In **MODSIM 2007 Int'l Congress on Modelling and Simulation**, (Oxley, L. and Kulasiri, D., Eds.), MSSANZ, Dec 2007, pp.1959-1965.

Using Ontology to Harmonize Knowledge Concepts in Data and Models

Wien, J.J.F. [1], M.J.R. Knapen [1], S.J.C. Janssen [2], P.J.F.M. Verweij [1], I.N. Athanasiadis [3], H. Li [3], A.E. Rizzoli [3] and F. Villa [4]

[1] Alterra, Environmental Sciences Group, Wageningen University and Research Centre, Wageningen
[2] Wageningen university, Plant Production Systems Group, Wageningen
[3] Dalle Molle Institute for Artificial Intelligence (IDSIA), Lugano
[4] University of Vermont, Vermont
Email: jan-erik.wien@wur.nl

Keywords: *ontology, modelling, model integration, integrated assessment*

EXTENDED ABSTRACT

Policy makers are confronted with complex, socially relevant problems. To increase the quality of their policy and realize a broader support for and understanding of actions, the policy making process has become a participatory process where policy measures need to be assessed in an integrated context (Rotmans and van Asselt ,1996).

In these integrated assessment studies, people work from different perspectives and domains, e.g. from an agricultural modelling perspective, an environmental perspective or an economic policy/problem perspective. Semantic interoperability is the key factor to integrate the knowledge of these different domains and perspectives in a computerized framework. Semantic interoperability is the ability of systems/components to share and understand information at the level of formally defined and mutually accepted domain concepts (Sølvberg, 1998). The specification of these concepts is done with ontology.

One of the most cited definitions of ontology is from Gruber (1993): “an explicit and formal specification of a conceptualization”. A “conceptualization” can be considered as an abstract model for a phenomenon identifying the relevant concepts. All concepts are explicitly described in a formal machine readable language.

The challenge in integrated modeling is the conceptual integration. To achieve this, we need explicit semantics and a shared conceptualization. For this we need to tackle the different perceptions and interpretations of people involved. Different modeling approaches, different formalism and last but certainly not least, the different integration requirements and ambitions need to be taken into account.

In the SEAMLESS project (Van Ittersum et al., 2007), the process used for creating a common ontology for models, indicators and raw data is based on a participatory and collaborative approach. A dedicated taskforce was created with participants from different parts of the project, and coordinated by the work package charged with integration. This task force envisages to develop a common knowledge base that represents a shared conceptualization between the different databases, models and indicators, with adequate meta-data.

A core component, called the SEAMLESS Knowledge Manager (KM) (Villa *et al.*, 2007), provides functionality as an extensible semantic modeling toolkit. It loads meaning through OWL (Ontology Web Language) ontology and transparently connects formal concepts to software objects and literals. Scalable use of machine reasoning effectively integrates an object-oriented framework, an object oriented database system and an ontology-based knowledge management environment into a SEAMLESS whole.

The development of the SEAMLESS common ontology was and still is a big challenge. By putting ontology in a central position in the project and the systems architecture, this shared conceptualization is the basis for generating (Java) source code for the object classes representing all the concepts and representing the objects in relational database tables. The use of ontology has proved to be very useful if not essential both for the technical integration of knowledge in the SEAMLESS Integrated Framework and in understanding the meaning of communicated words of the diversity of people within the project.

1. INTRODUCTION

Policy makers are confronted with complex, socially relevant problems. By handling these problems, it is a general tendency to increase participation of citizens and other stakeholders. In this way policy makers want to increase the quality of their policy and realize a broader support for and understanding of actions. In this new governance concept, the policy making process is the product of complex interactions between governmental and non-governmental organizations, each seeking to influence the collectively binding decisions that have consequences for their interest. Policy making is more and more a process of cooperation and participation in which the policy maker becomes a facilitator of the process. This concept is based on the assumption of the model of "co-production of knowledge" (Callon, 1999).

