EnviroInfo 2007 (Warschau)
Environmental Informatics and Systems Research
Copyright © Shaker Verlag, Aachen 2007. ISBN: 978-3-8322-6397-3

A Web-based System to Support e-Collaboration in the Design and Construction of Environmental Models

Halisson Brito[1], Julia Strauch[2], Jano Souza[3]

Abstract

Design and construction of environmental models are usually complex tasks that can be considerably improved with collaboration among design teams. Knowledge Management can also play an important role in these tasks, facilitating knowledge acquisition, dissemination and reuse. This paper presents a Web-based system to support e-collaboration in the design and construction of environmental models. The system is centered in Knowledge Management issues and also makes use of ontologies to increase the modeller's abilities to obtain, share and reuse knowledge about models in order to improve the design of environmental models. These tools are also valuable in model validation and reuse.

1. Introduction

Design of environmental models is usually a complex task. Sometimes it involves multidisciplinary teams to design, construct, test and validate a model of an event in the real world. An environmental model is a representation of what occurs during some event in nature.

Several tools to assist the design of environmental models are described in literature. However most of them are either pure geographical tools or CAD tools. The most semantic support they usually provide is the ability to register some metadata about the region being modeled.

Many times, the knowledge obtained in the design of an environmental model remains in the mind of the scientists and is neither easily shared nor reused in other modelling projects.

Knowledge Management systems may have a significant contribution to environmental modelling since it facilitates the acquisition, registration and reuse of the knowledge accumulated by a modeling team or even by an institution.

Collaboration may also play an important role in this scenario since it assists modellers in the solution of complex tasks, and permits sharing and reuse of the involved knowledge among all the participants/contributors of a modelling project. Furthermore, the development of information and communication technologies enabled the implementation of systems that are classified as Groupware. These systems aimed at providing better communication and collaboration processes among people engaged in the solution of some specific problem.

According to these principles, this paper presents ModManager, a Web-based system to support knowledge management on scientific models. It is part of the MODENA environment (Banito, et al., 2005b), which is being constructed to support the modelling activities of environment research teams. It is being validated through case studies in the GEOMA Network (Brazilian Thematic Network for Environmental Modelling of the Amazon) (GEOMA, 2006). The goal of this network is to model several aspects

[1] Systems Engineering and Computer Science Department, COPPE/UFRJ, Rio de Janeiro, Brazil.
Petrobras – Brazilian Petroleum Co., Rio de Janeiro, Brazil.
e-mail: hmbrito@cos.ufrj.br, internet: http://www.cos.ufrj.br, http://www.petrobras.com
[2] National School of Statistical Sciences, ENCE/IBGE, Rio de Janeiro, Brazil.
e-mail: juliast@ibge.gov.br, internet: http://www.ibge.gov.br
[3] Systems Engineering and Computer Science Department, COPPE/UFRJ, Rio de Janeiro, Brazil.
Institute of Mathematics, Federal University of Rio de Janeiro (IM/UFRJ), Rio de Janeiro, Brazil.
e-mail: jano@cos.ufrj.br, internet: http://www.cos.ufrj.br, http://www.im.ufrj.br

of the complex Amazon ecosystems. It consists of multidisciplinary researchers of several Brazilian institutions that are geographically distributed across the Brazilian territory.

It is important to notice that MODENA is not itself a modelling system. It is a computational environment to support e-Collaboration among research teams of the GEOMA Network. It aims at obtaining and sharing the knowledge of the researchers about the objects being modelled in a specific modeling project, as well as reusing that knowledge in other projects.

Besides design, these techniques help researchers in the construction (implementation), validation and use of environmental models.

In this paper we use the words 'scientist', 'researcher' and 'modeller' to represent the same agent in the system.

This paper is organized as follows: section 2 presents some related work in the area of design, and model management; section 3 presents our proposal and discusses some of its advantages; and section 4 contains final considerations and ongoing work.

2. Related Work

In general, the first approaches to model management were limited to model execution and/or data storage. An example is the prototype tool presented in Rinom and Kelner (1992). It was developed at NASA and called SIGMA Project. It aimed at supporting the construction and reuse of scientific models. That approach used the experience of the knowledge domain to assist model acquisition and execution through a high level graphic language for the specification of data flow. This approach ignored questions related to database access, efficiency and platform dependence.

Another similar approach is presented in Leonard (1993). The author described a prototype of a model management system for discrete events simulation using the relational approach to integrate data and models. This approach generated a variety of reports that were used to document the models.

