In **2007 European Simulation Interoperability Workshop**, Genoa, Italy, Simulation Interoperability Standards Organization, June 2007, pp.60.

Semantically Rich Interfaces for Simulation Interoperability

Andrea-Emilio Rizzoli
Hongtao Li
Ioannis N. Athanasiadis
Dalle Molle Institute for Artificial Intelligence
IDSIA-SUPSI, Galleria 2.
CH-6928 Manno, Switzerland
andrea@idsia.ch, hongtao@idsia.ch, ioannis@idsia.ch

François Marechal
Laboratory for Industrial Energy Systems,
École Polytechnique Fédérale de Lausanne,
CH-1015 Lausanne, Switzerland
francois.marechal @epfl.ch

Keywords:
Simulation interoperability, semantic, ontology, integrated modelling and assessment, agriculture, energy system.

ABSTRACT: *The size and scope of simulation models is constantly growing in order to keep up with the requirements imposed by increasingly complex scenario analyses. Moreover, advances in the performance of simulation hardware, of communication software, and development frameworks, also offer unprecedented opportunities. As a consequence, the complexity of the design, implementation and deployment of simulation models is also increasing, up to a point that it might soon become unmanageable.*
We describe an approach aimed at taming such complexity based on the use of ontologies to structure modelling and simulation knowledge, which can be manipulated through a knowledge manager to facilitate the development of models and tools for simulation.
We present two case studies and demonstrations of this approach. The first targets the integrated assessment of the common agricultural policies of the EU where models pertaining different domains (environmental, social, economic) and different scales (local, regional, continental) must be integrated. The second focuses on the integrated design of energy supply systems, where alternative models for energy supply systems must be integrated in order to select the best combination to produce energy minimizing costs and environmental impacts.

1. Introduction

Designing and implementing a simulation model has always been a complex activity, requiring multi-disciplinary skills, but in recent years it has become even more complex and demanding because of the constant increase in both *computational* and *dynamic* complexity of the models under investigation.

We thus distinguish among different types of model complexity, and at least between computational complexity and dynamic complexity. The former can be thought as a measure of the numbers of instructions that must be executed to simulate a mathematical model. For instance, if we increase the discretisation of the mesh in the simulation of a 3D fluid transport process, we increase the computational complexity of the simulation. Dynamic complexity on the other hand is more subtle, since it involves the way we represent the relationships among the variables in our models.

A model, being an abstraction, approaches its subject from a specific point of view; particular assumptions and hypotheses about the phenomena involved are made. We therefore tend to neglect the full extent of causal chains and driving forces affecting the phenomena of interest and we strive for simplification, focalization and modularisation of the model construction process. In other words, we tend to keep its dynamic complexity at a minimum.

Yet, during the last decades, a number of models have been designed and implemented, and it has become natural to assemble them together in order to try to address more and more complex problems, in order to explore the interdependencies among problems, which we preferred to ignore, for sake of simplicity, and for our inability to cope.

In the specific domain of management of environmental resources, integrated assessments are becoming increasingly common. Their purpose is to evaluate and assess the impact of various kinds of policies (land use, water allocation, and so on) from the perspective of sustainability; that is assessing the impact on the environmental, social and economic sectors.

The above mentioned sectors have previously been analysed in isolation, and to evaluate their interactions, we need to integrate models across scales and disciplines. This process is neither easy, nor straightforward.

At least three main software innovations came to help in the past few years. One is component-oriented software engineering; then we have distributed client-server applications, and finally ontologies for the representation and processing of modelling knowledge.

Software Engineering promotes the concepts of reusing "components-off-the-shelf" [1], distributed computing [2], agent-based computing [3], service-oriented architectures and web services [4] to support the development of modular applications. The very same concepts are meant to be used to develop modular and integrated environmental software applications.

Components can then be deployed as web services by means of platforms such as Sun's J2EE or Microsoft's .NET, which overcome the initial problems encountered by interoperability architectures such as CORBA and bring the facilities to develop client-server applications on everybody's desktops.

However, software integration is not the sole necessary condition the interoperability of simulation models. Even if a set of (good) software model implementations are working together, this is not at all a sign that the compound model makes any sense from a modelling point of view and generates credible results.