To account for this new governance, policy measures need to be assessed in an integrated context. Rotmans and van Asselt (1996) defined Integrated Assessment as "an interdisciplinary and participatory process combining, interpreting and communicating knowledge from diverse scientific disciplines to allow a better understanding of a complex phenomena. In this interdisciplinary and participatory process, information needs to be accessible in the way that all different types of stakeholders achieve a mutual understanding of the problems, objectives and solutions. But this mutual understanding across disciplines is often hindered by jargon, language, past experiences and presumptions of what constitutes persuasive argument, and different outlooks across disciplines or experts of what makes knowledge or information salient for policy makers or policy assessments (Cash et al., 2003).

This paper will describe the problems and (partial) solution of conceptual misunderstanding and knowledge integration in Integrated Assessments both from a theoretical perspective (section 2) and practical experiences with ontology based model integration (section 3). It will explain the challenge and the process used in a number of European integrated projects (e.g. SEAMLESS, AquaStress), for creating an ontology for (linking of) projects, models, indicators and raw data.

Section 4 gives a description of the role of ontology in the architecture and design of the integrated framework of SEAMLESS.

2. CHALLENGE IN INTEGRATING KNOWLEDGE

2.1. Theoretical perspective

Interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged (IEE,1990). There is often a distinction between syntactic, structural and semantic interoperability (Sølvberg, 1998).

• Syntactic interoperability is defined as the ability of two or more systems/components to exchange and share information by marking up data in a similar fashion (e.g. using XML).

• Structural interoperability means that the systems/components share semantic schemas (data models) that enable them to exchange and structure information (e.g. using RDF).

• Semantic interoperability is the ability of systems/components to share and understand information at the level of formally defined and mutually accepted domain concepts.

Semantic interoperability requires the correct interpretation and mutual understanding of all transferred information. In order to obtain mutual understanding of interchanged data, the actors have to share a model of what the data stand for. Semantic interoperability is about how to achieve such mutual understanding (Sølvberg, 1998).

One of the earliest theories dealing with understanding and remedies for misunderstanding is Richards's Meaning of Meaning Theory (Ogden and Richards, 1923). Instead of focusing on the information that is communicated, Richards wanted to study the meaning of the words. He felt that understanding is the main goal of communication and communication problems result from misunderstanding.

One of the ideas behind the Meaning of Meaning Theory is "The Proper Meaning Superstition" (Ogden and Richards, 1923). This is the false belief that every word has an exact, "correct" meaning. Richards says that the Proper Meaning Superstition is false because words mean different things to different people in different situations. This misunderstanding causes problems when two people believe they are talking about the same thing, but actually talk about different things.

Another concept that Richards uses is the idea of signs and symbols in communication. Words are examples of such symbols. Symbols have no natural connection with the things they describe (Griffen, 1997). Words are symbols of something because they have been given meaning. But very often words mean one thing in a certain context and mean another thing in a different context. This is why it is so important to study the context to get a better understanding of the meaning.

To come to this better understanding, Richards invented the Semantic Triangle. This triangle shows the relationship between symbols and their referent. One part of the triangle is the symbol, or the word. Another peak on the triangle is the thought or reference. This is the words that one would use to describe the referent. The referent, the last part, is the thing that one would picture in his mind.

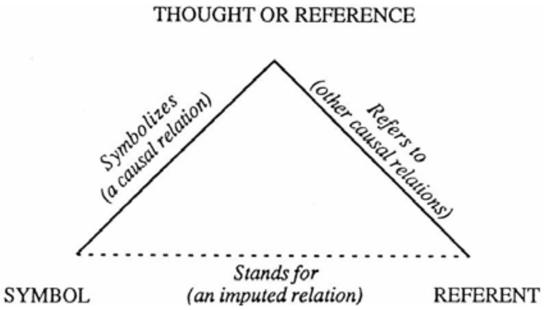

Figure 1. Semantic Triangle by Richards (Ogden and Richards, 1923).

An example in integrated agricultural modeling would be "wheat". The words to describe the referent can be for example "cereal crop, grain, flour". The referent, the last part, is the thing that one would picture in his mind.