In Cavalcanti et al. (2002) it was proposed an architecture to support scientific model metadata publication applied to the environmental area. It offered model sharing by using an extension of a heterogeneous and distributed database system called Le Select. The main feature of this proposal is the semantics of explicit representation of scientific models and their publication using XML. The main advantage of this architecture is the scientific experiment encapsulation for the user. A disadvantage is the model treatment in a static way: once published models can neither be extended nor composed to form new models.

Palma et al. (2000) described a collaborative spatial decision support system for environmental area. It provides several tools to support the decision-making process, ranging from problem definition, specification and documentation to database integration. However, these tools do not address specific issues of model management nor provide modelling life-cycle support.

Nowadays, in the semantic web context, a model management system can be a collaborative environment that combines techniques of knowledge management to support scientific modelling lifecycle in a way to increase modelling productivity and efficiency.

So, we propose an environment that allows the use of techniques of knowledge management to support scientific modelling lifecycle and that should have the following requirements:

- models should be available in heterogeneous and distributed platforms;
- models should be added to the database with minimum impact to the system;
- the model lifecycle and its evolution should be supported, including used resources, like data and programs, and derived knowledge of their use;
- the control flow of model execution should be distributed, so that several tools can cooperate and work interactively; and

Copyright © Shaker Verlag, Aachen 2007. ISBN: 978-3-8322-6397-3

- the use of knowledge management tools to support model identification, registration, utilization, reuse, learning, and sharing.

Another work Wäger and Hilty (2002) presented a simulation system for waste management that proposed the integration of modelling and simulation techniques into traditional planning and decision making procedures. It allowed the development of a complex model to simulate recovery and recycling facilities of plastics waste. However, it was intended to solve a specific problem. Furthermore, they addressed some weakness in their work that can be extended to many model management and simulation systems, like:

- the difficulty of identification and communication of underlying concepts and assumptions;
- the modelling languages unfamiliar to typical key players (that leads to difficulty in understanding the model and developing it);
- the systematic validation of the model is not supported.

In the next section we present how knowledge management techniques used in the proposed system help minimizing the effects of those problems.

3. ModManager: a system to support e-Collaboration in environmental modelling

In this paper we describe the use of MODENA (Scientific Models Management Environment) (Brito et al., 2005b) to support e-collaboration among environmental modellers in their tasks in model design.

MODENA (Scientific Models Management Environment) is a model management system that is focused on the knowledge issues of scientific model management. The system is also used to manage the 'execution' of models, that is, the simulations taken with model instances.

MODENA consists of two subsystems: ModManager (Bito et al., 2005a), that deals with issues of model management and knowledge management about the design, construction and validation of models; and ModRunner (Vassalo et a., 2006), that deals with issues of simulations (the use of models) and knowledge about them.

This paper focuses on the first subsystem, which is a human-centered collaborative system to support the design and construction of environmental models. One of its goals is to extract, as much as possible, the knowledge that researchers have on the object being modelled, throughout the whole modelling lifecycle. To accomplish that, the system has several tools in order to capture the knowledge on a model from the moment it is first imagined by a researcher to the time it is used in simulation tasks. Among the possibilities of these tools to enrich the knowledge base of a model, we can highlight:

- textual description of knowledge, where modellers can register relevant notes about the model or the whole development process;
- diagrammatic representation of knowledge, like mind maps and hyperbolic trees (that can be used to represent the relationship among the concepts concerning the model), beyond conventional diagrams (to represent internal processes or data flow inside the model or the relationship among models), all of them stored in the system database;
- chats/discussion forums about model design, construction, validation and use, that can enrich considerably the development process and the knowledge base of the model;
- registration of model metadata, data, equations, files, references, among many other features, that are also used in model simulations and to compose workflows;
- interchange of model data, metadata and knowledge among researchers to improve model development and use.

Copyright © Shaker Verlag, Aachen 2007. ISBN: 978-3-8322-6397-3

In order to understand how the system can help scientists in the modeling process, we describe below a usage scenario of the system in the context of the GEOMA network.

3.1 A general usage scenario of ModManager

A typical GEOMA research team is composed of researchers from several parts of the Brazilian territory. They may be involved, for instance, in the design of a dynamics population model to simulate the impact of the introduction of genetically modified mosquitoes into the environment in order to reduce the number of cases of malaria in the Amazon region (Wyse, Bevilacqua and Rafikov, 2004). This is usually a multi-disciplinary and geographically dispersed team that must work together to reach the expected result.

