A knowledge base system for multidisciplinary model-based water management

Ayalew Kassahun[a] **and Huub Scholten**

[a]*Information Technology Group, Department of Social Sciences, Wageningen University, The Netherlands (ayalew.kassahun@wur.nl)*

Abstract: To improve the quality of multidisciplinary model-based water management studies and provide guidance on best practices, a user-friendly guidance and quality assurance knowledge-based system has been built within the EU funded HarmoniQuA project. The knowledge base system provides experts and experienced persons in model-based water management studies a web-based software tool for developing knowledge base for processes related to managing model-based water management studies. The knowledge base is used in the Modelling Support Tool, a tool developed within the HarmoniQuA project, to provide multidisciplinary team members targeted and actionable information. MoST also uses the knowledge base as a process specification in managing model-based study projects. In the HarmoniQuA project, quality assurance procedures has been developed with the help of this system and currently we are extending our KBS to develop a water stress mitigation process for the EU funded AquaStress project. The knowledge base system is based on ontological formalisation that consists of knowledge models for specifying process definitions and knowledge models for the execution of the process. This paper focuses on the knowledge models and a web-based software system we developed to edit and manage the knowledge base. The knowledge models are developed by a team of experts and experienced persons in model-based water management studies and a team of knowledge and software engineers. In a separate paper, we look at how these knowledge models are used in managing multidisciplinary model-based water management projects.

Keywords: process knowledge base, knowledge acquisition, quality assurance procedure

1. INTRODUCTION

The HarmoniQuA (Harmonising Quality Assurance in model based catchments and river basin management) project is an EU funded project with the objective of providing a user-friendly quality assurance (QA) framework. This quality assurance framework includes: (1) a scientifically based methodology for generic and domain specific QA procedures and (2) software tools for organising knowledge in a structured database to support modellers and water managers throughout the QA process [Scholten and Osinga, 2002, HarmoniQuA, 2006].

Problems related to methodology in model-based water resources management that the HarmoniQuA project set out to address include, among others, lack of mutual understanding due to ambiguous terminology, insufficient consideration of socio-economic, institutional and political issues, and lack of documentation and transparency [Refsgaard and Henriksen, 2004, Refsgaard *et al*., 2005, Scholten *et al*., 2006]. To solve these problems a QA procedure and software tools were developed

In the HarmoniQuA project, first a number of state-of-the-art modelling guidelines and procedures were evaluated. These include Good Modelling Practise [Van Waveren, *et al*., 1999], Bay-Delta water and environmental modelling protocols [BDMF, 2000] and Murray-Darling basin's commission groundwater flow modelling guideline [Middlemis, 2000]. Then a new QA procedure was developed based on the existing ones and new experiences and insights gained in recent years.

Most of those QA procedures evaluated and the new HarmoniQuA procedures use a process-oriented approach in specifying QA procedures. The process approach is based on the belief that the desired goal is achieved more efficiently when activities are managed as process [ISO 9000, 2004]. In fact, organizations that wished to be certified for

attaining the ISO quality standard must identify and manage the processes that make up their quality management systems [ISO 9001:2000, 2001].

We developed a knowledge-based system that supports the development of QA procedures. We have since then extended the system to support any processes in model-based study projects. In the context of the HarmoniQuA project, the knowledge base system allows experts to organise expert knowledge on the QA procedures required to guarantee the quality of model-based studies. The HarmoniQuA Modelling Support Tool (MoST), on the other hand, serves other types of users, namely users of the expert knowledge. The tool serves these group of users by guiding them in there work and managing their activities by allowing them to record their activities and generating customizable reports [Kassahun *et al*., 2004].

The knowledge base system is a general-purpose process knowledge base tool that has successfully been applied in developing quality assurance procedures. Currently we are using the system to develop the water stress mitigation process for the EU funded AquaStress project. In this paper, we explain the requirements, design and implementation of the knowledge base system. A separate paper explains in detail how MoST supports multidisciplinary model-based water management projects.

2. PROCESS KNOWLEDGE SYSTEM AND ITS ONTOLOGICAL REPRESENTATION

There is no commonly accepted definition of knowledge-based systems [Hendriks and Vriens, 1999, Gallupe, 2001.] For the purpose of this paper, we define *knowledge base* as a machine readable and –interpretable collection of information and a *knowledge-based system* as a software system that facilitates acquisition and maintenance of knowledge from experts and makes that knowledge accessible to novice users and other experts [Kassahun *et al*., 2004.]