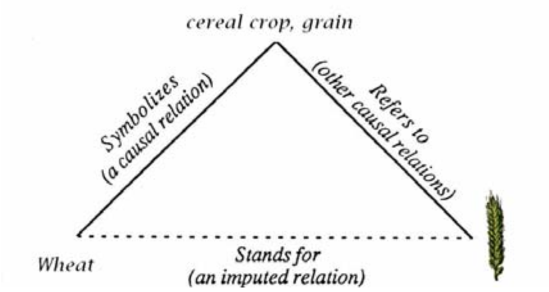

Figure 2. Semantic Triangle for wheat.

Understanding that people mean different things when they say the same thing is an important concept for people to understand. Richards gives ways to solve this problem of ambiguity. One of them is to give a definition. Definitions are words used in place of another word to explain the thought in a person's mind. Another option to understand the meaning by using a metaphor. A metaphor can help to clarify what each person is saying.

Feed forward is also an important factor when trying to avoid misunderstanding. Feed forward is when the speaker thinks of how his audience will react to what he is about to say and adjusts his words accordingly (Ogden and Richards, 1923). Feed forward forces the speaker to consider the experiences of the audience in order to better explain what they are saying.

To enable semantic interoperability in integrated modeling, the problem of semantic conflicts or semantic heterogeneity needs to be solved. For this ontology will be used. The term ontology is borrowed from philosophy, where ontology is a systematic account of Existence.

One of the many definitions of ontology is from Neches et al. (1991) "an ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary". One of the most cited definition is from Gruber (1993): "an explicit and formal specification of a conceptualization".

A "conceptualization" is explained as an abstract model of a phenomenon, identifying the relevant attributes. A conceptualization is an abstract, simplified view of the world that we wish to represent. Every knowledge base or knowledge-based system is committed to some conceptualization. (Gruber 1995)

"Formal specification" refers to the fact that the language semantics are machine readable. Often this is done by use of W3C OWL (Ontology Web Language, Patel-Schneider et al., 2004). A formal specification helps to communicate the definition of terms in a context independent ways and formal language semantics allows some automated consistency checks.

2.2. Practical experiences of ontology based knowledge integration

In integrated assessment projects a large number of scientists co-operate working on a wide range of issues, like different type of models, data and databases, indicators, assessment problems and user interaction. This wide range of activities needs to be brought together, as the models should be using the data, the indicators should be based on model outputs and data, while the assessment problem needs to be consistently defined to parameterize or configure the models coherently. Also indicators need to link to the assessment problem to be clearly presented to the end-users. This leads to a complex integration problem in which many scientists from different domains should contribute to achieve one shared understanding of the integrated assessment procedure.

In such a complex integration task, different types of misunderstandings around the meaning of concepts can occur:

1. as the same concepts might be used for different meanings, for example area in a model and area in the database,
2. as different concepts might be used, which have the same meaning, for example an internal user and an integrative modeler,
3. as concepts might be used with an ambiguous meaning, for example scenario (Shoemaker, 1993),
4. As relationships between concepts might be understood in a different way, for example between the different spatial scales and administrative regions.

The challenge in integrated modeling is the conceptual integration. To achieve this, we need explicit semantics and a shared conceptualization. For this we need to tackle the different perceptions and interpretations of people involved. Different modeling approaches, different formalism and last but certainly not least, the different integration requirements and ambitions need to be taken into account.

3. USE OF ONTOLOGY

3.1. Common Ontology

Ontology helps to formalize the knowledge captured in and/or between models, in order to subsequently facilitate model development, testing and documentation (Scholten and Kassahun) and model re-usability and exchangeability (Rizzoli et al., 2005) and separates knowledge captured in the model from the actual implementation in a modelling language or software e.g. Java, FORTRAN, Mathlab, STATA, etc. (Gruber, 1993; Villa et al., 2006) or from the data in a database (Zander & Kächele, 1999).

