International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Towards integrated environmental model-based problem solving 2.0

Huub Scholten[a]
[a]*Wageningen University, Information Technology Group,*
Wageningen, The Netherlands, Huub.Scholten@wur.nl

Abstract: Environmental problem solving requires that problem-solving resources such as tools, models, data and its documentation can easily discovered and seamlessly cooperate, even if they are distributed and originate from different disciplines. The main hurdles to take include problematic collaboration within multidisciplinary teams, finding appropriate resources, and gluing them to a single virtual specific problem-solver. The first hurdle can be lowered by making knowledge explicit by organizing knowledge related to the problem, to the problem solving process and to the problem-solving model in a repository of domain ontologies and metadata, which is structured in layers, ranging from generic to detailed and specific. A second hurdle can be lowered by developing a framework that enables finding appropriate resources and compose them to end-user applications with workflows that guide team members through the problem solving process. Such a framework implicates interoperability of the resources in a technical, syntactic and semantic sense. Semantic interoperability can be based on semantic annotation of the resources and organizing this annotated knowledge in a Repository of Domain Ontologies and Metadata.

***Keywords**:* model-based problem-solving; web based services; ontologies; integrative modelling framework; semantic interoperability

1 MODEL-BASED ENVIRONMENTAL PROBLEM-SOLVING

While electronic exchange of information and remote collaboration are becoming the norm in most research communities, the circulation, communication and reuse of scientific resources, such as data, models, modelling tools, and its documentation remain cumbersome. Despite much talk about “intelligence-enabled” information (e.g. the semantic web), it is still quite common to share scientific resources often in conventional formats that are not meant for integration and self-consistency. Moreover, because of the ever increasing amount of knowledge produced, it has become extremely difficult to find and reuse data, models, and tools of other research groups that can be useful for someone’s own research.

A traditional approach in integrated modelling is to rely on a common integrated modeling framework (examples are OpenMI and CSDMS in the hydrologic domain, IGSM for global climate change, Cape-OPEN in the chemical engineering domain, and many others). This requires the various research teams to adhere to specific formalisms of the platform of choice and basically to cooperate in a single location at a given time in collaborative modeling sessions. Such dedicated solutions have limited value and can scarcely be reused by researchers that are geographically distributed in so called Virtual Research Communities (VRCs) that collaborate in model-based problem-solving and rely on information and communication technology.

As the complexity of the problems studied increases and the scale grows from local to global, a much wider collaboration among international research groups is required and to meet these requirements, web-enabled environments have been developed on a generally *ad-hoc* basis, using diverse standards for data, models, and problem-specific solutions. An example of such an approach can be found in water resource modelling, where a service oriented computing paradigm was followed in a single application domain (Granell et al., 2010, Goodall et al., 2011). This in turn results in serious challenges to reuse existing resources (especially data and models) for other applications or by other teams.

An additional complicating factor is associated with the increased need for supporting multidisciplinary problem solving. Over the last decades the character of model-based problem solving projects changed from single discipline, single person, academic oriented research model studies to multidisciplinary, decision support oriented projects, in which teams of members with different backgrounds and roles must collaborate. Such projects typically integrate subprojects belonging to various application domains and are executed by distributed teams.

Successful projects that aim to overcome (parts of) these difficulties have focused too much on a specific type of research or research activity that can be addressed by a dedicated approach (formats, tools, shared knowledge). This is however the very reason why such dedicated solutions can hardly be re-used by other VRCs or for other applications. On the other hand, generic (i.e. more ambitious) approaches, for instance those based on a complete ontology-based formalization of the knowledge across multiple disciplines, that can support a wide range of communities and applications, are typically difficult to use. An example of the latter is the SEAMLESS project (http://www.seamless-ip.org), where a dedicated team of researchers worked on the definition and management of a shared ontology, which was successfully used to formalize the interfaces across different modelling domains (Rizzoli et al., 2008). Yet, the practical knowledge of how to operate with the ontologies remained confined to a restricted group.

This paper proposes to overcome the difficulties and problems for integrated environmental model-based problem solving by developing an Environmental Modelling Infrastructure (EMI), consisting of resources (models, data, documents, tools) and of a framework with functionality to handle resources that are compliant with EMI. The resources will be deployed as registered web services, which users can find and compose to end-user applications that run on web servers or the European Grid Infrastructure.