In this paper we advocate that the interoperability of simulation models must be supported by an adequate management and processing of the modelling knowledge, which can successfully take place by means of ontologies. This work argues that sound integration of environmental models also requires automated coupling of the knowledge hidden behind each software implementation.

The paper is organised as follows: in Section 2 we briefly review the role of ontologies in knowledge representation, with a certain attention to modelling knowledge. In Section 3 we describe how ontologies can be used to facilitate model integration and simulation interoperability, by the semantic annotation of model interfaces. In Section 4, we introduce our case study: the development of a framework for the integrated assessment of agricultural policies at the European level. In Section 5 we outline the application of the proposed methodology for the integrated design of power supply systems. We finally draw the conclusions, outlining future perspectives.

2. Knowledge representation and ontology

The term of 'ontology' originates from philosophy. It was given a specific technical meaning in computer science rather than what it originally refers to. Thus '*an ontology*' instead of 'ontology' often used in such cases. *An ontology*, according to Gruber's definition later refined by Studer [5], is *a formal explicit specification of shared conceptualization*.

Conceptualization refers to the definition of an abstract system model by means of a representation language, capable of capturing the objects involved and their classifications, typologies of entities, and interactions and relationships among those.

Explicit means that the type of concepts used, and the constraints on their use are defined clearly and in detail, leaving no room for confusion or doubt. Formal refers to the fact that an ontology should be expressed in a (preferably machine-readable) mathematically based representation system. This means that a candidate conceptualization defined using an ontology is subject to a formal verification procedure that can prove if the it is sound with respect to the specifications. Shared reflects the notion that an ontology captures consensual knowledge, that it is not private of some individual, but accepted by a group [5].

Ontologies are best suited for applications that require interface interoperability in open environments, through loose integration of components and services. In such open environments, ontologies can be considered as a mediator among heterogeneous information sources and services. With respect to scientific computing and the modelling and simulation activities, we identify that ontologies may be valuable in application aspects, as the following:

(a) Domain-specific formalizations through structuring model specific knowledge as ontologies. Such activities may come along with the semantic annotation of the input and output requirements of simulation models.

(b) Ontology-based integration of models, tools, and data for cross scale and cross discipline integration;

(c) The automatic generation of meta-model at run time, through the interpretation of logical constraints.

In modelling and simulation activities for scientific computing, modellers are considered as communities of "knowledge workers" in the fashion of Warren [6]. Knowledge captured in scientific models can be specified using ontologies and through the semantic web, it can be ultimately exposed, shared, and reused, and combined properly for integrated studies, as we will indicate in the following sections.

3. Increasing simulation interoperability through semantically rich interfaces

Easy model linking and integration is a key feature that is advertised by most modelling frameworks. However, we advocate that simple integration in software terms is not enough for sound model integration. A software implementation of an environmental model does not take into account the full semantics of the model interface. Model's assumptions are not captured in a components' software interface. The information associated with the inputs, states, outputs and parameters is limited to their data type. For instance, a typical software implementation exposes as model interface arrays of doubles, integers, and strings, whose context is described in the software documentation, or, even worse, only in the variable names. However, this practice requires that someone have to read the documentation in order to understand how to reuse this model properly. This is because the model's knowledge related to its interface is not encapsulated in the actual interface of the software implementation in a self-explained fashion.

The vision of reusing model software implementations as off-the-shelf components requires the assumptions on the model interface to be represented in a rich, machine-readable format. In order to achieve sound model integration, each linkage should be verified not only at the low level of data type matching (which is the unique requirement for software integration), but also against the actual semantics (context and assumptions) related to model interface.

A rich model interface is required to capture not only datatypes of the interface variables, but also the modellers knowledge related to characteristic times, units, pre- and post- conditions, temporal or spatial dimensions and sampling rates. Ontologies can be utilized for expressing such complex specification of models interfaces.

As discussed also in [7,8], we can build upon ontology-based tools for automatically generating model interface code, which in turn may be used for wrapping model implementations. Embedded or plug-in capabilities on (semi-) automated processing of domain specified knowledge represented by ontologies could further enhance the capability of model/tool development in a declarative fashion such as runtime meta-model generation. Along this line of research we find the work of Tolk et al. [9], who demonstrate the usefulness of an ontological representation for the conceptual interoperability of models.