The development of a common ontology by a group of researchers is a complex, challenging and time-consuming task (Farquhar et al, 1995; Gruber, 1993), that still remains a scientific challenge. Tools are available that help in ontology development and store the ontology once it was developed. To achieve ontological commitment, i.e. the agreement by multiple parties to adhere to a common ontology, when these parties do not have the same experiences and theories (Holsapple & Joshi, 2002) a collaborative approach should be used. Other approaches for ontology development are the inspirational approach, the inductive approach, the deductive approach and the synthetic approach (Holsapple & Joshi, 2002). A collaborative approach has the advantages that researchers from different disciplines are diverse in their contributions, which avoids blind spots and which has more chances of getting a wide acceptance (Holsapple & Joshi, 2002) and that it can incorporate the other approaches, e.g. synthetic approach, as required for development of parts of the ontology.

3.2. Ontology Engineering

SEAMLESS is an integrated assessment modelling project (Van Ittersum et al., 2007), which aims to provide a computerized framework to assess the sustainability of agricultural systems in the European Union at multiple scales. The process used in the SEAMLESS project, for creating a common ontology for models, indicators and raw data is based on a participatory and collaborative approach. A dedicated taskforce was created with participants from different parts of the project, and coordinated by the Work Package charged with integration. This task force, called the DOT.force (Data and Ontology Taskforce) envisages developing a common knowledge base that represents a shared conceptualization between the different databases, models and indicators, with adequate meta-data. The DOT.force has knowledge engineers and domain members. These knowledge engineers have knowledge and experiences in the design and content of either databases or ontology. Domain members will be flexibly involved by this core group of knowledge engineers to develop specific parts of the ontology or database. The domain members hold knowledge about a specific domain, like a model, a database, indicators, scenarios, or the SEAMLESS-IF, which should be captured by the knowledge base and the database.

The knowledge engineers in the DOT.force initiate on a number of actions that should ultimately lead to a complete ontology. These actions are:

1. integrating the different databases into one SEAMLESS database,
2. clarifying the interfaces between the models, while adding relevant meta data,
3. linking indicators to model outputs in the ontology and retrieving required data from the database,
4. supplementing the ontology with additional meta-data on the concepts it holds, like units, minimum and maximum value, source and references,
5. developing a common ontology to cover concepts relevant for runs with the SEAMLESS-IF like scenarios, projects, scales, problems, running time, etc

Different methods are used to construct the ontology for the different actions. For actions 1 and 2 on databases and models, dedicated meetings are organized to develop the ontology, while for action 3 on indicators a proposal was made by the knowledge engineers, which was then evaluated by relevant domain members. Action 4 on metadata is carried out independently by domain members, once agreement on the common ontology has been reached between domain members and knowledge engineers. For action 5 on project and scenario definition an iterative process was used to develop a document, which is synchronized with a project ontology after each iteration.

4. SYSTEM ARCHITECTURE

This chapter describes how ontology plays a central role in the architecture of SEAMLESS-IF.

A core component, called the SEAMLESS Knowledge Manager (KM), provides functionality as an extensible semantic modeling toolkit. It loads meaning through OWL ontology's and transparently connects formal concepts to software objects and literals. Scalable use of machine reasoning effectively integrates an object-oriented framework, an object oriented database system, and an ontology-based knowledge management environment into a SEAMLESS whole. A tight API supports storage, search, retrieval and advanced management of semantically explicit software objects.

The use of explicit semantics in software allows attaining integration goals that are relevant to many disciplines and applications. By means of plug-in packages, the Knowledge Manager can be extended with knowledge and software functionalities to support specific semantic modeling tasks. The plug-in packages include for example:

1. Ontology and software support to handle measurement units, datasets, observations and re-definable spatial and temporal contexts;

2. Ontology and software support to handle several representations of time and space, including support for GIS data formats in both raster and vector form;

3. The OPAL framework to facilitate specification of semantic objects in automatically generated, customizable XML schemata;