2 CONCEPT AND OBJECTIVES

The *Environmental Modelling Infrastructure* (EMI) will consist of three parts: the *EMI Engine*, the *EMI Knowledge Repository* and a series of *EMI-compliant web services* (see Figure 1). It aims to enable and facilitate collaborative, participatory, environmental research by designing and deploying an Environmental Modelling Infrastructure that focuses on modelling. It has to become a virtual infrastructure for multidisciplinary environmental model-based problem solving, which has to be built as much as possible on existing technologies, methods and tools, but with innovative qualities on how to make knowledge explicit and how it supports discovery of resources and facilitates interoperability between them.

There will be two types of users: *Resource Providers* and *Application End-users*. Resource Providers are developers or owners of models, data, tools and research documents. The EMI should appropriately handle ownership, although sharing of resources is a central intention. Application End-users are typically modelers that use models to solve some environmental problem.

In order to make EMI a useful instrument for environmental model-based problem solving, a sufficient number of environmental resources (data, , models, documentation modelling tools, collaboration-support tools) have to be made available as EMI-compliant web services. Resources will be *semantically annotated* to facilitate discovery and use and will be *published* in the *EMI Knowledge Repository*.

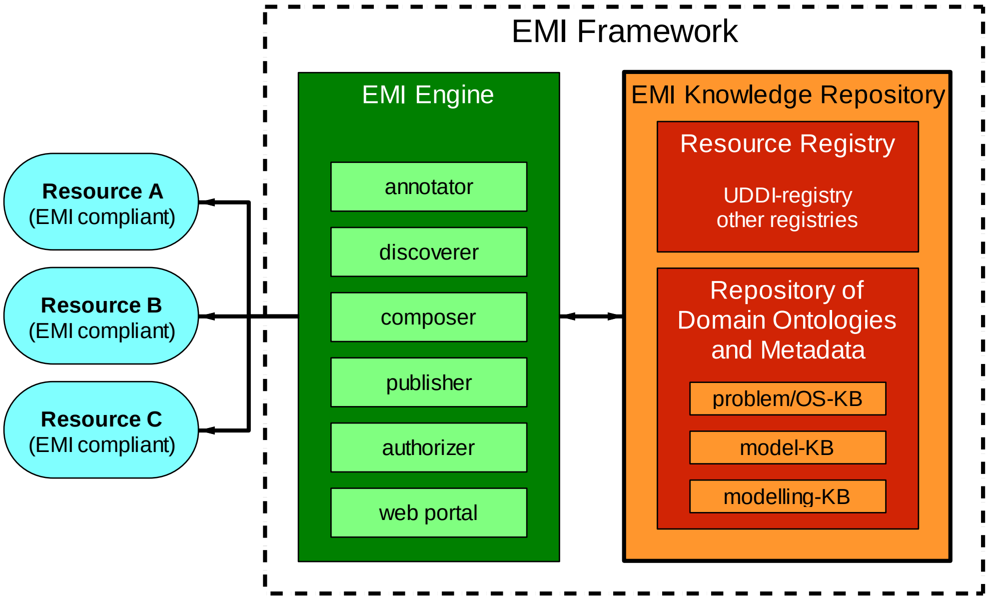

Figure 1. Basic outline of an Environmental Modelling Infrastructure (EMI).

The *EMI Engine* will help end-users to find resources, by supporting to define problems and will use the resulting problem definition to discover appropriate web services that can be integrated as resources of an end-user application, based on the information of the resources in the *EMI Knowledge Repository*, including its function and well-described interfaces. The *EMI Engine* will generate the End-user application and an associated workflow that sets the resources in the right order, allows automatic invocation of the web services and enables human interaction.

3 STATE-OF-THE-ART TECHNOLOGIES

The OASIS Open Reference Model for Service Oriented Architecture describes a Service Oriented Architecture (SOA) as a paradigm for organizing and utilizing distributed resources that may be under the control of different owners (OASIS, 2006). A service provider publishes his resource to a registry, where a potential user of a service can find it. After the user has found a service and its well-defined service interface, it can be bound and used to solve real world problems.
Throughout the last decade, the Web Service technology became the most widespread technology used to realize SOAs. EMI will be strongly based on web service technology and SOA, and will use existing formal and *ad-hoc* standards and protocols, originating from various organizations, e.g. OASIS, Open Geospatial Consortium (OGC) and others. These protocols and standards include but are not restricted to SOAP, WSDL, UDDI, WS-BPEL, BPEL4People and OWL.

4 WEB SERVICES

To make EMI successful a sufficient number of relevant resources should be available as web services and registered in the EMI Knowledge Repository (see section 6). First a distinction can be made between new resources that can be developed as EMI-compliant web services directly and legacy software that has to be converted to EMI-compliant web services, which can be done manually and semi-automatically (Kassahun et al., 2010).

