Capturing Modeling Processes – Towards the MoDial Modeling Laboratory

S.J.B.A. (Stijn) Hoppenbrouwers, L. (Leonie) Lindeman, and H.A. (Erik) Proper

Institute for Computing and Information Sciences, Radboud University Nijmegen
Toernooiveld 1, 6525 ED Nijmegen, The Netherlands, EU
S.Hoppenbrouwers@cs.ru.nl, E.Proper@cs.ru.nl

Abstract. This paper is part of an ongoing research effort to better understand the process of conceptual modeling. As part of this effort, we are currently developing a modeling laboratory named MoDial (*Mo*deling *Dial*ogues). The main contribution of this paper is a conceptual meta-model of that part of MoDial which aims to capture the elicitation aspects of the *modeling process* used in creating a model, rather than the model as such.The current meta-model is the result of a two-stage research process. The first stage involves theoretical input from literature and earlier results. The second stage is concerned with (modest) empirical validation in terms of interviews with modeling experts.

1 Introduction

Conceptual modeling is at the core of information systems engineering. In view of deliverables produced during information systems engineering, in [12] a distinction is made between *usage world*, *subject world*, *system world* and *development world*. Understanding each of these worlds requires considerable modeling efforts to define the requirements on the system and to produce the design of a system. The work reported in this paper is part of an ongoing effort by our group to better understand the act of conceptual modeling [6,] in the context of information system engineering. One of our longer term goals is to turn the *art* of modeling into a *science* of modeling.

As put forward during the panel session of the ORM 2005 workshop, one of the strategies of our research group is to be able to more explicitly capture conceptual modeling cases, where the aim is to not only include the resulting models, but also to capture the modeling process leading up to the model (and the strategies shaping it; [9]). To better understand modeling processes, and to also be in a position to gather direly needed empirical data on the details of modeling processes, we are developing a modeling laboratory called MoDial, which stands for *Mo*deling *Dial*ogues.

In line with our earlier results [8, 7] we regard modeling processes as involving two related dialogues: for *elicitation* and for *formalization*. The elicitation dialogue takes place between an informer (presumably a domain expert) and a model mediator (typically an information analyst). The formalization dialogue, on the other hand, takes place between the model mediator and the model builder (usually some tool

R. Meersman, Z. Tari, P. Herrero et al. (Eds.): OTM Workshops 2006, LNCS 4278, pp. 1242 – 1252, 2006.

used to capture and verify the actual model). In the past we have already provided a theoretical model for the formalization dialogue [6] based on preliminary work reported in [14]. This dialogue is characterized by the use of strict controlled language [3] which lends itself very well to a transformation towards a formal language. In the elicitation dialogue between an informer and the model mediator, however, the use of a strict controlled language for making initial statements is not very realistic. The properties of the dialogue language is dictated by the abilities of the informants to express themselves (and for model mediators to validate their understanding of the informants' statements) rather than by the needs of formalization.

The work reported in this paper is primarily concerned with the *elicitation* dialogue. Many important modeling decisions emerge, or are at least resolved, in this dialogue rather than in the formalization dialogue. We eventually want to capture modeling decisions and the processes in which they take place. Our aim here is therefore to derive a meta-model of the concepts needed to capture elicitation-related parts of the modeling processes for use in a modeling laboratory such as MoDial.

The current meta-model as reported in this paper is the result of an MSc research project [11], and has been created in two stages. In the first stage, an abstraction was made of three pre-existing conceptual modeling procedures described in literature. During the second stage, three experienced modelers were interviewed in order to validate the meta-model resulting from the first stage. In the remainder of this paper, we consecutively discuss these stages, followed by a discussion of the resulting model and planned follow-up research activities.

2 First Stage – Described Modeling Procedures

Several descriptions of conceptual modeling processes exist. Each of these documented procedures can be regarded as a guidebook for the creation of a specific kind of conceptual model. Such guidebooks are typically based on commonly used modeling practices. Below we will discuss three procedures, which will then be generalized. The three procedures compared are aimed at the creation of ORM diagrams [4], UML class diagrams [13] and ER diagrams [1], respectively.