4. Support for repositories of semantic objects that wrap common data formats and RDBMS/SQL database front-ends.

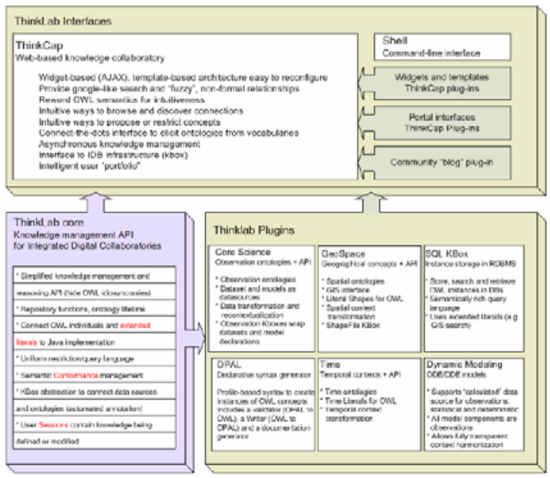

Figure 3. Knowledge manager (Villa *et al.* 2007, SEAMLESS PD 5.4.2.2).

The concepts relevant to the SEAMLESS domain (mainly the agricultural domain) and the models used in the SEAMLESS project have been put into ontology. This ontology, together with related ones (e.g. for measurement units), are loaded into the Knowledge Manager component. The Knowledge Manager can make the links between the models. The actual exchange of information is based on the Open Modeling Interface (OpenMI) (Gijsbers et al., 2005; Gijsbers et al. 2006) that provides a standardized interface to define, describe and transfer data between software components that run sequentially. This choice was made based on technical and functional requirements and the possibility to re-use legacy models.

In the current second prototype version of the SEAMLESS software the facilities of the Knowledge Manager are used off-line (not in runtime) to generate (Java) source code for the object classes representing all the concepts, and matching object-relational mapping files for use with Hibernate (www.hibernate.org).

In the generated source code Java annotations are used to record the connection to the ontology, like in this example:

```
@ConceptURI("http://localhost/ontologies/c
rop.owl#Crop")
public class Crop implements Serializable
{
        ....
        @PropertyURI("http://localhost/onto
logies/crop.owl#hasCropSoilRequirements")
        Public CropSoilRequirements
getCropSoilRequirements()
        ....
}
```

Since the annotations are accessible at runtime the ontology information can be used for reasoning, e.g. to validate whether an output of a model can be used as an input for another model.

The generated object-relational mapping files contain the necessary information for representing objects in relation database tables. For example names of tables and columns to use, data type of the fields, where to store properties of the objects and relations between objects.

In all practicality it is a fully automatically generated persistence layer for the SEAMLESS system. Higher layers of the system, and some of the models themselves, use it to retrieve, work with and store instances of the concepts as defined in the ontology.

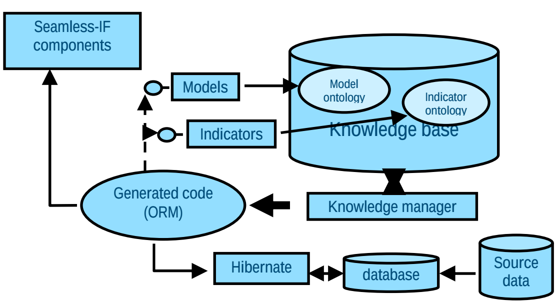

Figure 4. SEAMLESS integrated framework architecture.

The objective for the final version of the SEAMLESS system is to move from off-line use of the Knowledge Manager to a more integrated use at runtime, for example to be able to dynamically add concepts and to help users of different domains to perform integrated assessment tasks.

5. CONCLUSION

The challenge in integrated modeling is the conceptual integration. To achieve this, we need explicit semantics and a shared conceptualization. A participatory and collaborative approach is a key success factor for the creation of a common ontology for models, indicators and raw data.

The development of the SEAMLESS common ontology was and still is a big challenge that is performed by a dedicated taskforce. By putting the ontology in a central position in the project and the systems architecture, this shared conceptualization is the basis for generating (Java) source code for the object classes representing all the concepts and representing the objects in relational database tables.