In order to do that, researchers can use the system described here to support their work in the design, construction and use of a model. First of all, the team leader or the coordinator of a modelling project creates a project in the system and associates to it the modelling team and the contributors. Even with no model included yet in the project, the modellers can start to include in it the knowledge about what we can call the "pre-model". This knowledge can be, for instance, modellers' discussions or design issues that may be important to the following phases of the project or even to other projects.

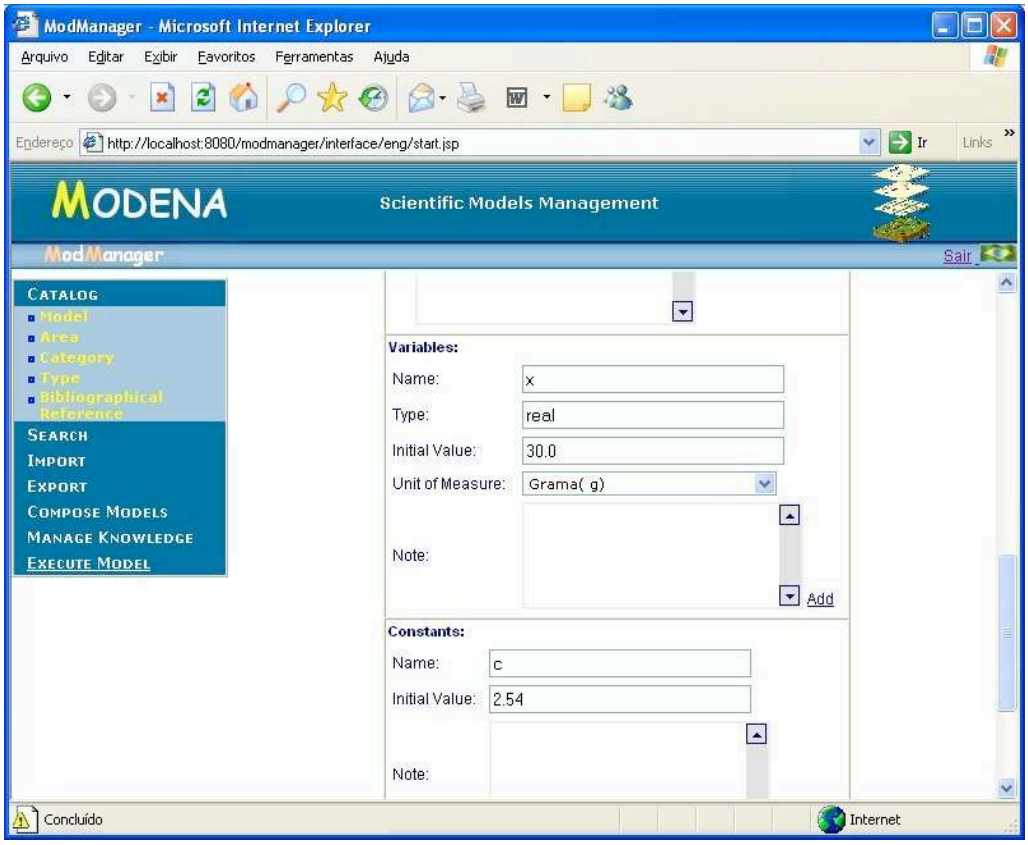

Figure 1: Model metadata registration

Copyright © Shaker Verlag, Aachen 2007. ISBN: 978-3-8322-6397-3

With the first sketch of a model, modellers can collaboratively register it in the system and then append metadata and knowledge as the model is being designed and constructed. As this system is a Web-based system, researchers belonging to different institutions distributed across the country can work together to improve the model development. Furthermore they may use interactive tools inside the system, like chats and forums, to improve their collaboration in the design and construction of the model, along the whole modelling process. Figure 1 shows the system screen for model cataloguing.

In order to validate the model, researchers can submit it to execution using ModRunner. They can also use chats and discussion forums to discuss the results and refine the model to reach the point where it becomes usable, adding more knowledge to the model and so enriching it.

3.2 Other tools to support e-collaboration

Another collaborative tool is the exchange of model information and knowledge in formats like CSV, XML and KO (Knowledge Objects) (Oliveira and Souza, 2004), to be used in other Knowledge Management systems. This tool allows researchers to import and export all model metadata, equations, constants, variables, algorithms, programs, workflow definitions, and all the knowledge about the model, beyond all the data used in or resulting from the execution of model instances.

