Structuring Multidisciplinary Knowledge for Model Based Water Management: The HarmoniQuA Approach

Huub Scholten[a], Jens Christian Refsgaard[b] and Ayalew Kassahun[a]

[a]*Information Technology Group, Department of Social Sciences, Wageningen University, The Netherlands (huub.scholten@wur.nl)*
[b]*Geological Survey of Denmark and Greenland, Denmark*

Abstract: In the last decade a strong need for Quality Assurance has emerged among professionals in model based water management. The manifold of reasons for this can be summarised as follows. There is a lack of mutual understanding between modellers, model study clients, model auditors, stakeholders and concerned members of the public. Furthermore, malpractice (careless handling of input data, insufficient calibration/validation and model use outside of its scope) promotes a growing disbelief in the competence of models to support decision-making. This growing disbelief is fuelled by the tendency of modellers to oversell model capabilities. Several initiatives claim to support Quality Assurance in model based water management, but these focus on single domains and usually have chosen a textbook approach to provide guidelines and fill-in forms to monitor modelling. The EC funded project HarmoniQuA aims at supporting the full modelling process by offering computer based guidance for all water management domains, different types of users, different types of modelling purposes (planning, design and operational management) and different levels of modelling complexity. In addition to the guidance, users are helped to record all what they do and to produce for a diverse audience dedicated reports. Finally, HarmoniQuA intends to facilitate co-operation between groups within and between modelling studies and let its users learn from other modelling studies. The HarmoniQuA system consists of a knowledge base and a toolbox. This paper will discuss scientific, technical and organisational considerations behind the HarmoniQuA approach for integrated modelling support. These considerations can be divided into two groups, one related to the expertise of modellers, the other to knowledge engineering. An ontological approach, originating from knowledge engineering technology, is perfectly equipped to design a structure for the modelling knowledge base, which is filled with the expertise of modellers and to design the structure for recording the modelling process in so called model journals. The HarmoniQuA toolbox uses the knowledge base to generate dedicated guidelines for the variety of users and fills (instances of) the model journal ontology with actual data of a modelling process.

***Keywords**:* Modelling knowledge; Model based water management; Ontologies; Multidisciplinary knowledge.

1. INTRODUCTION

Mathematical models have been applied for several decades in solving problems in many domains of water management. With the requirements imposed by the EU Water Framework Directive the trend to base water management decisions to a larger extent on model studies and to use more sophisticated models is likely to be reinforced. At the same time insufficient attention is generally given to documenting the predictive capability of the models. In the last decade a growing need for Quality Assurance (QA) has emerged among professionals in this field [Refsgaard, 2002]. Quality Assurance is defined by NRC [1990] as the procedural and operational framework used by an organisation managing the modelling study to assure technically and scientifically adequate execution of all tasks included in the study and to assure that all modelling-based analysis is reproduced and defensible.

Refsgaard [2002] gives a summary of reasons for the growing interest in quality assurance:

- Ambiguous terminology and a lack of mutual understanding between key-players;
- Malpractice (careless handling of input data, inadequate model set-up, insufficient calibration/validation and model use outside of its scope);
- Lack of data or poor quality of available data.
- Insufficient knowledge on the processes hindering ecological (biota) modelling.
- Miscommunication of the modeller to the end-user on the possibilities and limitations of the

modelling project and overselling of model capabilities;

- Confusion on how to use model results in decision making;
- Lack of documentation and transparency of the modelling process, leading to projects, which hardly can be audited or reconstructed.
- Insufficient consideration of socio-economic, institutional and political issues and a lack of integrated modelling.

As a result of these problems, the credibility of the models is often questioned, and sometimes with good reason [Refsgaard, 2002]. The need for improving the quality of the modelling process has regularly been emphasised by the research community, e.g. Klemes [1986], NRC [1990], Anderson and Woessner [1992], Forkel [1996] and Rykiel [1996]. The recommendations made in this respect mostly focus on scientific/technical guidance in how the modeller should carry out the various steps of the modelling work in order to achieve the best and most reliable results.