Once a resource is available as web service and registered in the EMI Resource Registry, users can browse/search the registry to find resources that fit their needs. Although technical interoperability (i.e. technically correct connecting web services) and syntactic interoperability (i.e. exchange of information in the right format) are guaranteed using the highly standardized web service technology, the question of semantic interoperability (i.e. what is the meaning of exchanged information or does it make sense to feed a certain model with certain data) remains. Semantic interoperability can be based on semantic annotation of the resources (Villa et al., 2009, Rizzoli, 2010). Subsequently, this annotated knowledge has to be organized in the EMI Repository of Domain Ontologies and Metadata as is described in section 6.

Semantic annotation serves various purposes, as it enables discovery of the web services, its composition in a workflow and its execution. Semantic annotation is an essential resource of the Semantic Web concept. Semantic annotation brings together the social aspects of the world wide web with the enhanced representational capabilities provided by formal description of resources where the knowledge is structured and formalized in a set of *ontologies* (Gruber, 1993; Villa et al. 2009, Rizzoli, 2010). Semantic mediation defines a strategy to properly match model inputs and outputs to independent data sources and to enable reproducibility, type safety and uncoordinated extensibility to data, models and tools (Villa et al., 2009). It is evident that the use of semantic annotation and mediation can greatly facilitate the retrieval and operational reuse of scientific resources.

5 EMI ENGINE

The EMI Engine consists of a series of functional parts, depicted in Figure 1 and described in this section.

5.1 Annotator

Semantic annotation aims at adding rich semantics to resources (Villa et al., 2009, Rizzoli et al., 2010). It is based on the Semantic Web technology for annotation of unstructured information with standardization for data exchange, e.g. Resource Description Format (RDF), and ontology languages, e.g. Web Ontology Language (OWL). This approach has been extended to describe web services with semantics, e.g. Semantic Annotations for WSDL (SAWWSDL) and XML Schema (Rizzoli et al., 2010). This enables registering web services into a UDDI registry (see section 6.1) and storing domain specific knowledge of the web service into a Repository of Domain Ontologies and Metadata (see section 6.2).

WSMO and OWL-S are reference frameworks for semantic modelling of Web services that enable to model complementary domain specific WS ontologies using the WSML family of languages without annotating syntactic-based descriptions, but by providing grounding mechanisms between their semantic descriptions and syntactical descriptions of services (e.g. WSMO offers grounding to WSDL, SAWSDL, etc.).

The fast developing semantic annotation technology resulted in frameworks and tools that can enable annotating environmental resources for model-based problem solving, but these will not be described here. EMI’s annotator will be selected from or based on existing tools.

5.2 Discoverer

Once a resource is available as a EMI-compliant Web Service and registered in the EMI Knowledge Repository, the *EMI discoverer* enables users to browse/search the EMI Resource Registry in order to find resources adequate to their research endeavors. Technical interoperability is guaranteed using the highly standardized Web Service technology, using the description of the resource interfaces in the UDDI-based EMI Resource Registry (see section 6.1). In addition to simple browsing and searching to discover appropriate resources, the knowledge stored in the EMI Knowledge Repository can also be used to find the best resources and to advise the *EMI composer* (see section 5.3) on which resources to use and how to compose them to an end-user application. for environmental model-based problem solving.

5.3 Composer

After finding appropriate web services these have to be connected to an end-user application by a *Composer* using the well described interfaces stored in the EMI Knowledge Repository (section 6). The so formed chain of web services (data, model(s), tools, documentation) should be depicted in a *scientific workflow*, which is obviously based on the results of the discoverer and on the knowledge and information on the resources in the EMI Knowledge Repository. Scientific workflows to organize modeling components and web services have an increasing popularity, which appears among others from the Kepler scientific workflow system (Ludäscher et al., 2006) and from the Open Modelling Interface (OpenMI) Configuration Editor (Gregersen et al., 2007). EMI's scientific workflows will be based on WS-BPMEL technology and BPEL4People, the latter to enable human interactions with the workflow. Workflow and end-user application GUI elements can be presented to users as Web Portal (see section 5.4).

5.4 Other functional parts of the EMI Engine

Resource owners can use the *EMI Publisher* to provide their resources to EMI. First resources have to be annotated and then converted to EMI-compliant web services and registered in the EMI Knowledge Repository (section 6).

An increasing part of resources is publicly available, but the rest is proprietary. This requires an *Authorizer* to handle ownership and other authorization issues.

The main function of a *Web Portal* is to be the personalized central gateway to the rest of the functionality, i.e. annotator, discoverer, composer, scientific workflow and GUI's from end-user applications. Several *portlets* can be plugged into the overall portal to enable interacting with presentation-oriented web services, e.g. web server GUIs. This will be based on WSRP technology.