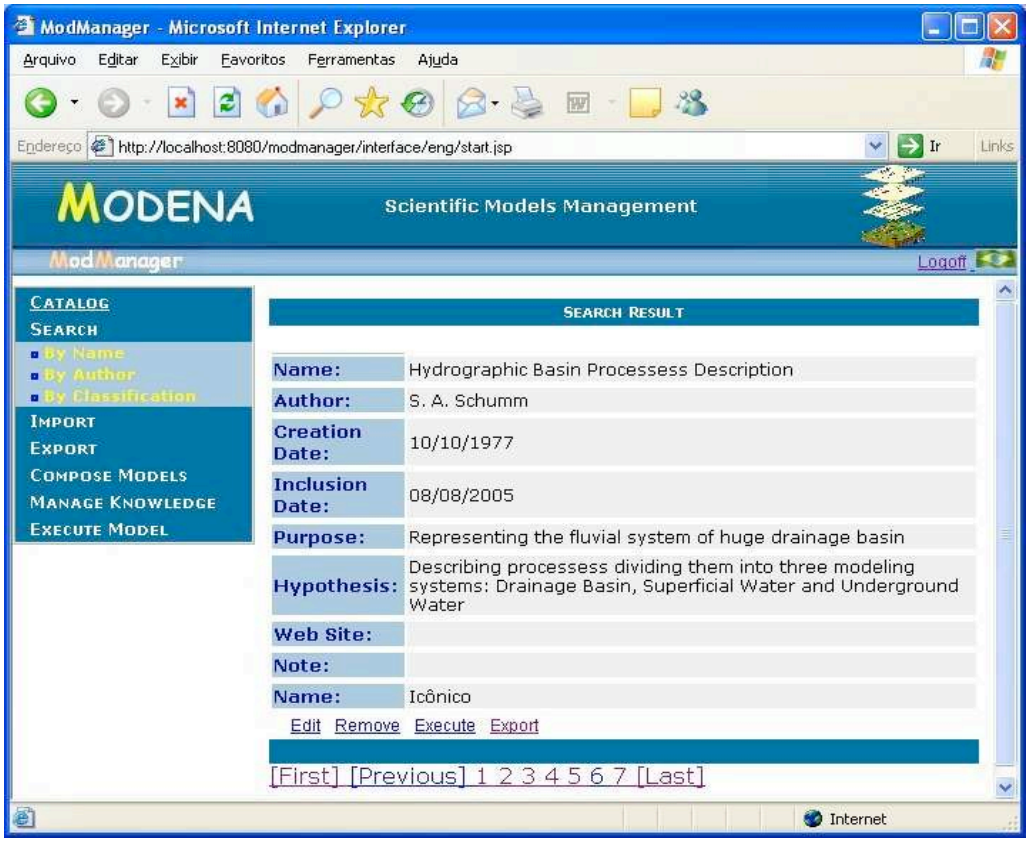

Figure 2: Result of model search

Copyright © Shaker Verlag, Aachen 2007. ISBN: 978-3-8322-6397-3

Model composition is another tool to be available in MODENA. It will permit the collaborative development of new models from existing ones. It is possible to chain the output of some models with the input of other ones in order to create new models for the solution of larger problems. So it enables models to be registered in the system in a modularized form (that is, sub models or smaller pieces of models), then integrating some pieces to form larger models, introducing modularization and reuse in model manipulation. Another important tool is model search. This tool, beyond its utilization in pure model localization for model use, is also important for model composition and model exchange. Apart from that it allows the localization of existing models before the construction of a new one, avoiding the construction of a model that already exists in the system database. Figure 2 presents the screen with the result of a model search

Another important feature of ModManager is its ability to register model taxonomies. Model taxonomies are important steps for model cataloguing and search since they allow grouping of models by similarity, which may be useful in model reuse or in model composition. However there is no agreement on model taxonomies in science neither in a single area (like environmental research, for example). In environmental research there is neither a single taxonomy to classify a model nor even a consensus on the best taxonomy. So, ModManager allows the researcher to include the taxonomy that he/she uses to classify his/her models or uses a taxonomy that is a consensus in a major group in his/her area. Furthermore, ModManager is proposing the use of ontologies for taxonomy mapping so that models classified as of one taxonomy may be used in other areas of research via taxonomies' semantic similarity. These ontologies must be collaboratively constructed so that models can be used in different areas. Even the own taxonomy may be collaboratively constructed and used by the research group in the system, in order to obtain a kind of consensus in model classification.

