFSEM'98 Theory and Practice of Informatics: 25th Conference on Current Trends in Theory and Practice of Informatics, Jasna, Slovakia, November 21-27, 1998 Proceedings (Lecture Notes in Computer Science S.), 60-62
- [9] Elmasri, R., Navathe, S., B., Fundamentals of Database Systems, Third Edition, Addison-Wesley, December 5th, 2002, 569-583.
- [10] Gaarder, J., Sophie's World, Farrar, Straus and Giroux Inc, 1994.
- [11] Grabt, J., Litwin, W., Roussopoulos, N., d Sellis, T. 1993. Query languages for relational multidatabases. VLDB J. 2, 153-171.
- [12] Halpin, Terry, An ORM Metamodel, Journal of Conceptual Modeling, October 2000, issue 16
- [13] Halpin, terry, An ORM metamodel of Barker, Journal of Conceptual Modeling, December 2000, issue 17.
- [14] Henderson-Sellers, Brian, Atkinson, C., Kühne, T., Gonzalez-Perez, C., Understanding metamodelling, ER 2003 15 october 2003.
- [15] Heylighen, F., Complexity and Information Overload in society: Why increasing efficiency leads to decreasing control, April 12, 2002, The information society, <http://66.249.93.104/search?q=cache:7Ismf4nNZg8J:pespmc1.vub.ac.be/Papers/Info-Overload.pdf+information+overload+%2B+waddington&hl=en&ct=clnk&cd=9&client=firefox-a>
- [16] Hodge, Gail, Metadata Made Simpler, Published by NISO Press National Information Standards Organisation, 2001, www.niso.org

Plan van Aanpak: Demystifying Metadata & Metamodel

- [17] Hodge, Gail, Understanding metadata Published by NISO Press National Information Standards Organisation, 2001, www.niso.org
- [18] Jernst, what are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology and a meta-model?, January 15th, 2003, www.metamodel.com
- [19] Lakshmanan, L.V.S., Sadri, F., en Subramanian, S.N. 1997. Logic and algebraic languages for interoperability in multidatabase systems*, The journal of logic programming.
- [20] Lakshmanan, L.V.S., Sadri, F., en Subramanian, S.N. 2001. SchemaSQL-An extension to SQL for multidatabase interoperability. ACM trans.Datab.Syst. 26, 4 (Dec), 476-519.
- [21] Lans, R., F., van der. 1999. Het SQL leerboek. 65-67, 385-389.
- [22] Oei, J., L., H., en Hemmen, L., J., G., T., van. en, Falkenberg, E., D., en, Brinkkemper, S., The Meta Model Hierarchy: A Framework for Information Systems Concepts and techniques. July 29th, 1992. 2-21.
- [23] OMG, Meta Object Facility (MOF) specification, version 1.3, new edition march 2000, 10 mei 2006, www.omg.org.
- [24] Pelzer, T., System Tables, 12 juni 2006, <http://www.dbazine.com/db2/db2-disarticles/pelzer4>
- [25] Proper, E., Weide, Th., P., van der, Modelling as Selection of Interpretation, Institute for Computing and Information Sciences, Radboud University of Nijmegen.
- [26] Seiner, S., Robert. DBAs Don't need Meta Data. TDAN.com Issue7.0 Articles – December 1998.
- [27] Seiner, S., Rober. Meta Data Themes: The Basics. The Data Administrator Newsletter, June 1998.
- [28] SIMIN, Begrippenlijst metadata, 17 april 2006, <http://www.simin.nl/index.php>
- [29] Smit, Frans, Het Historisch Data Warehouse, February 2002, <http://www.cultivateint.org/issue6/warehouse-d/>
- [30] Song, H., Metamodels for Object Role Modeling., march 2005.
- [31] Tannenbaum, Adrienne, Metadata Solutions, Using Metamodel, Repositories, XML and Enterprise Portals to Generate Information on Demand, Addison-Wesley, 2nd printing, February 2006.
- [32] Waddington, P., Dying for information? A report on the effects of information overload in the UK and WorldWide, Reuters united Kingdom, 1996, <http://www.cni.org/regconfs/1997/ukoln-content/repor~13.html#34>
- [33] Website British Library, Exchange Formats, 17 April 2006, <http://www.bl.uk/services/bibliographic/exchange.html>
- [34] Website Codd's 12 rules, 12 juni 2006, http://en.wikipedia.org/wiki/Codd%27s_12_rules

Plan van Aanpak: Demystifying Metadata & Metamodel

- [35] Website Obituary: Dr. Edgar F. Codd, 12 juni 2006, http://66.249.93.104/search?q=cache:NvMiVDcQ36YJ:www.nao.org.uk/intosai/edp/intoit_articles/18p60top62.pdf+Dr+F+edgar+Codd+%2B+12+rules&hl=en&ct=clnk&cd=2&client=firefox-a
- [36] Website Dublin Core, 17 april 2006, <http://dublincore.org>.
- [37] Website, EXIF metadata, <http://exif.org/specifications.html>
- [38] Website, Federated database system, 26 juli 2006, http://en.wikipedia.org/wiki/Federated_database
- [39] Website Glossary, Define Integration, 17 april 2006, http://searchcrm.techtarget.com/sDefinition/0,,sid11_gci212359,00.html
- [40] Website IBM Federated Database Technology, 13 juni 2006, <http://www-128.ibm.com/developerworks/db2/library/techarticle/0203haas/0203haas.html>
- [41] Website ID3v2, 10 mei 2006, <http://www.id3.org/>
- [42] Website Metadata, 9 maart 2006, <http://en.wikipedia.org/wiki/Metadata>
- [43] Website METS, Metadata Encoding & Transmission Standard, 17 April 2006, <http://www.loc.gov/standards/mets/>
- [44] Website Semantics, www.whatis.com
- [45] Website TDAN.com Issue 17.0 Articles, Tannenbaum, Adrienne, A Disaster Crying for Solutions, juli 2001, www.tdan.com
- [46] Website, UML Semantics, 4 juli 2006, <http://etna.int-evry.fr/COURS/UML/semantics/semant2.html>
- [47] Website Unmanaged metadata, 3 mei 2006, <http://www.e-consultancy.com/newsfeatures/22715/unmanaged-metadata-threatens-quality-of-corporate-decision-making-says-ovum.html>
- [48] Website, What is metamodeling and what is it good for, <http://www.metamodel.com>
- [49] Website, WinFS overview, http://www.ntfs.com/winfs_basics.htm
- [50] Wyss, M., Catharine. and Robertson, L., Edward. 2005. Relational Languages for Metadata Intergration. ACM Transactions on Database Systems, Vol. 30, No. 2, June 2005, 624-600
- [51] Zhao, J., L., Schema coordination in federated database management: a comparison with schema integration., Decision Support Systems 29, (1997), 243-257.