Ontologies can therefore be used for specifying model interfaces, as described in Athanasiadis et al. [8] where model inputs, outputs and parameters are defined with respect to an upper model interface ontology. In this way, algorithms computing the numerical solution of the model equations are kept separated from the declarations of the model interface.

In an open simulation environment, all model components can be made available as services that comply to such a common specification of their interfaces, and ontology-based facilities can be built for discovering available models and data, taking benefit of the rich information specified in their interfaces. Furthermore, model knowledge stored in the ontology can be used both for software documentation and provide functionalities which go beyond the computation of model variables.

4. Development of a component-based agricultural and environmental policy assessment tool that implements a semantically enriched integrated modelling framework

Integrated Modelling and Assessment (IMA) provides a systematic, inter-disciplinary approach to inform coherent and holistic decision-making, by means of flexible integration of cross scale and dimension 'reusable' model components and datasets, based upon state-of-the-art software development strategies, architectures and tools that implement proper computational and artificial intelligences [10].

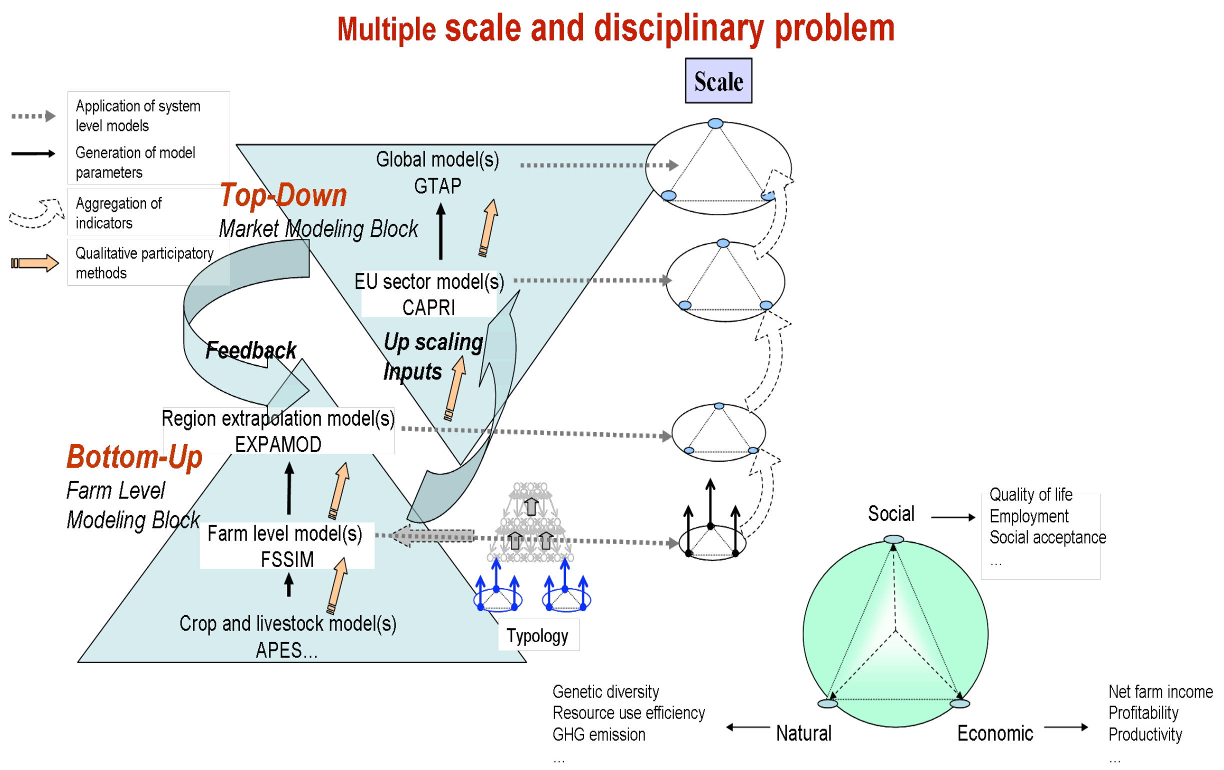

Fig. 1. SEAMLESS *IMA Framework* (SEAMLESS-IF) – A framework tackling complex multi-scale and disciplinary agriculture and agro-forestry assessment problem

The EU FP6 funded SEAMLESS[1] project is developing an *IMA framework* (SEAMLESS-IF) which integrates approaches from economic, environmental and social sciences to enable the assessment of the impact of policy and behavioural changes and innovations in agriculture and agro-forestry at different scales from farm level up to regional and global levels [11]. Within SEAMLESS-IF, various multi-paradigm model components and databases have been/will be developed or adapted, and integrated as illustrated in Figure 1. An integrated modelling framework – SEAMFRAME is therefore developed to facilitate these needs.