Existing modelling guidelines, mostly nationally based, focus on a single domain in contrast to multi-domain and integrated models [Refsgaard, 2002]. Furthermore, these guidelines vary throughout Europe. The resulting models and decisions based on them are therefore often: not transparent, irreproducible, non-auditable and not fully comparable between different countries.

The Water Framework Directive (WFD) provides European policy at the river basin scale. It explicitly states that water resource models should be applied. The EU-financed project HarmoniQuA aims at improving the quality of model based water management at catchment and river basin scales by providing guidance throughout the modelling process and by supporting all persons involved (water managers, modellers, auditors, stakeholders and concerned members of the public) in their activities. The guidelines are based on accepted and common methodology and practices of experienced modellers. This knowledge is collected, completed, improved and made available in the form of a Knowledge Base, using state-of-the-art knowledge engineering technology, which uses an ontological approach. MoST, the software tool of HarmoniQuA provides guidance from the Knowledge Base, it supports monitoring of the modelling activities and reporting to various audiences. In the future MoST will use expertise collected in previous modelling studies to advise on how to perform the model study at hand.

This paper focuses on how HarmoniQuA handles and improves existing knowledge on modelling for water management.

2. KNOWLEDGE BASE (KB)

2.1. PREVIOUS EXPERIENCES

Quality management is quite common practice in software engineering, but in the field of modelling and simulation quality management is often restricted to verification and validation issues. Scholten and Udink ten Cate [1999] have proposed a Simulation Maturity Model (SMM), comparable to Humprey's Capability Maturity Model (CMM), which was developed to improve software engineering [Humphrey, 1989]. Just like CMM, SMM distinguishes five stages of maturity: *ad hoc*, repeatable, defined, managed and optimised. For QA in the context of model based water management the definition stage is the most essential stage. Such a definition can be used as a base for modelling guidelines as has been done to produce a (Dutch) Good Modelling Practice Handbook [Scholten *et al.*, 2001, Van Waveren *et al.*, 2000].

The HarmoniQuA approach in developing a knowledge base with guidelines made use of the experiences in realising the Dutch GMP Handbook. Furthermore it was based on other water management related guidelines, especially on the Murray-Darling groundwater flow modelling guideline in Australia [Middlemis, 2000] and the Bay-Delta modelling protocol for water and environmental modelling in Californian [BDMF, 2000].

2.2. DESIGN CRITERIA

HarmoniQuA will support model based water management in general, we had to decide to realise support at a very generic level only or to aim at serving all key players, covering a wide range of water management domains, suited for various purposes and for modelling jobs of different complexity. In this way the knowledge has been dedicated to a specific modelling study and to the roles of its key players. The term 'domain' refers here to the disciplines of water management, making water management, as required by WFD, a multidisciplinary topic. The following domains / disciplines and other dedication aspects are distinguished:

- *Domains*: groundwater, precipitation-runoff, hydrodynamics, flood forecasting, surface water quality, biota (ecology) and socio-economics;
- *User types*: modeller (e.g. consultant), water manager (e.g. client), auditor, stakeholders and (concerned members of the) public;
- *Application purpose*: planning, design and operational management;

- *Job complexity*: basic, intermediate and comprehensive.

A major design criterion in the development of the modelling knowledge base is the granularity of the decomposition. Three decomposition levels are distinguished. At the highest level the modelling process has been divided in *steps*, which are groups of *tasks*. To perform a *task* one or more *activities* have to be performed. An *activity* is related to the actor, being the smallest 'doing' in the process. A *task* related to what has to be done and it refers therefore to the modelling process. By performing *activities* a *task* will be completed. *Steps* are logical groups of *tasks* and have only an organisational purpose for human actors involved in the process.