2.1 ORM Diagrams

Object-Role Modeling (ORM) is a fact oriented method, which has initially been developed for modeling information systems at a conceptual level. ORM makes use of natural language statements by examining them in terms of elementary facts. In [4] a *Conceptual Schema Design Procedure* (CSDP) is provided, consisting of seven steps:

1. Transform familiar information examples into elementary facts, and apply quality checks.
2. Draw the fact types, and apply a population check.
3. Check for entity types that should be combined, and note any arithmetic derivations.

4. Add uniqueness constraints, and check arity of fact types.
5. Add mandatory role constraints, and check for logical derivations.
6. Add value, set comparison and sub-typing constraints.
7. Add other constraints and perform final checks.

2.2 UML Class Diagrams

The Unified Modeling Language (UML) is an object modeling and specification language that is designed to make object oriented analyses and designs for information systems. UML 2.0 defines thirteen types of diagrams, divided into three categories. The UML class diagram belongs to the category of "structure diagrams" [13]. The UML, being a language, does not provide a procedure for the creation of UML class diagram. However, procedures do exist for creation of a class diagram. We have used the procedure described in [15]. This procedure has the following outline:

1. Identify all possible candidate classes.
2. Select classes from the list of candidates.
3. Make a model dictionary.
4. Identify associations.
5. Identify attributes.
6. Identify operations.
7. Generalize with the help of inheritance.
8. Add business rules with Object Constraint Language (OCL).
9. Divide classes into packages.
10. Repeat over executed steps.

2.3 ER Diagrams

The Entity Relationship (ER) model was introduced in 1976 by Peter Chen [1]. There is no generally accepted standard for ER, but there are different corresponding components that exist in most of its variants. An ER diagram is the graphical result of the ER model and shows entities and the coherence between them [2]. Various guidelines exist for the creation of an ER diagram. We have used the guidelines as provided in [5]. Before this procedure can be followed, information has to be collected: functional requirements, forms, reports or existing models. The procedure starts from studying facts or sentences:

1. Identify possible entity types.
2. Identify possible relationship types between identified entity types.
3. Determine cardinalities of the relationship types.
4. Identify and associate attributes with entity types.
5. Identify and associate attributes with relationship types.
6. Determine domains of attributes.
7. Determine potential identifiers for every entity type.
8. Think about the use of specialization/ generalization (sub-types/ super-types).
9. Check the model on redundancy.
10. Validate the ER model.
11. Plan a review about the model with users.

2.4 Generalization of Modeling Procedures

In [11], the above procedures were scrutinized to infer what kind of information was handled in each of the involved steps, leading to an initial meta-model. This meta-model has been validated by populating it with a few examples of models and modeling sessions. Generalizing, we observed that each of the three modeling procedures essentially proceeds through six stages:

1. Identify requirements-on/goal-of the model to be produced.
2. Identify the modeling language to use.
3. Select and gather sources.
4. Scope the domain.
5. Engage in a question and answers process with an informer(s).
6. Conduct final quality checks.

Needless to say, iteration between these stages is likely to occur. We will now briefly discuss the stages, providing an indication of the kind of information that will be involved in capturing the part of the modeling process under concern.

Before a modeling process starts, a valid goal for the creation of the model should be identified [8] and the requirements on the model to be produced. Once the requirements on the model have been identified, it is important to choose the right kind of model. Every artifact has its own strengths and weaknesses, and every kind of model is only applicable in particular situations.

All three procedures require collection of information about the Universe of Discourse before actual modeling starts. There are different ways to do this, for example by gathering reports, forms, tables, diagrams, or texts. When a description of the problem is available, then this is also a useful source of information.

The next stage is to take decisions about the concepts that play a part in the domain, also known as scoping the domain. After scoping the domain, the determination of the relevant concepts occurs, e.g. by organizing modeling sessions. This is where we enter the elicitation dialogue between an informant and the model mediator [6, 7, 9]. This dialogue is dominated by a "game" of questioning and answering. The informant is a domain expert who harbors knowledge of some domain, and is presumed to know about the target domain (or can find out more if needed). It is assumed that domain experts can express and validate statements about the domain in a language which is suitable to *them* (probably disqualifying, for example, to the strict controlled language used during the formalization dialogue). Statements can take the form of free-hand drawing, (partial) diagrams, semi-formal expressions, texts, etc. The model mediator is not required to have any specific knowledge of the domain, but is assumed to know how to create a verifiably correct model [6].