It is commonly accepted that integrated modelling frameworks offer a powerful tool for modellers, researchers and decision makers, since they allow the management, reuse and integration of models from various disciplines and at different spatial and temporal scales. However, the actual reusability of models depends on a number of factors such as the accessibility of the source code, the compatibility of different binary platforms. As explained, a remedy to these problems involves the specification of rich model semantics, by means of ontologies.

In the next subsections we detail how the SEAMFRAME architecture is organised and we introduce its key parts: the modelling environment, the processing environment, and the knowledge base.

4.1. Model components and the modelling environments

The interfaces of SEAMLESS models *have been defined explicitly in an ontology-enabled knowledge base*. The semantically rich definition of the interface can be used to generate the source code of *domain classes*, which are the interface definition of the models, according to the component-based software engineering approach.

The interface definitions are implemented in model wrappers, which deliver the execution of the model codes.

In this way, models are seen as components that can be automatically integrated into the SEAMFRAME modelling framework, as shown in Figure. 2.

The SEAMLESS model components are developed in different *modelling environments* by using different

[1] SEAMLESS denotes "System for Environmental and Agriculture Modelling, Linking European Science and Society", see http://www.seamless-ip.org.

programming languages – the agriculture management module (FSSIM-AM) in Java, the farm economic mathematical programming model (FSSIM-MP) in GAMS; the agricultural production and externalities simulator (APES) in C#, etc.

Thanks to the above mentioned domain classes, models are wrapped up as *linkable components* (also implementing the *OpenMI*[2] interface) so in order to be connected together in a workflow and communicate with other linkable model components. This happens in the processing environment.

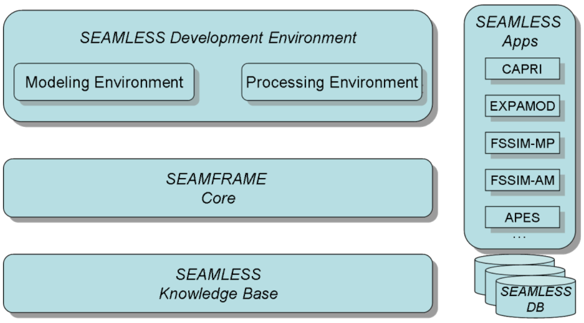

Fig. 2. Semantically enriched ontology-enabled SEAMFRAME architecture

4.2. The processing environment

The *processing environment* is a software application built on top of SEAMFRAME. Its purpose is to let the user apply processing tools to models and data in order to execute operations such as simulations, optimisations, scenario analyses and so on. The core of the processing environment is the *composition engine*, which is a piece of software that orchestrates the ordered execution of the *linkable components*. It also has the task of initializing all the linkable components according to their configurations.

4.3 The knowledge base

SEAMLESS Knowledge Base (KB) is a knowledge repository, which embraces the ontologies describing the data structures of SEAMLESS model components.

Protégé was used as ontology editor and Figure 3 displays part of the topological structure of the SEAMLESS ontology.

[2] OpenMI stands for Open Modelling Interface and Environment, a standard for model linking in the water domain. Details on OpenMI can be found at http://www.openmi.org/.