Besides making an appropriate choice in the granularity of the decomposition, several other design criteria are relevant for the KB:

- Explicit structure of the modelling process;
- Easy to update by web based access;
- Easy to maintain;
- Flexible structure;
- Authorisation management for knowledge editors;
- Portable to other operating platforms than MS Windows™;
- Other, normal software engineering criteria.

To deal with most of these criteria we chose for an ontological approach in the design of the KB.

2.3. AN ONTOLOGICAL APPROACH

The most used definition in knowledge engineering of the term 'ontology' is of Gruber [1993, 1995]: an ontology is an explicit specification of a conceptualisation, referring to what can be represented in terms of concepts and the relationships among them. Borst [1997] added to this definition that there should be consensus of the concepts and the relations between them, resulting in the following definition: an ontology is a formal specification of a shared conceptualisation.

Uschold *et al.* [1998] distinguishes three groups of uses of ontologies: *communication* (of structured knowledge between people and between organisations), *interoperability* (understanding knowledge between machines and between men and machines) and *systems engineering* (for software systems and knowledge based systems, facilitating re-use and making knowledge explicit).

Developing an ontology is a part of a process to build a knowledge base for some purpose. This process is typically composed of the following steps:

- An *ontological structure* is made which is the frame of the intended knowledge base;
- A tool based on this ontology is used for *knowledge acquisition*;
- The acquired knowledge is stored *as instance of the ontology* in a knowledge base;
- A *software application* is developed which uses this knowledge base.

An ontology can be seen as a framework that represents the syntax and the semantics of data structures for a certain domain, in a formal, machine processable way. In order to describe the syntax and semantics of data structures, ontologies must provide one or more standard vocabularies, defining the terms (*concepts*) and *relations* that are used to describe this specific knowledge domain (subject area). To describe a certain piece of knowledge an ontology contains sentences describing the *concepts* and *relations* between them. *Concepts* can be discussed and have to be represented. The term *concept* has thus a broader meaning than 'entity' and it encompasses abstract and concrete things, but also processes, tasks and ambitions or goals. *Concepts* are used to define and explain terms. *Relations* organise *concepts* in a hierarchical or in some self-defined structure. Often ontologies contain other elements e.g. *properties*, *functions*, *axioms*, but these are not essential to understand what ontologies are. *Instances* are also parts of an ontology, as they contain the actual knowledge. If *task* is a concept in an ontology, the instances of *task* can be *go shopping*, *cook a meal*, *eat the meal*. A comprehensive and clear introduction on what ontologies are and why we need them is given in Chandrasekaran *et al.* [1999].

In HarmoniQuA we use Protégé2000 as tool to build the ontology and to include the collected knowledge as instances of the ontology. The functionality of Protégé2000 has been extended by building a plug-in for XML-export, according to a predefined interface for the tools that have to co-operate with the KB.

2.4. FROM KNOWLEDGE TO KB

In a small group of 5 modelling experts with know-how in knowledge engineering techniques and/or experience in the development of one of the existing guidelines, of the modelling process has been decomposed. In this first draft the modelling process has been divided in 5 *steps* and at a lower decomposition level into around 50 *tasks*. Each task was further decomposed in the following *task describing components*: name, definition, explanation, one or more activities (most with one or more associated methods), references, software aspects, links to other tasks and some other aspects. This first draft consisted of *structure diagrams*, which determined the order of the task and *spreadsheets* with rows for the task determining components, a column to fill these in and columns to indicate relevance of each task determining component for various types of users, domains, job complexity and application purpose. In the decomposition three types of tasks are distinguished: *normal tasks*, *decision tasks* (to decide on advancing to the next task or going back to a previous one) and *review tasks* (i.e. special decision tasks emphasising the negotiating interaction between water manager and modeller).