The final stage of the modeling process is a quality check of the model. The question that needs to be answered here is whether the model produced fits the requirements/goals as stated in the first stage. The procedures discussed propose different checks for redundancy, derivations, consistency and completeness, and to validate the model. When a problem is found, the modeler can choose to update the model at that point, or to accept the model the way it is now. In some situations, a model is judged "good enough", and need not be perfect.

Our initial meta-model was refined on the basis of (modest) empirical validation, as described in the next section. We will not present the initial model, only the meta-model resulting after the round of interviews of modeling experts (see Section 4).

3 Second Stage – Experiences from Practitioners

Using as a starting point the meta-model derived from the first stage, interviews were conducted with three expert modelers. To ensure that these based their input on their own experiences, and not on company standards, independent consultants/architects where selected as interviewees. Each of them has ample experience in information and/or knowledge modeling:

1. Mr. De Vries, Zetetic, http://www.zetetic.nl
2. Mrs. Bleeker, Gemara-Consulting, http://www.gemara-consulting.nl
3. Mr. Crompvoets, Bommeljé Crompvoets and partners, http://www.bcp-software.nl

3.1 Interview with Mr. de Vries

Mr. de Vries specializes in consultancy and architecting in the field of knowledge management and knowledge technology. In these assignments, he produces many models, for example of processes, information flows and information systems. In his experience, the first step of a modeling process is trying to clarify what it is he's hired for. This allows him to so he can identify the goal of the model. Next, Mr. de Vries chooses a suitable modeling technique. This selection depends on both the goal of the model and the techniques already used by the organization.

Mr. de Vries usually organizes workshops with people who are involved with some process, system, or specific domain. At the beginning of the modeling process these workshops are at an overall level and aim to discover which concepts matter. In later sessions the understanding of key concepts is deepened. To get the concepts clear, it is important to talk to the right stakeholders, so these have to be identified at the beginning of the modeling process.

When an issue arises during a workshop a few scenarios can be thought up which potentially resolve it. The issue is then 'put on hold' and the stakeholder commissioning the model is requested to select the desired scenario.

3.2 Interview with Mrs. Bleeker

Mrs. Bleeker is an independent information architect and requirements engineer, who uses models to clarify which information or business concepts play a part in a domain. She uses ORM, primarily as a way of structuring her thoughts. For documenting, however, UML and natural language are also used.

In modeling sessions, Mrs. Bleeker prefers interviews to workshops. In interviews Mrs. Bleeker frequently uses common interviewing techniques, such as summarizing and a drill-down style of questioning. Her general way of working is to take a top-down approach. She first determines the boundaries of the domain and sub-domains. This is followed by an elaboration on the identified sub-domains.

Mrs. Bleeker enquires with the commissioning party to discover promising informants. However, these are generally the busiest people around, so sometimes she has to make do with people who know less about the domain. Besides interviewing domain experts, she also discusses the key issues with the decision makers. This way she ensures commitment to the resulting model from these decision makers.

Before starting a modeling process, Mrs. Bleeker first determines the goals for producing the model. In addition, she determines beforehand the way in which the model will be documented. She has a number of standard ways of documenting that work well for her. However, some projects need a creative approach to documenting.

Mrs. Bleeker stops modeling when she has a clear overview of the domain. When there is little time available, she aims to achieve as much clarity about the domain as possible in the given time. In such cases, Mrs. Bleeker also speaks explicitly about the ambition level with regards to the depth and breadth of the resulting model. It is not always possible to produce the 'perfect' model in view of resources available. These constraints are added as disclaimers to the documentation accompanying the models.

3.3 Interview with Mr. Crompvoets

Mr. Crompvoets is consultant and uses various techniques to create models, but his choice of technique depends on the organization an situation in which the project takes place. He often uses modeling sessions to adjust concepts and to create a common view. Before starting a modeling session, he usually studies some documents, so he can steer a session as he sees fit.

A modeling session starts by shaping a picture of the strategy and the vision of the future of the company. After that, he focuses on the products and services of the company. He zooms in more and more, paying attention to the different views of the participants, so every concept becomes clear for every participant. Asking questions is very important to realize clarity. When a concept is not clear for everyone, discussions arise, unveiling different interpretations of a concept. Once a concept is clear for (enough) participants, the next concept is discussed.