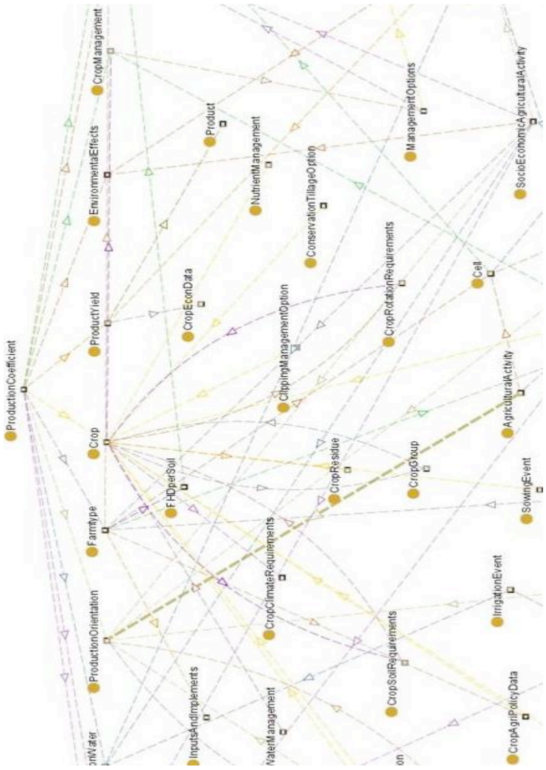

Fig. 3. Part of SEAMLESS ontology

The domain manager, processing the knowledge contained in the KB, has been developed to facilitate the development of OpenMI compliant model components. It provides assistance in domain-specific model development through the automatic generation of code templates for model interfaces (domain classes) and the model wrappers, which implement the interface required by the processing environment. An example is reported in detail in [8].

As previously said in Section 4.1, the model component wrappers both implement the interface that allows model linking and provide with access of legacy code, that allows the execution of models written in diverse programming languages. Model wrappers can access an instance of a domain class at runtime to feed the model component with the appropriate inputs before invoke the model engine [10]. The model wrapper therefore can (a) initialize the model component right after the start of the execution of the workflow; (b) scale model inputs in order to generate internal model execution controller; (c) dynamically prepare the meta-models that describe the model specifications (e.g. modules and equations to be used, sets definitions, how selected modules are structured, etc) according to a generic model template and the inputs that the wrapper received at run time; (d) dynamically prepare the model inputs data in an exact format the model needs; (e) retrieve model outputs and scale them to the required level (e.g. aggregation) to be persistently stored in the database or to be communicated with other linkable model components.

It is therefore thanks to the domain manager that semantically-enriched SEAMLESS model components can be linked and executed by the processing environment.

4.4 Current state of development and future activities

SEAMFRAME has been successfully implemented for facilitating the model development and integration of the first prototype of SEAMLESS-IF. In the new prototype, currently under development, the SEAMLESS ontology-enabled knowledge base will be integrated with conventional domain-centric data models (Enterprise Java Beans) and object-relational mapping toolkits (Hibernate).

5. Ontology-enabled knowledge processing for integrated design of energy supply systems

Energy can be generated from a palette of difference sources and using a number of different conversion technologies, to transform electrical energy into thermal energy, solar energy into electrical energy and so on. It is therefore extremely important to achieve the optimal design of energy conversion technologies and systems at different scales, in order to optimize a number of objectives such as the quality of life, the maximal saving of energy, and of the local and global environment protection.

Pursuing these objectives is a great challenge [12] given the volatile and stochastic nature of the environment in which the problem is set. The dynamics of environmental legislations for emission trading are unclear, the fuel and electricity prices due to the market liberalization or resource shortages are extremely volatile, and the innovations and integration of energy conversion technologies are constantly advancing.

Designing an optimized energy supply system requires a deep appreciation of the increasing complexity of the system and of its components, constraints and objectives.

It is therefore highly needed to develop an integrated modeling and assessment (IMA) framework that could facilitate the researchers, the technology suppliers, the local communities, the public service utilities and investors, and the policy makers, to perform a holistic assessment of energy supply system designs.

The framework should provide modelers with a variety of different modeling paradigms (process-based versus mechanistic black-box models, deterministic versus stochastic, and so on), different model domains (thermodynamic, economic, emission, etc.). The framework should also provide a number of tools to perform system analyses (single objective mix integer linear programming and multi-objective evolutionary algorithm based optimization), and finally, it should also provide domain-specific knowledge on how to match a problem description with the tool able to solve it.