Ontologies provide means of modelling and representing knowledge in formats that can be easily understood by humans and interpreted by computer programs. One of most widely used development platforms for ontology-based systems is Protégé. Originally developed to reduce the knowledge-acquisition bottleneck, Protégé has developed into a general-purpose environment for knowledge modelling [Gennari et. al, 2003].

We choose Protégé as our base framework for a number of reasons, among others: it's modular and extensible architecture, the availability of several ready made extensions called plug-ins and the fact that it is an 'open source' project backed by a large community of active users and developers.

Our process knowledge system is a web-based application that uses Protégé inside a server application. The server application provides a number of functionalities that were required by users but that are on available in Protégé.

2.1. Requirements to the system

At the beginning of the HarmoniQuA project, a group of experts in the field of water management, software and knowledge engineers defined the structure and the knowledge items to be collected. We started knowledge acquisition using spreadsheets and manually processing the spreadsheets to create a common understanding of the system that is going to be implemented [Scholten, *et. al*., 2006]. This step has been vital in making it clear to software and knowledge engineers what the needs and expectations of the domain experts are. In addition, domain experts were also asked to specify their needs and specify requirements on how the system should work. At the end of this stage, we identified a number of requirements; the most important were:

- The process knowledge system should be simple and user-friendly. Domain experts are not trained in the use of general-purpose knowledge-based systems and may not be familiar with concepts that are specific to general purpose knowledge-based systems. Thus, a new interface should be built that directly supports their work.
- In the HarmoniQuA project, as well as in other projects, domain experts are located in different geographical locations. Therefore, the system should be accessible remotely and allow users to work collaboratively.
- Access to the system should be managed. An authorisation and authentication mechanism is required. Moreover, experts should be able to edit knowledge items only in their field of expertise.
- The system must be adaptable as new information and insights become available.

Based on the above requirements, it was decided to build a web-based interface to the Protégé knowledge-based framework. The web-based interface should shield users from the intricacies of working with Protégé. The system should provide functionalities that are essential but missing in Protégé such as authentication and authorisation and collaborative development.

2.2. Process ontologies

Two sets of generic ontologies have been constructed as part of the process knowledge

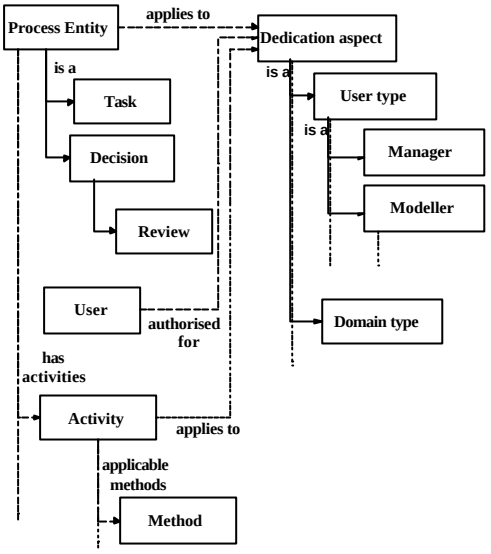

Figure 1. Part of the process knowledge ontology. Modified after Kassahun et al. (2004).

models (Figure 1). These are the *task* and the *dedication aspect* ontologies [Kassahun *et al.*, 2004]. Terms of the *task* ontology are concepts related to actions done, such as tasks, activities, methods and users. Terms of the *dedication aspect* ontology are qualifiers of instances from the *task* ontology. Dedication aspects are used in the authorisation module of the knowledge base authoring tool to grant or deny editing authorisation. In MoST, dedication aspects are used to filter instances of knowledge items based on user roles and preferences. They are also used in MoST's project management component to authorise users in performing tasks.

3. PROCESS KNOWLEDGE AUTHORING

Traditionally knowledge acquisition consists of a tedious task of elicitation (via interviews, etc.), analysis, modelling and validation of knowledge. One of the purposes of Protégé is to solve this knowledge-acquisition bottleneck. Protégé is a "meta-tool" that enables knowledge engineers to construct domain-specific knowledge-acquisition systems [Musen, *et al.*, 2000, Gennari et. al, 2003].