The use of ontology has proved to be very useful if not essential both for the technical integration of knowledge in the SEAMLESS Integrated Framework and in understanding the meaning of communicated words of the diversity of people within the project.

6. ACKNOWLEDGEMENTS

We thank all scientists of the SEAMLESS project who contributed to the concept of using ontology and the development of the SEAMLESS common ontology. This work has been carried out as part of the SEAMLESS Integrated Project, EU sixth Framework Programme for Research and Technological Development, Contract No. 010036-2.

The sixth framework programme (FP6) is a collection of the actions at EU level to fund and promote research. The main objective of FP6 is to contribute to the creation of the European Research Area (ERA) by improving integration and co-ordination of research in Europe.

7. REFERENCES

Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., et al. (2003). Science and Technology for Sustainable Development Special Feature: Knowledge systems for sustainable development. PNAS, 100(14), 8086-8091.

Farquhar, A., Fikes, R., Pratt, W., & Rice, J. (1995) Collaborative Ontology Construction for Information Integration (No. KSL-95-63): Knowledge Systems Laboratory, Department of Computer Science, Stanford University.

Gijsbers, P.J.A., Gregersen, J.B., (2005) OpenMI: glue for model integration. In: Zerger, A., Argent, R.M. (Eds.), MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2005. pp. 648-654.

Gijsbers, P.J.A., Wien, J.E., Verweij, P., Knapen, R. (2006) Advances in the OpenMI. In: Proceedings of 7th International Conference on Hydroinformatics, HIC 2006. Nice, France, pp. 72-81.

Griffen, E.M. (1997). A first look at communication theory. New York: McGraw-Hill Companies Inc.

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5, 199-220.

Gruber, T. R. (1995) “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”, International Journal of Human and Computer Studies, 43(5/6), 907–928.

Holsapple, C. W., & Joshi, K. D. (2002). A collaborative approach to ontology design. Communications of the ACM, 45(2), 42-47.

IEE, (1990) Institute of Electrical and Electronics Engineers. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York, NY: 1990.

Van Ittersum, M. K., F. Ewert, T. Heckelei, J. Wery, J. Alkan Olsson, E. Andersen, I. Bezlepkina, F. Brouwer, M. Donatelli, G. Flichman, L. Olsson, A. Rizzoli, T. Van Der Wal, J.-E. Wien and J. Wolf (2007), Integrated assessment of agricultural systems- a component based framework for the European Union (SEAMLESS), Agricultural Systems In Press.

Neches, Robert, Richard Fikes, Tim Finin, Thomas Gruber, Ramesh Patil, Ted Senator, and William R. Swartout. (1991) Enabeling technology for knowledge sharing. AI Magazine, Vol.12, No. 3, Fall 1991.

Ogden, C. K. & Richards, I. A. (1923) "The Meaning of Meaning." 8th Ed. New York, Harcourt, Brace & World, Inc.

Patel-Schneider, Peter F., Patrick Hayes, Ian Horrocks. (2004) OWL Web Ontology Language Semantics and Abstract Syntax. W3C Recommendation. [Available at: http://www.w3.org/TR/owl-semantics/]

Rizzoli, A., Donatelli, M., Athanasiadis, I., Villa, F., Muetzelfeldt, R., & Huber, D. (2005). Semantic links in integrated modelling frameworks. Paper presented at the MODSIM 2005 International Congress on Modeling and Simulation. Melbourne, Australia

Sølvberg, A. (1998) Data and what they refer to. In P. P. Chen, editor, Concept Modeling: Historical Perspectives and Future Trends. Springer Verlag.

Schoemaker, P. J. H. (1993), Multiple Scenario Development: Its Conceptual and Behavioral Foundation, Strategic Management Journal 14(3), 193.

Scholten, Huub, Ayalew Kassahun, Jens Christian Refsgaard, Theodore Kargas, Costas Gavardinas and Adrie J.M. Beulens (2007) A methodology to support multidisciplinary model-based water management Environmental Modelling & Software, Volume 22, Issue 5, Pages 743-759