6 EMI KNOWLEDGE REPOSITORY

Two core parts of EMI Knowledge Repository can be distinguished, a registry, based on UDDI and a Repository of Domain Ontologies and Metadata with information that does not fit into the UDDI.

6.1 UDDI registry

Universal Description, Discovery and Integration (UDDI) is a platform-independent, XML-based registry for web service providers and a mechanism to register and locate web services. It was originally proposed as a core Web service standard, designed to be interrogated by SOAP messages and to provide access to WSDL documents describing the protocol bindings and message formats required to interact with the web services listed in its directory. It consists of three resources: *white pages* with information on the service provider, *yellow pages* providing the functionality of the service and *green pages* that describe how to access a web service with technical information on the service bindings.

6.2 Repository of Domain Ontologies and Metadata

The EMI Repository of Domain Ontologies and Metadata will contain data that does not fit into the UDDI registry, but that is related to the semantics of the interfaces of the EMI-compliant web services, registered in the UDDI registry. Preferably and/or in a later stage, more than just the interfaces are described in the Repository of Domain Ontologies and Metadata, especially more of the content of the registered models. Also scientific and other publications have to be included in the Repository of Domain Ontologies and Metadata.

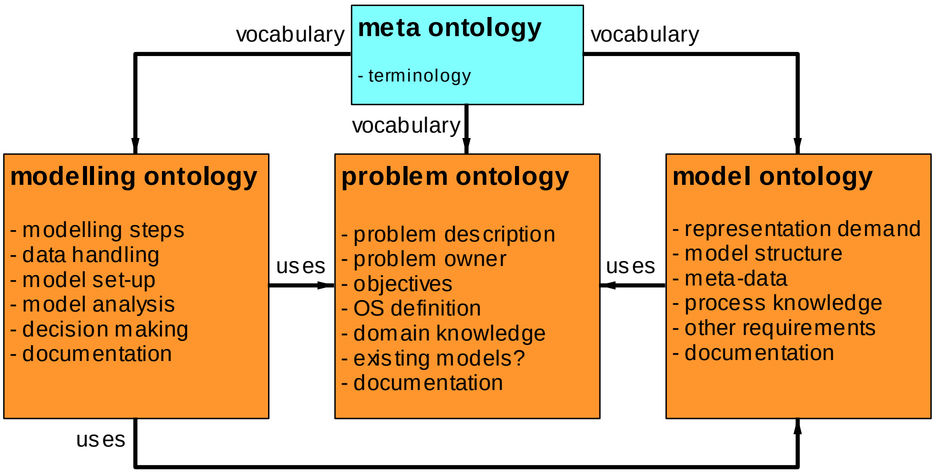

Figure 2. A structured ontology for multidisciplinary problem-solving, consisting of 4 interrelated ontologies. At a highly abstract level, the ontologies are concepts (presented as rectangles) and the relationships between the ontologies depicted as arrows.

The environmental domain that uses models for problem solving, is very complex with several contributing disciplines and several subdomains. Therefore, a clear distinction has to be made between bodies of knowledge to be stored in the Repository of Domain Ontologies and Metadata (Scholten and Beulens, 2012). The main distinction is between what is known and relevant for the problem at hand from the real world, referred to as the *problem ontology* and the knowledge included in the model, referred to as the *model ontology* (Figure 2). Furthermore, modelling is complex and environmental models can belong to a variety of modelling paradigms, each with its own approach and solving techniques. Therefore a third pillar is added concerning how to model according to Good

Modelling Practices. The fourth ontology provides basic terminology, needed to develop the other parts.

Each of the three main ontologies are structured in layers ranging from generic to detailed and specific (Scholten and Beulens, 2012).

6.3 Other registries

Besides semantically annotating and registering resources, existing registries (including D4Science, TaToo and GEON registries) and respective semantically annotated resources will be integrated into a meta-registry and, in case necessary, semantically enriched through indirect tagging.

7. DISCUSSION

The outlined Environmental Modelling Environment (EMI) is a functional blueprint of what can be developed from present technologies to realize an integrated environmental model-based problem solving environment. It should facilitate reuse of models and data, will allow using modelling tools developed by others and gives access to scientific and technical documentation. In this way it makes a more profound analysis with alternative models, tools and data within the reach of all virtual research communities tackling complex, real-world environmental problems that ask for multidisciplinary approaches.