The process knowledge authoring tool uses Protégé internally to manage knowledge items. We develop software components around Protégé that will fulfil the requirements pursued. To allow collaborative knowledge acquisition and analysis the system should be accessible via Internet. The required simplicity means that the system should encapsulate the knowledge engineering aspects of Protégé and allow experts to concentrate on entering knowledge items. Unlike relational databases, ontological structures are described in the same way as the actual data (*instances*). Therefore, one can simply combine multiple ontological data sources together. This makes ontology-based system easily extensible and adaptive.

In the HarmoniQuA project, domain experts were actively involved in the modelling of the knowledge system at an early stage. Therefore, there was no need of a tedious work of knowledge elicitation and analysis since the experts enter knowledge item into the system that manages the authoring process based on the ontology of the specific process problem domain.

3.1. Design of the system

Figure 2 shows the overall framework of the process knowledge system that utilises web technology. The *command handler* module handles communication with the users of the system. Users access the system either using a web-browser or MoST. Standard web browsers are used to edit the knowledge base. MoST downloads the knowledge base in XML format and allows users to browse the knowledge base offline or manage projects based on the process knowledge base both offline and online.

Protégé is a single user desktop application with little support for collaboration. The authentication and authorisation module fosters collaboration in a number of ways:

1. All users of the system are assigned dedication aspects upon registration. In the HarmoniQuA knowledge base, users are assigned to a number of dedication aspects, such as *user types* (General Public, Modeller,

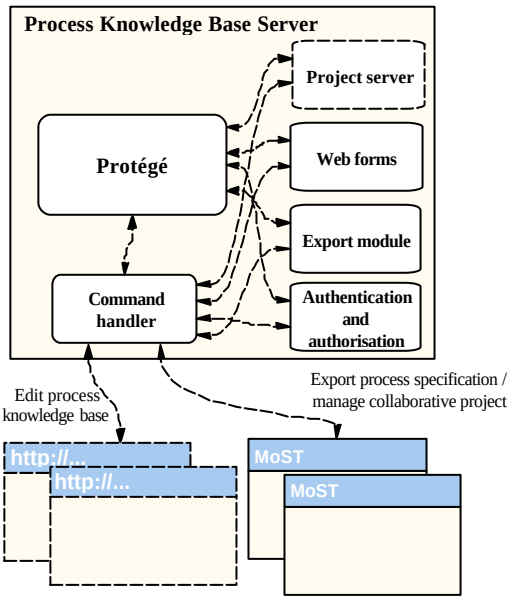

Figure 2. Framework for the implementation of the process knowledge system.

Manager, etc) and *domain types* (Generic, Groundwater, Hydrodynamics, etc). Knowledge items are also tagged with dedication aspects. Users can edit the knowledge base items when their dedication aspects match with or include the dedication aspects of the knowledge item.

2. Users can work in the system concurrently. The system has a locking mechanism to overcome authors entering conflicting updates.

3. Users can attach comments to knowledge items.

The *export* module and the *project server* are used to export and store the knowledge base in XML format. MoST uses the export module to provide the knowledge base offline.

3.2. Implementation

The process authoring tool is built using the Java programming language and the Java EE servlet technology. The Java servlet application on the server, therefore, needs to be installed in a servlet container. The server application uses the Protégé API to access and manage knowledge bases.

HTML and JavaScript is used to generate user interfaces. The authoring tool can be used by any standard web browser. In addition, extra functionalities are built such as backup-and-recovery, spelling checking and correction facilities to give domain experts extra incentive to adopt the system.