4 Resulting Model

The resulting meta-model is presented as an ORM model, divided in three parts. The partial meta-model shown in Fig. 1 concerns the contextual aspects of a modeling process. A modeling process is presumed to be initiated by some stakeholder, aims to produce a model, at a certain level of ambition, in order to meet some goal that has a pre-determined quality check, making use of some modeling technique, limiting itself to some scope, and possibly using additional documentation. The model produced during a modeling process must indeed be a model that is allowed in the selected modeling technique. Note that the latter relationship is a derived one (see Fig. 3).

The partial meta-model depicted in Fig. 2 deals with actual modeling processes. A modeling process is regarded as consisting of a number of modeling sessions. Each modeling session involves a number of session involvements: the actual involvement of some actor in a particular modeling session. Currently we identify three kinds of involvement: *informing*, *mediating* and *deciding*[1]. The informing and mediating involvement kinds correspond to what has been stated in the introduction about the elicitation part of the modeling process. The deciding role has been added based on input from the interviewees. As it turns out, during a modeling session, several issues arise that cannot be answered within the session. An external party is needed to answer these questions. Usually the stakeholder initiating the modeling process needs to be consulted to resolve such issues.

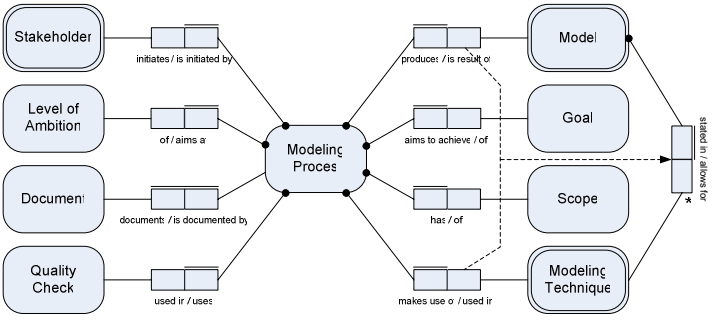

Fig. 1. Modeling context

The involvement of the actors in a modeling session leads to a number of questions, and statements answering these questions. In capturing modeling processes we are not only interested in questions put forward by the mediator, but also by questions put forward by the informers. We therefore allow all kinds of session involvements to be the originator of questions and answers.

Questions can not only be posed, but can be withdrawn as well. However, we would want to record withdrawals explicitly as they constitute interesting junctions in a modeling process. Statements are either made in response to questions, are a position statement, or are based on a decision made by a deciding involvement. Furthermore, each proposed statement can be accepted, rejected, or withdrawn during the modeling process.

The area enclosed by the rectangle adorned with a clock identifies that for the enclosed area a history needs to be maintained. For example, the order in which questions are put forward and answered is relevant to us.

The final partial model is depicted in Fig. 3. It concerns models, modeling techniques, and meta-models. The approach taken is this fragment aims to be neutral

[1] Note that the open circle at the base of the sub-type arrow signifies it to be a "self defining" sub-type. This can best be regarded as a graphical abbreviation.