2.5. IMPLEMENTATION

In the second stage, a modelling expert for each of the 7 domain filled in spreadsheets for each of the 50 tasks and indicated the relevance for the user types, application purpose and job complexity. Two parties, not involved in providing the knowledge, assessed the quality of the knowledge. All HarmoniQuA partners discussed their findings and proposed improvements. In a third step most of these comments were incorporated in the spreadsheets, resulting in a first, spreadsheet based, prototype, which was not very suited to be used in actual modelling studies. The 350 spreadsheets were imported in Protégé2000 as instances of the ontology structure for the KB (see Figure 1) with a tool developed in the project.

Spreadsheets were also the preliminary front-end interface in the development of a glossary for the terminology in model based water management. The domain experts delivered a prototype glossary of almost 1000 entries, which has been included in the Protégé2000 KB.

A series of tools has been built in HarmoniQuA to work with the KB. As extra front-end for experts in model based water management, a web based knowledge editor and a glossary editor assist in completing and improving the modelling knowledge and the glossary. The major tool, especially for end-users, provides guidance, supports monitoring of the modelling activities and helps in reporting to various audiences. In the future it will give advice based on previous modelling studies. This tool will be described briefly in section 2.6 and more comprehensively by Kassahun et al. [2004].

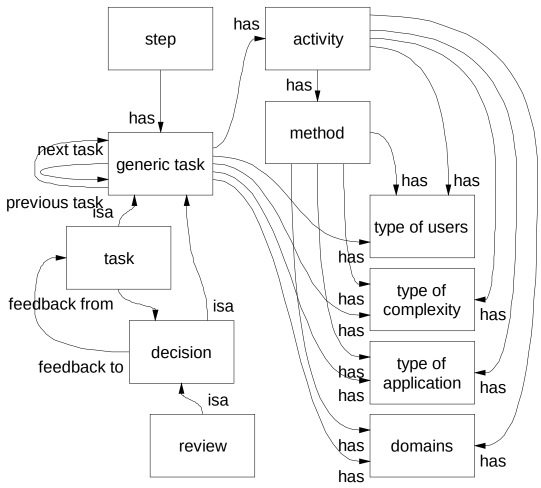

Figure 1. Part of the ontological structure of the KB. Rectangles are concepts and arrows relations.

2.6. CONTENT

The knowledge base is too large to be discussed here in detail. It consists of the following *steps*: (1) purpose and conditions, (2) conceptualisation, (3) model set-up, (4) calibration-validation and (5) prediction. The first step (purpose and conditions) focuses on the interaction between the water manager and the modeller (or the modelling team). This step starts with a series of tasks for water manager in the following order: *describe problem and context*, *identify data availability*, *define objective*, *determine requirements*, *prepare terms of reference*. In the next task (*prepare / evaluate tender*) there are roles for the water manager, the modeller and the auditor. The last task in this step is the decision task *agree on model study plan*, where water manager and modeller have to discuss how the model study has to be continued. If the parties come to an agreement, one has to continue with the conceptualisation step and do the first task *describe system and data availability*, typically a modeller's task.

The tasks briefly presented here do not consist of activities, which have complex methods to be used, as is the case for the complex task *parameter optimisation* being a task in the step *calibration and validation*. This task consists of several activities to do and a number of methods to use. The HarmoniQuA guidelines recommends to 'use expert knowledge' or – in case the modeller is not familiar with the model code – to 'do a sensitivity analysis' to select optimisation parameters. For a sensitivity analysis several methods are recommended, including, but not restricted to analytical sensitivity analysis, manual sensitivity analysis, Response Surface Methods, Monte Carlo Methods. Furthermore, the guidelines give short introductions of the suggested methods.

3. MOST

The end-user part of the HarmoniQuA system, called MoST (Modelling Support Tool), provides guidance, monitors modelling activities and helps in reporting. An outline of the functionality of MoST is shown in **Figure 2**.