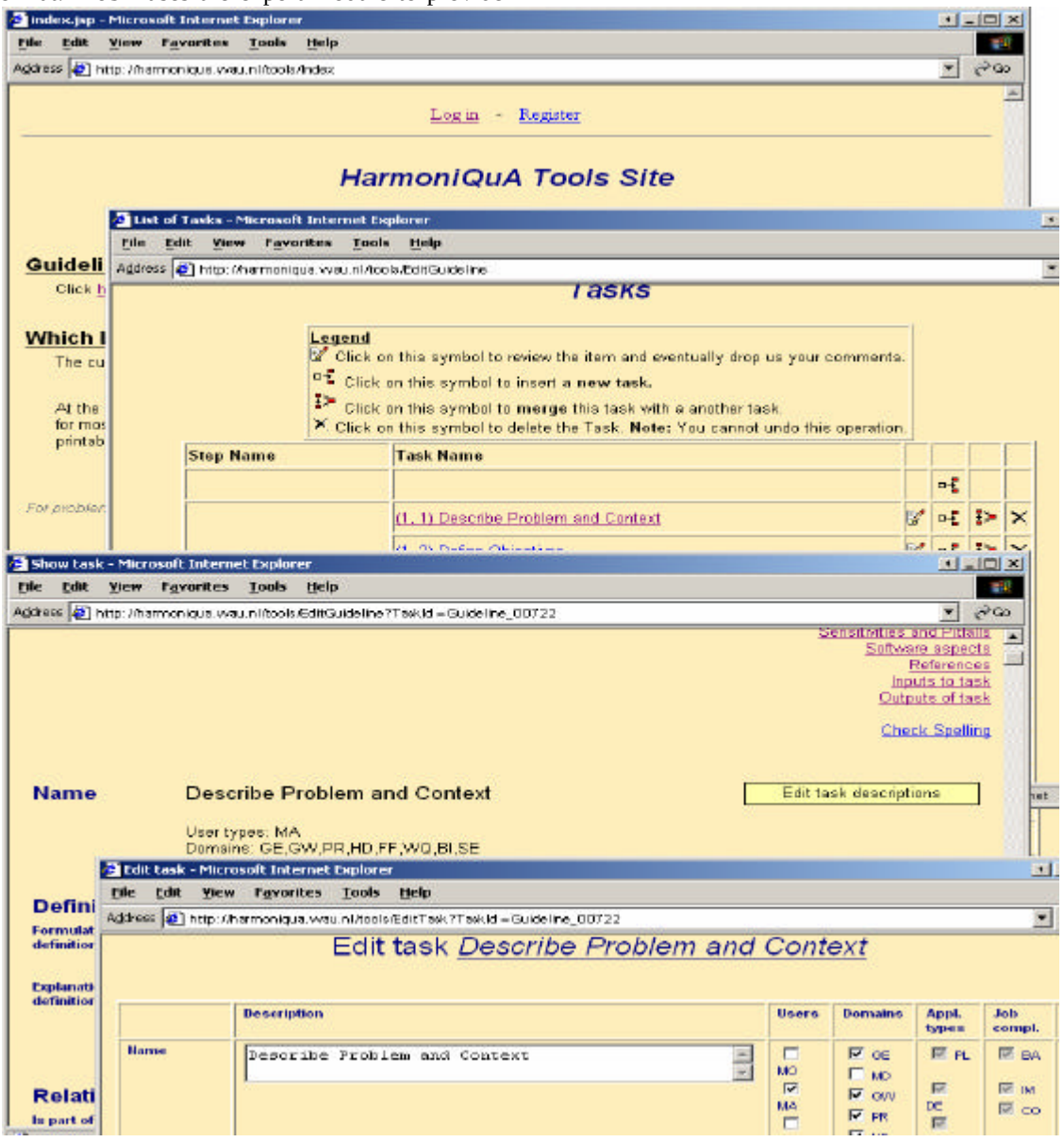

Figure 3. Few screen shots of the user interface of the process knowledge authoring tool.

3.3. User interaction

The process authoring tool generates user interfaces in HTML and JavaScript so that the only requirement on the users' computer is a standard web-browser. Few screen shots of the user interface are shown in Figure 3.

Users must be registered to enter the knowledge base and will have strict permissions on the operations they are allowed to perform. All registered users have *read* and *comment* authorisation rights.

To edit the knowledge base users must be assigned to a dedication aspect. Dedication aspects are assigned to users as well as to knowledge items. When assigned to users, dedication aspects function as *roles;* when assigned to knowledge items they mean: "this item is *applicable to ...".* When a user, e.g. a domain expert in groundwater, wants to edit an instance of the task "Describe Problem and Context", the user needs to be assigned to all dedication aspects that this *task* has. In this case, if the task is a modeller's task for the domain of groundwater, the user needs to be a modeller in the domain of groundwater.

The user interface provides a web interface for administration of users and making backups.