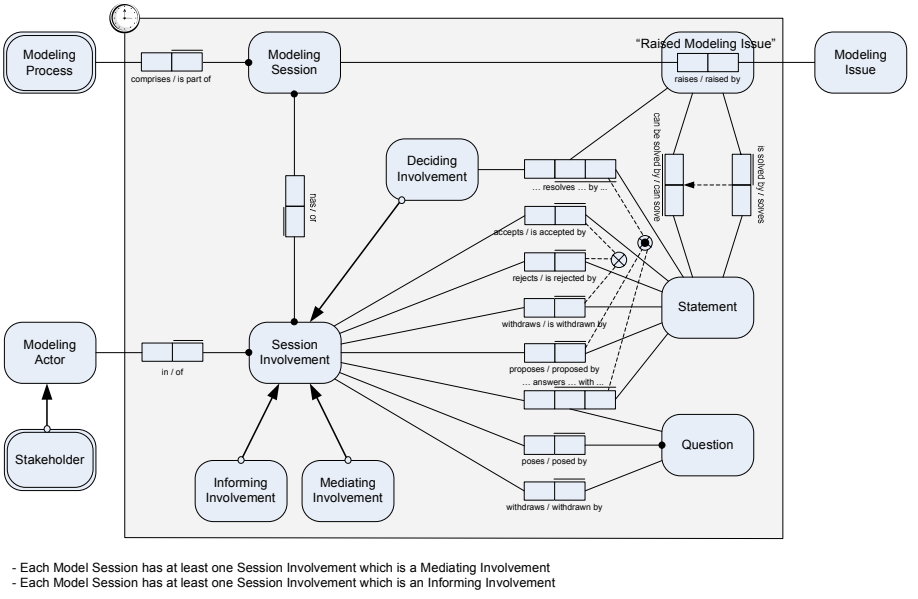

Fig. 2. Modeling dialogue

with regard to specific modeling languages. A model is presumed to comply with a meta-model, while a modeling technique has an underlying meta-model. We assume meta-models to consist of a set of meta-types, while a model is presumed to consist of model elements (which can be decomposed). The model elements are tied to the meta-model by typing the elements in terms of meta-types.

The model elements, which are the building blocks of the models being produced, are implied by the statements produced in the elicitation dialogue. In the formalization dialogue, the relationship between these statements and the model elements can be made much more rigid [6, 7], and can to some extend even be derived automatically.

5 Conclusions and Further Research

We presented a first version of the part of MoDial's meta-model which deals with the elicitation dialogue between informers and model mediators. The main focus of the meta-model was the process of modeling rather the creation of the actual model in isolation.

The meta-model resulted from a two-stage development: a theory-based stage (literature) and a practice-based stage (modest empirical work in terms of interviews and discussions with practitioners). Further theoretical embedding and empirical

validation is yet called for. The current schema is merely a first result. As a next step we aim to:

1. Enhance the meta-model with strategies for modeling processes. The many modeling "work-flows" as suggested by the guidebooks that are available on modeling essentially constitute a pre-defined work-flow based on some strategy. We aim to make these strategies in modeling explicit in our modeling laboratory. We also aim for dynamic, situation-specific workflow development.
2. The current meta-model also lacks elaborate motivations for modeling decisions. We hope to perform more empirical validation of the model through more interviews with modeling experts, but also by using the meta-model during modeling sessions in the laboratory. Minimally, the model will only be used as a documentation standard, but we hope it will be a key structure in the MoDial system we are developing.

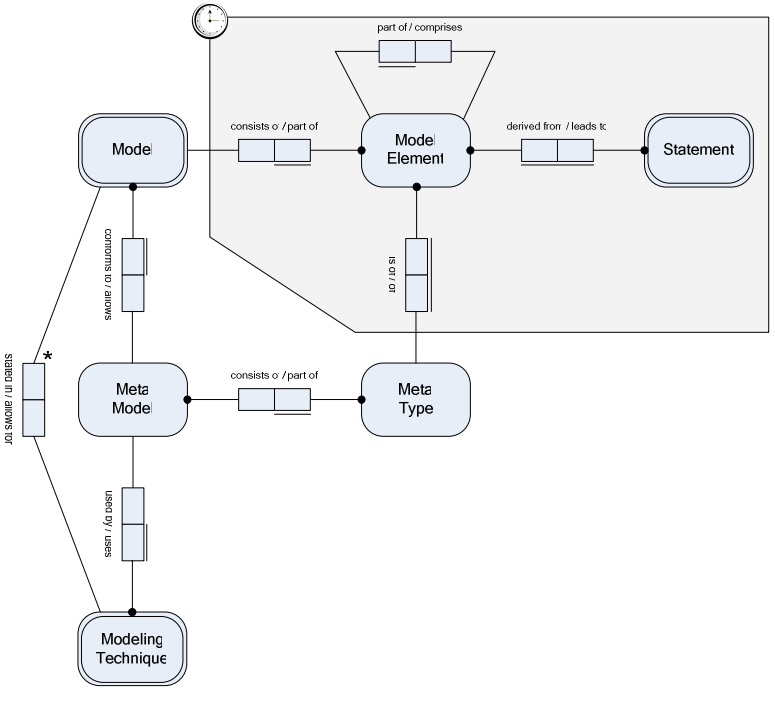

Fig. 3. Models, meta-models and modeling techniques

3. Currently, we focus on modeling dialogues where the informers are presumed to have a homogenous view on the domain to be modeled. This is, however, rather unrealistic from a practical point of view. Once we have conducted modeling experiments in a controlled homogenous setting, we aim to also cater for the negotiation processes in creating a unified model when dealing with heterogeneous views of different informers on the same domain.