For example, when designing an energy supply system, the feasible technological supply options and their integration possibilities could be formulated based on spatial and temporal specific conditions (demands, fuel availabilities, etc) and domain specific expert knowledge. Then, the interaction with energy flow models (process simulation models or performance meta-models), process integration and optimization tools are implemented to generate the optimal (or 'Pareto' optimal) design of the system and generate its 'superstructure' (i.e. corresponding market available energy supply technologies involved and their sizes and integrations). Based on generated 'superstructure' meta-models, further multi-criteria assessment of the energy supply system to be designed, or the coupling with demand side options, or the extrapolation of the designed system to a higher spatial scale (e.g. regional) can be performed.

Attempt on sharing energy flow model components throughout the web based on a so-called distributed object-oriented modeling environment (DOME) can be dated back to 1999 – within the frame of a collaborative project between the Industrial Energy Systems Laboratory of the Swiss Federal Institute of Technology Lausanne (LENI-EPFL), the CADlab of Massachusetts Institute of Technology, and laboratories from the University of Tokyo. DOME is used as information based collaboration platform for supporting the decentralized integration of models built independently under various application tools and operating systems by worldwide developers in this domain. It provides system engineers with the ability to access experts' models over the internet through user-friendly interfaces, and was successfully implemented to simulate the design of a district heating system [13]. Development of web services was then proposed as the means for remote interoperable access of model and data components further improved the accessibility of the domain specified knowledge and data. However, the problems of "lack of semantic annotations" still remain. The platform and web services only provide protocols and interfaces descriptions for services in a rigid way (e.g. concepts mapping) that is difficult to adapt to changing environments without human intervention. Runtime meta-model (e.g. superstructure) generation of an energy supply system and the automated/semi-automated model selection and integration specified by these meta-models are not supported either, which requires the domain specified expert knowledge.

Ontologies have the potential to address this deficiency effectively. A semantically-enriched IMA framework for energy system design built upon *structured domain specified knowledge* has been recently proposed by the authors, based on the development experiences on SEAMFRAME and other mentioned approaches.

Thanks to an ontology-based approach, models, tools and data components within this framework can be implemented as self-described web services conforming to open interfaces and strict contracts. The ontology can be used to annotate not only the interfaces of the model components but also the interfaces of data components and tools. Meanwhile, (part of) domain specific expert knowledge for superstructure generation will be incorporated within the knowledge base represented by description logics, e.g. the relation between concepts and restrictions on specific instances of an ontology class, etc.

Model components can thus be easily extended, reused, or integrated and executed with other model and data components or evaluated by e.g. process integration or multi-objective optimization tools to generate ‘superstructure’ meta-model, based upon which, automated/semi-automated model components selection, integration and execution throughout the web can be implemented.

Such an IMA framework is under conceptual design and evaluation, and will be further developed and implemented in later projects.

6. Discussion

In this paper we have presented an approach to enhance the interoperability of simulation models by the semantic annotation of their interfaces. This is achieved by the use of ontologies to represent knowledge about the model interfaces, that is, the set of inputs and outputs that need to be connected to exchange information among models.

By means of two case studies, one related to agriculture and the second to the energy domain, we show how the use of semantically enriched model interfaces can facilitate the interoperability of models, which have been originally developed for independent deployment.

7. References

[1] C. Szyperski, D. Gruntz, S. Murer: Component Software: Beyond Object-Oriented Programming, 2nd Edition, ACM Press, 2002.

[2] H. Attiya, J.L Welch: Distributed Computing – Fundamentals, Simulations, and Advanced Topics, John Wiley & Sons, New York, 2004.

[3] M. Luck, P. McBurney, O. Shehory and S. Willmott: Agent Technology: Computing as Interaction (A Roadmap for Agent Based Computing), AgentLink, 2005.

[4] T. Erl: Service-Oriented Architecture : A Field Guide to Integrating XML and Web Services, Prentice-Hall, 2004.

[5] R. Studer, V.R. Benjamins, D. Fensel, Knowledge Engineering: Principles and Methods. IEEE Transaction on Knowledge and Data Engineering 25(1-2):161-197. 1998.

[6] P. Warren: “Knowledge Management and the Semantic Web: From Scenario to Technology”, IEEE Intelligent Systems, Vol. 21, No. 1, pp. 53-59, 2006.