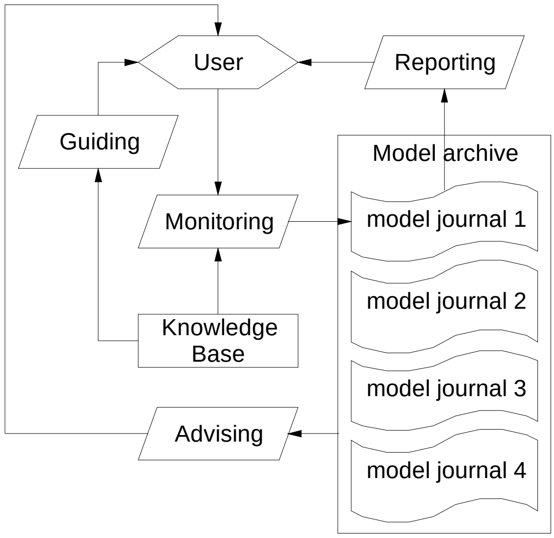

Figure 2. Outline of the functionality of the HarmoniQuA system.

The **Guideline Tool** presents guidance from the knowledge in the KB dedicated for a specific modelling study and dedicated for the various users involved. Users can browse through the *steps*, *tasks* and read what *activities* have to be done and which *methods* are available to carry out the activities. Those that are interested can download a prototype of the Guideline Tool after registration and the guidelines from the website www.harmoniqua.org/deliverables.htm. This tool is also integrated in MoST. In a model study different actors are involved in a mixture of *tasks* and *activities*. These actors can learn what they have to do according to the guidelines provided by the guideline part of MoST. Further MoST supports the actors in the modelling study by helping **monitoring** what they do. What actually has been done and what decisions have been made is monitored and stored in *model journals*. The latter can be stored on a local computer or at some central network server. The structure of the *model journals* is also based on an ontology, which is comparable to the one of the KB. A prototype of this monitoring functionality of MoST has been completed and tested in 10 test case. The first results show the usefulness of this part of MoST, although many shortcomings have to be repaired and wishes for extra functionality have to be realised in a next version, which will be released at the end of 2004. So far MoST has a primitive **reporting functionality**, but the next version will enable to write reports for different audiences.

The **advisor functionality** is still in the design phase and implementation may appear not within the scope of the HarmoniQuA project.

The HarmoniQuA suite for model based water management will be completed with training material in the form of multi-medial course-ware for students and workshop-ware for professionals.

4. CONCLUSION

The HarmoniQuA Modelling Support Tool (MoST) aims at supporting model based water management. This support consists of providing guidance to all involved in the modelling process. This guidance is provided by a knowledge base (KB), which contains expert knowledge on many facets of modelling, including specific knowledge on seven domains of water management and of a glossary with many entries from the jargon used in water management. Many efforts are made to ensure that the knowledge in the KB is accepted by a wide group of key players. Furthermore, the KB has been designed and implemented using an ontological approach, which is a state-of-the-art knowledge engineering technology. This approach has been chosen to guarantee that design criteria are met. These include a proper choice of the level of detailing, easy maintenance and updating of the structure and the content of the KB. Furthermore, the knowledge has been made as explicit as possible and specific for seven *domains*, five *user types*, three *application purposes* and three level of *job complexity*. This approach allows to provide guidance specific for a model study and the persons involved. In this way HarmoniQuA intends to contribute a methodological part to an infrastructure for implementing WFD and for model based water management in general.

It is too early to draw final conclusions, as two new versions of the HarmoniQuA product are planned for the next two years. We are in the middle of testing, evaluating and improving what we achieved so far. Is the KB complete, consistent and has an appropriate level of detail been chosen? Can all experts agree on the methodology provided? Does the toolbox MoST meet the needs of all intended users? Even if the last question may be answered positively, we have to find out, if the wider professional community will use it.

The achievements of our approach so far suggest potentials for a wider use, such as for modelling in multidisciplinary problem solving projects in general.