References

1. P.P. Chen. The Entity-Relationship Model: Towards a Unified View of Data. *ACM Transactions on Database Systems*, 1(1):9-36, March 1976.
2. Kroenke D.M. *Databases, principles, design and implementation (Databases, beginselen, ontwerp en implementatie)*. Pearson Education Benelux, Amsterdam, The Netherlands, EU, 9th edition, 2004. In Dutch.
3. Norbert E. Fuchs, Uta Schwertel, and Rolf Schwitter. Attempto Controlled English - Not Just Another Logic Specification Language. *Lecture Notes in Computer Science*, 1559: 1-20, 1999.
4. T.A. Halpin. *Information Modeling and Relational Databases, From Conceptual Analysis to Logical Design*. Morgan Kaufmann, San Mateo, California, USA, 2001.
5. B. van Hoof, H. van Seters, and J.P. Zwart. *Lecture Notes on Proces and Data Modeling (Syllabus Proces- en Datamodellering)*. HAN University of Applied Science, Arnhem, The Netherlands, EU, 2002. In Dutch.
6. S.J.B.A. Hoppenbrouwers and H.A. (Erik) Proper. A Fundamental View on the Process of Conceptual Modeling. In *Conceptual Modeling - ER 2005 - 24 International Conference on Conceptual Modeling*, volume 3716 of *Lecture Notes in Computer Science*, pages 128-143, June 2005.
7. S.J.B.A. Hoppenbrouwers and H.A. (Erik) Proper. Formal Modelling as a Grounded Conversation. pages 139-155, June 2005.
8. S.J.B.A. Hoppenbrouwers, H.A. (Erik) Proper, and V.E. van Reijswoud. Navigating the Methodology Jungle - The communicative role of modelling techniques in information system development. *Computing Letters*, 1(3), 2005.
9. S.J.B.A. Hoppenbrouwers, H.A. (Erik) Proper, and Th.P. van der Weide. Towards explicit strategies for modeling. In T.A. Halpin, K. Siau, and J. Krogstie, editors, Proceedings of the Workshop on Evaluating Modeling Methods for Systems Analysis and Design (EMMSAD'05), held in conjunctiun with the 17th Conference on Advanced Information Systems 2005 (CAiSE 2005), pages 485-492, Porto, Portugal, EU, 2005. FEUP, Porto, Portugal, EU.
10. S.J.B.A. Hoppenbrouwers, H.A. (Erik) Proper, and Th.P. van der Weide. Understanding the Requirements on Modelling Techniques. In O. Pastor and J. Falcao e Cunha, editors, *17th International Conference on Advanced Information Systems Engineering, CAiSE 2005, Porto, Portugal, EU*, volume 3520 of *Lecture Notes in Computer Science*, pages 262-276, Berlin, Germany, EU, June 2005. Springer-Verlag.
11. L. Lindemans. Modeling Processes (Modelleerprocessen). Master's thesis, Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands, EU, June 2006. In Dutch.

12. J. Mylopoulos. Techniques and Languages for the Description of Information Systems. In P. Bernus, K. Mertins, and G. Schmidt, editors, Handbook on Architectures of Information Systems, Berlin, Germany, EU. Springer, Berlin, Germany, EU, international handbooks on information systems edition, 1998.
13. OMG. UML 2.0 Superstructure Specification - Final Adopted Specification. Technical Report ptc/03-08-02, August 2003.
14. Patrick van Bommel, P.J.M. Frederiks, and Th.P. van der Weide. Object-Oriented Modeling based on Logbooks. *The Computer Journal*, 39(9):793-799, 1996.
15. J. Warmer and A. Kleppe. *Practical UML (Praktisch UML)*. Pearson Education Benelux, Amsterdam, The Netherlands, EU, 3 edition, 2004. In Dutch.