A part of the proposed infrastructure is useful for environmental model-based problem solving, but also for other domains. Most of the functional parts of the EMI Engine are – to some degree – generic, as they are not specific for the environmental application domain. Moreover, although essential for EMI, they are not new and their development can strongly rely on existing software. An exception may be the EMI discoverer, as it will not only browse resources and search in EMI's UDDI-based Resource Registry, but will also use the content of EMI's Repository of Domain Ontologies and Metadata, which contains knowledge on (environmental) problems, models and modelling. This knowledge can be used to advise on which resources to select and how selected resources best can be composed to useful end-user applications.

A substantial part of EMI is specific for environmental model-based problem solving. This part will be determined by the series of environmental resources (models, data, tools and documentation) that are made EMI-compliant, registered in the EMI Resource Registry and which knowledge is made explicit and stored in the EMI Repository of Domain Ontologies and Metadata. Gathering and structuring this knowledge is probably most difficult, as it requires semantically annotating relevant environmental resources. This will cost a lot of effort and is – at least for the moment – largely a manual process.

A roadmap to realize environmental model-based problem solving 2.0 by developing EMI is difficult to sketch. In the beginning a sufficient number of relevant and useful resources have to be annotated and transformed to EMI-compliant web services in order to make it attractive for potential users. Initially, most of this work has to be done manually. In future stages, existing and semantically annotated resources will be (semi-)automatically wrapped to web services, fully registered in the EMI Knowledge Repository and made available on local servers of their owners/providers or on a public GRID.

Opposite to other e-science environments, whether single domain specific or more generic, EMI will be using ontologies that contain content related knowledge, which facilitates to deal more effectively with semantic interoperability.

This functional blueprint is not an attempt for a design of EMI. Especially the functions of the EMI Engine overlap partly and should be clearly redefined and based on state-of-the-art technology.
Developing EMI will accelerate moving towards environmental modelling 2.0.

ACKNOWLEDGMENTS

Most of the ideas in this paper have been discussed in the context of writing a proposal for FP7 funding by the European Commission with major contribution of Andrea Rizzoli (IDSIA, CH), Ioannis Athanasiadis (DUTH, GR), Marek Makowski (IIASA, A), Christoph Pflügler (Uni-Augsburg, D) and Ayalew Kassahun (Wageningen University, NL) and further input of researchers from Alterra (NL), PIK (D), IHE (NL), NIT (P), BC3 (E) and Seeconsult (D).

REFERENCES

Goodall, J.L., B.F. Robinson and A.M. Castronova, 2011. Modeling water resource systems using a service-oriented computing paradigm. *Environmental Modelling & Software*, 26:573-582, 2011.

Granell, C., L. Díaz and M. Gould, Service-oriented applications for environmental models: Reusable geospatial services. *Environmental Modelling & Software*, 25:182-198, 2010.

Gregersen, J.B., P.J.A. Gijsbers and S.J.P. Westen. OpenMI: Open modelling interface. Journal of Hydroinformatics, 9:175-191, 2007.

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

Kassahun, A., I.N. Athanasiadis, A. Rizzoli, A. Krause, H. Scholten, M. Makowski and A.J.M. Beulens, Towards a service-oriented e-infrastructure for multidisciplinary environmental research. In: D.A. Swayne, W. Yang, A.A. Voinov, A. Rizzoli and T. Filatova (Editors), Proceedings of the iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software (iEMSs 2010). EMS, Ottawa, Ontario, Canada, 2010

Ludäscher, B., I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E.A. Lee, J. Tao and Y. Zhao, Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience*, 18:1039-1065, 2006.

OASIS. Reference model for service oriented architecture 1.0. Organization for the Advancement of Structured Information Standards, http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.pdf, 2006.

Rizzoli, A.E., M. Donatelli, I.N. Athanasiadis, F. Villa, and D. Huber, Semantic links in integrated modelling frameworks. *Mathematics and Computers in Simulation*, 78:412-423, 2008.

Rizzoli, A.E., G. Schimak, M. Donatelli, J. Hrebicek, G. Avellino, J.L. Mon and I. Athanasiadis, TaToo: tagging environmental resources on the web by semantic annotations. In: D.A. Swayne, W. Yang, A.A. Voinov, A. Rizzoli and T. Filatova (Editors), Proceedings of the iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software (iEMSs 2010). EMS, Ottawa, 2010.

Scholten, H. and A.J.M. Beulens, An ontological framework for model-based problem-solving. In: D.R. Dolk and J. Granat (Editors), *Modeling for Decision Support in Network-based Services*, Lecture Notes in Business Information Processing, 42, Part 4, Springer, Berlin / Heidelberg, pp. 226-256, 2012.

Villa, F., I.N. Athanasiadis and A.E. Rizzoli, Modelling with knowledge: a review of emerging semantic approaches to environmental modelling. *Environmental Modelling & Software*, 24:577-587, 2009.