4. PROCESS KNOWELDGE BASE AND MoST

The Modelling Support Tool, MoST, has the functionality to guide users, monitor what they do in their modelling projects, and let them generate reports on the projects. MoST relies on the process knowledge that is collected and organised using the process knowledge system. The system keeps track of the different versions of the process knowledge base and makes that available to MoST. MoST can be used offline by a single user and online by a users working as a team. When used online, MoST accesses projects managed by the knowledge base system and allow users to collaborate with team members in a project.

5. CONCLUSION

The process knowledge base system and associated web-based authoring tool were developed for the HarmoniQuA project. The process knowledge base is based on generic knowledge models to describe project processes. The authoring tool is based on web technology. It is allows experts to collect and organise process knowledge for use in water management in collaborative way. The authoring tool and the associated support tool have been successfully applied in HarmoniQuA. Currently we are using the system to the water stress mitigation process for the EU funded AquaStress project. The process knowledge authoring tool has greatly simplified the knowledge acquisition process that traditionally required substantial amount of work for collecting and analysing knowledge. The knowledge base authoring tool in combination of MoST supports all users involved in model-based water management studies. Experts use the tool to organise and manage knowledge-based process while other experts and novice users use the tools for guidance and managing their model-based studies.

6. ACKNOWLEDGEMENT

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 Energy, Environment and Sustainable Development programme (Contract EVK2-CT2001-00097).

7. REFERENCES

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

Gallupe, B. (2001). "Knowledge management systems: surveying the landscape." International Journal of Management Reviews **3**(1): 61-77.

Gennari, J. H., Musen, M. A., Fergerson, R. W., Grosso, W. E., Crubézy, M., Eriksson, H., Noy, N. F. and Tu, S. W. (2003). "The evolution of Protégé: an environment for knowledge-based systems development." International Journal of Human-Computer Studies **58**(1): 89-123.

HarmoniQuA (2006). http://harmoniqua.org, Accessed on Februay 11, 2006.

Hendriks, P. H. J. and D. J. Vriens. 1999. "Knowledge-Based Systems and Knowledge Management: Friends or Foes?" Information and Management 35:113-25.

ISO 9000 (2004). Introduction and support package, Guidance on the concept and use of the process approach for management systems, International Organisation for Standardization.

ISO 9001:2000 (2001). Quality Management Systems: Requirements., International Organisation for Standardization.

Kassahun, A., Scholten, H., Zompanakis, G. and Gavardinas, C. (2004). Support for model

based water management with the HarmoniQuA toolbox. Proceedings of iEMSs 2004: "Complexity and Integrated Resources Management". C. Pahl, S. Schmidt and T. Jakeman. Osnabrück, Germany, June 2004, International Environmental Modelling and Software Society**:** 1282-1287.

Middlemis, H. (2000). Murray-Darling Basin Commission. Groundwater flow modelling guideline. South Perth. Western Australia, Aquaterra Consulting Pty Ltd.

Musen, M. A., Fergerson, R. W., Grosso, W. E., Noy, N. F., Crubézy, M. and Gennari, J. H. (2000). Component-Based Support for Building Knowledge-Acquisition Systems. Proceedings of the Conference on Intelligent Information Processing, (IIP 2000) of the International Federation for Information Processing Sixteenth World Computer Congress (WCC2000), Beijing, China, pp.18-22

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

Refsgaard, J. C., Henriksen, H. J., Harrar, B., Scholten, H. and Kassahun, A. (2005). "Quality assurance in model based water management - review of existing practice and outline of new approaches." Environmental Modelling & Software **20**: 1201–1215.

Scholten, H., Kassahun, A., Refsgaard, J. C., Kargas, T., Gavardinas, C. and Beulens, A. J. M. (2006). "A methodology to support multidisciplinary model-based water management." Environmental Modelling & Software **(in press.)**

Scholten, H. and Osinga, S. A., Eds. (2002). Requirement analysis report of the HarmoniQuA knowledge base and tools. Wageningen, Wageningen University.

Van Waveren, R. H., Groot, S., Scholten, H., Van Geer, F., Wösten, H., Koeze, R. and Noort, J. (1999). Vloeiend modelleren in het waterbeheer. Handboek Good Modelling Practice. Utrecht/Lelystad, STOWA/RWS-RIZA.