5. ACKNOWLEDGEMENTS

The present work was carried out within the Project 'Harmonising Quality Assurance in model based catchments and river basin management (HarmoniQuA)', which is partly funded by the EC EESD-programme, Energy, Environment and Sustainable Development (Contract EVK1-CT2001-00097).

6. REFERENCES

Anderson, M.P. and W.W. Woessner, The role of postaudit in model validation. *Advances in Water Resources, 15*, 167-173, 1992

BDMF, Protocols for Water and Environmental Modeling. Bay-Delta Modeling Forum. Ad hoc Modeling Protocols Committee. www.sfei.org/modelingforum/, 2000

Borst, W. N., Construction of engineering ontologies for knowledge sharing and reuse. University of Twente, Enschede, The Netherlands, , 1997

Chandrasekaran, B., Josephson, J., Benjamins, V., What are ontologies and why do we need them? *IEEE Intelligent Systems 14*, 20-26, 1999

Forkel, C., Das numerische Modell - ein schmaler Grat zwischen vertrauenswürdigem Werkzeug und gefährlichem Spielzeug. Presented at the 26. IWASA, RWTH Aachen, 4-5 January 1996, 1996

Gruber, T. R., A translation approach to portable ontology specifications. *Knowledge Acquisition 5(2)*, 199-220, 1993

Gruber, T. R., Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies 43*, 907-928, 1995

Humphrey W.S., *Managing the software process*, Addison-Wesley, Reading, 494 pp., 1989.

Kassahun, A., H. Scholten, G. Zompanakis and C. Gavardinas, Support for model based water management with the HarmoniQuA toolbox, Proceedings iEMSs2004.

Klemes, V., Operational testing of hydrological simulation models. *Hydrological Sciences Journal, 31*, 13-24, 1986.

Middlemis, H.. Murray-Darling Basin Commission. Groundwater flow modelling guideline. Aquaterra Consulting Pty Ltd., South Perth. Western Australia. Project no. 125, 2000.

NRC, Ground Water Models: Scientific and Regulatory Applications. National Research Council, National Academy Press, Washington, D.C., 1990.

Refsgaard J.C. and H.J. Henriksen, Modelling guidelines – terminology and guiding principles. *Advances in Water Resources 27(1)*, 71-82, 2004.

Refsgaard, J.C. (ed.), State-of-the-art report on quality assurance in modelling related to river basin management. HarmoniQuA Report, D-WP1-1, Copenhagen, Denmark, 2002. www.HarmoniQuA.org/deliverables.htm,

Rykiel, E.R., Testing ecological models: the meaning of validation. *Ecological Modelling, 90*, 229-244, 1996.

Scholten H. and A.J. Udink ten Cate, Quality assessment of the simulation modeling process. *Computers and Electronics in Agriculture, 22*,199-208, 1999.

Scholten, H., R. H. Van Waveren, S. Groot, F. Van Geer, J. H. M. Wösten, R. D. Koeze, and J. J. Noort, Improving the quality of model based decision support: Good Modelling Practice in water management. Pages 223-230 in: A. Schumann, J. Xia, M. Marino, and D. Rosbjerg (Eds). Regional Management of Water Resources. Proceedings of Symposium S2 of the 6th Scientific Assembly of the IAHS, Maastricht, The Netherlands, 2001.

Uschold, M, King, M, Moralee, S. and Y. Zorgios, The Enterprise Ontology, *The Knowledge Engineering Review*, Vol. 13, Special Issue on Putting Ontologies to Use (eds. Mike Uschold and Austin Tate), 1998.

Van Waveren, R.H., S. Groot, H. Scholten, F. Van Geer, H. Wösten, R. Koeze, J. Noort,. Good Modelling Practice Handbook, STOWA, Utrecht, RWS-RIZA, Lelystad, The Netherlands (in Dutch, English version from http://www.landandwater.tudelft.nl/downloads/GMP-uk.pdf), 2000.