Thus, beyond simple search using taxonomies, the use of ontologies to map different taxonomies may lead to an increase in the power of model search, introducing some kind of semantic search based on taxonomies. It would allow, for instance, a search result containing models of different areas that could be used to compose a new model for another research area.

As an example, the model of terrain gradient may be classified in a taxonomy of the Cartography area, but may be used in several other areas, like Hydrology, Agriculture, Urban Planning, and so on.

Another use of ontologies in MODENA is the semantic search for models. A search like that could work beyond exact word search and retrieve models similar to that being searched or models with similar aim, properties, programs, workflows, data, and so on.

According to what is stated above, all knowledge relating to the model can be associated to it in several steps since model pre-design and design to model construction, validation and use, during the whole modelling lifecycle. Moreover, the system also permits knowledge validation, that is, all the knowledge associated with the model can be validated by specialists to assure its content. In this case chats and discussion forums can also be used to improve and/or validate the knowledge.

As a form of increasing the knowledge base of a model, the system also registers the history of modification and execution of the model, so that researchers can use it to learn more about the model and use that knowledge in future utilization of the model or even in the development of new ones.

4. Concluding Remarks

One of the main features of MODENA is its easy-to-use Web interface that provides a human centered collaborative environment for researchers that are geographically distant to interact and improve the design, construction, validation and use of their models.

The usage scenario mentioned here also shows the application of the characteristics of collaboration among distributed and interdisciplinary researchers in order to reach the best model for their study.

Copyright © Shaker Verlag, Aachen 2007. ISBN: 978-3-8322-6397-3

Bibliography

Brito, H.M., Strauch, J., Souza J.M. (2005a): ModManager: A Web-based system for Knowledge Management of Scientific Models (in Portuguese). Proceedings of the IV Brazilian Congress of Knowledge Management (KMBrasil). São Paulo, Brazil, Nov. 09-11, 2005a.

Brito, H.M., Strauch, J., Souza J.M. and C. Osthoff (2005b): Scientific models management in computational grids, in: Proceedings of the 17th International Conference in Scientific and Statistical Database Management, Santa Barbara, USA, June 27-29, pp. 113-116.

Cavalcanti, M.C., Mattoso, M., Campos, M.L., Llirbat, F. and E. Simon (2002): Sharing scientific models in environmental applications, in:. Proceedings of the 2002 ACM Symposium on Applied Computing, Madrid, Spain, 2002. pp. 453-457.

GEOMA Network 2006 http://www.geoma.lncc.br. (In Portuguese)

Lenard, M.L. (1993): A Prototype Implementation of a Model Management System for Discrete-Event Simulation Models. Proceedings of the 1993 Winter Simulation Conference (G. W. Evans, M. Mollaghasemi, E.C. Russell, W.E. Biles eds.).

Oliveira J. and J. M. Souza (2004): "Improving Knowledge Sharing through Knowledge Objects Representation". Proceedings of the IV International Conference on Knowledge Management, Graz, Austria, 2004.

Palma, S.J.M., Strauch, J.C.M, Souza, J.M. et al. (2000): Specs - A spatial decision support collaborative system for environment design, in: Proceedings of the V International Conference on Computer Supported Cooperative Work in Design, Hong Kong, Nov. 29 - Dec. 02, 2000.

Rimon, M., Keller, R.M. (1992): A Knowledge-Based Software Development Environment for Scientific Model Building. IEEE.

Vassalo, A., Oliveira, C., Osthoff, C., Brito, H.M., Strauch, J. and J. M. Souza (2006): "Execution Management of Scientific Models on Computational Grids". Proceedings of the I International Workshop on High-Performance Data Management in Grid Environments (co-located with VECPAR 2006). Rio de Janeiro, Brazil, July 13, 2006.

Wäger, P., Hilty, L.M. (2002): A Simulation System for Waste Management – From System Dynamics Modelling to Decision Support. In: Rizzoli, A.E.; Jakeman, A.J. (eds.): Proc. Integrated Assessment and Decision Support, Lugano, 2002, pp. 174-179.

Wyse, A.P., Bevilacqua L. and M. Rafikov (2004): Population Dynamics of An. darlingi in the Presence of Genetically Modified Mosquitoes with Refractoriness to Malaria. Proceedings of the IV Brazilian Symposium on Mathematical and Computational Biology / I International Symposium on Mathematical and Computational Biology, Ilhéus, Brazil, Nov. 27 – Dec. 01, 2004.

Copyright © Shaker Verlag, Aachen 2007. ISBN: 978-3-8322-6397-3