[7] I.N.Athanasiadis, S. Janssen, D. Huber, A.E. Rizzoli, M. van Ittersum: “Semantic modelling in farming systems research: The case of the Agricultural Management Definition Module”. In J. Marx-Gomez et al (eds.) Information Technology in Environmental Engineering (ITEE 2007) Springer Verlag, pp.417-432 2007.

[8] I. N. Athanasiadis, A.E. Rizzoli, M. Donatelli, L. Carlini: “Enriching environmental software model interfaces through ontology-based tools”, International Journal of Applied Systemic Studies, accepted for publication, to appear in 2007.

[9] A. Tolk, C. Turnitsa, S. Diallo: Implied ontological representation within the levels of conceptual interoperability model, Journal for Intelligent Decision Technologies, to appear.

[10] H. Li, K. Louhichi, S. Janssen, A. E. Rizzoli, I. N. Athanasiadis, E. Meuter, D. Huber: “Development and Application of a Component-based Generic Farm System Simulator Implementing a Semantically Enriched Integrated Modelling Framework”. In proceedings of the International Symposium on Environmental Software Systems 2007, May, 2007, Prague.

[11] M. van Ittersum, F. Ewert, et al: Integrated assessment of agricultural and environmental policies – towards a computerized framework for the EU, In: Proceedings of 3rd Biennial meeting of

the International Environmental Modeling and Software Society. Vermont, USA, 2006.

[12] Li H., "Environomic Modeling and Multi-objective Optimisation of Integrated Energy Systems for Power and Cogeneration". PhD thesis, Swiss Federal Institute of Technology Lausanne (EPFL), No.3657, 2006.

[13] M. Burer, H. Li, D. Favrat, S. Kraines, D. Wallace: "District heating in the city of Tokyo: a case study using a new integrated simulation environment" ASME International Mechanical Engineering Congress and Exposition: Symposium on Thermodynamics and the Design, Analysis, and Improvement of Energy Systems', November 11-16, 2001, New York, NY

Acknowledgements

This publication has been partially funded under the SEAMLESS integrated project, EU 6th Framework Programme for Research, Technological Development and Demonstration, Priority 1.1.6.3. Global Change and Ecosystems (European Commission, DG Research, contract no. 010036-2). For further reference, the SEAMLESS project website is: http://www.seamless-ip.org.

Author biographies

Dr ANDREA E. RIZZOLI is senior research scientist at IDSIA and Professor of Computer Science at SUPSI. He was awarded the "Laurea" (Master) degree in Control Engineering in 1989 and the Ph.D. in Control Engineering and Informatics in 1993 from Politecnico di Milano (Italy). His interests are in simulation and in decision support systems. Recipient of the "Early Career Research Excellence" award by the Modelling and Simulation Society of Australia and New Zealand. Editor of the Software section of the international journal Environmental Modelling and Software.

Dr HONGTAO LI is research scientist at IDSIA. Dr Li was awarded the Master in Engineering in 1995 from North China Electric Power University (China), Master in Energy in 2003 and Ph. D. in Mechanical Engineering in 2006 from Swiss Federal Institute of Technology Lausanne (Switzerland). His research interests are in multi-criteria integrated modelling and assessment, with special focuses in energy domain. He has worked on industrial and academic research projects, and successfully developed approaches and models for highly complex energy conversion system design evaluations and operation management, under uncertain economic and environmental legislating circumstances.

Dr IOANNIS N. ATHANASIADIS is research scientist at IDSIA. His areas of expertise include environmental informatics, software agents, semantic modelling, decision support systems and machine learning. Ioannis holds a Diploma and a Ph.D. in Electrical and Computer Engineering, both from the Aristotle University of Thessaloniki. In June 2004, he has been awarded the iEMSs Student Prize in Integrated Systems. He is serving as an associate editor of the International Journal of Advanced Systemic Studies (InderScience) and reviewer for several scientific publications.

Dr FRANCOIS MARECHAL is currently 'Maître d'Enseignement et de Recherche' in the Industrial Energy Systems Laboratory (LENI) at the Swiss Federal Institute of Technology Lausanne (EPFL). He was award Ph. D. in engineering with outstanding distinction. His interests include development of methodologies on computer aided process and energy conversion systems analysis and synthesis; and applications of computer based methods for the analysis, the design and the optimisation of industrial processes and efficient energy conversion systems.