

Semantic mediation of an integrated assessment tool for agricultural systems

Sander Janssen¹, Ioannis N. Athanasiadis², Erling Andersen³, Andrea E. Rizzoli², Jan-Erik J.F. Wien¹ and Martin K. van Ittersum⁴

¹Alterra, Stichting DLO, sander.janssen@wur.nl, ²Dalle Molle Institute for Artificial Intelligence, ioannis.andrea@idsia.ch, ³Forest and landscape, University of Copenhagen, eran@life.ku.dk, ⁴Wageningen University, Plant Production Systems Group, martin.vanittersum@wur.nl

Abstract: Semantic mediation based on ontologies has been proposed as a method to achieve integration in multi-disciplinary and multi-scale integrated assessment tools. In the development of an integrated assessment tool, SEAMLESS-IF, agricultural models, data sources, scenarios and indicators had to be integrated in one conceptual schema, providing the shared definition of concepts and their relationships. This paper describes the different methods used in the semantic mediation of the various cases (i.e. models, data sources, scenarios and indicators) in SEAMLESS-IF. By contrasting the methods, this paper argues that the method used for semantic mediation has to be flexibly adapted to the case and the researchers involved. Common aspects in the semantic mediation are i. the organization in several prototypes, which were gradually improved, ii. a collaborative approach to discuss ontology content between domain experts and knowledge engineers and iii. the editing of the ontology itself was done by a knowledge engineer and not by a whole community of researchers. To succeed in building coherent and comprehensive ontologies, these common aspects (i.e. prototypes, collaborative approach, ontology editing) need to be adapted to the case (i.e. researchers, available information, mediation question).

Keywords: ontology, models, data sources, indicator, collaborative

1. Introduction

Semantic mediation has been proposed as a method to achieve integration in multi-disciplinary and multi-scale integrated assessment tools (Rizzoli, et al., 2008, Villa, et al., 2009). Such integrated assessment tools are often based on linked quantitative models into model chains. Through semantic mediation an agreed conceptual schema can be built between researchers for linking their models and the model interfaces are enriched with meaningful metadata on their behaviour (Villa, et al., 2009). In this case, the agreed conceptual schema serves as an additional knowledge layer enriching the original models and model chains and does not necessarily require the enrichment of the models themselves with tags or ontological constructs. Semantic mediation is a crucial challenge for any integrated modelling project (Jakobsen and McLaughlin, 2004, Bracken and Oughton, 2006, Scholten, 2008), as it provides consistent and transparent building blocks in definitions and terms required for the methodological and technical linking of models, data sources and indicators. Models from different disciplines have a different representation of data, space and time, and linking them implies that the outputs of one model have to be matched to the inputs of another model, while the modellers and their models should have a common understanding of the space and time, in which they operate.

Usually semantic mediation is based on ontology, which is a specification of a conceptualization for a system (Gruber, 1993). An ontology consists of concepts and their relationships (Antonioni and van Harmelen, 2004). Other options for building a joint conceptual schema are ad-hoc variable mappings, mathematical formalism (Hinkel, 2008)

and concept maps (Novak and Cañas, 2006). This paper focuses on the use of ontologies for semantic integration and model linking, since i. ontologies are in machine readable format, i.e. as the Web Ontology Language (McGuinness and Van Harmelen, 2004), ii. ontologies are based on first order logic, upon which a computer can reason, iii. the developed ontologies are a separate product, that are independent of the models on which they are originally based and that can be used in developing new models and iv. both modellers and non-modellers can contribute to the ontology development.

The SEAMLESS Integrated Framework (IF) is an integrated assessment tool for agricultural systems, in which agricultural and environmental policies are assessed on several dimensions of sustainability (Van Ittersum, et al., 2008). SEAMLESS-IF has a chain of linked agricultural models at its core in the tradition of Integrated Assessment and Modelling. In the development of the SEAMLESS-IF, agricultural models, data sources, scenarios and indicators had to be integrated in one conceptual schema, providing the shared definition of concepts and their relationships. Ontologies were used as a method to capture this shared definition, and these ontologies had to be build from scratch through semantic mediation.

Shared conceptual schemas for SEAMLESS-IF were developed for data (Janssen, et al., 2009a), scenarios (Janssen, et al., 2009b), models (Athanasiadis, et al., 2009, Janssen, et al., In review) and indicators (Théron, et al., 2009), which demonstrate that it is possible to develop meaningful ontologies from a scientific content point of view. To build these shared conceptual schemas different methods were used for each case (i.e. models, data sources, scenarios and indicators). This paper describes, compares and reflects on these different methods used in the semantic mediation in SEAMLESS-IF, and aims to investigate how semantic mediation in a research project can best occur or be organized. Among others, this paper argues that the method used for semantic mediation has to be flexibly adapted to the case (i.e. models, data sources, scenarios and indicators) and the researchers involved. This paper focuses on the process of semantic mediation and not on annotating the models, data sources, scenarios or indicators with ontology constructs or tags.

The next section provides background information to SEAMLESS and the use of ontologies in SEAMLESS. The third section 'Results' describes the process of ontology mediation used for each of the cases (i.e. models, data sources, scenarios and indicators). The fourth section 'Discussion' compares the different methods for ontology development and finally, in the section 'Conclusions and recommendations' lessons learned are described relevant to other efforts of semantic mediation of agri-environmental tools, models and data sources.

2. METHODS

2.1 Ontologies and SEAMLESS-IF

In the development of SEAMLESS-IF ontologies were used to achieve semantic agreement between a group of researchers on the shared scientific content. The integration for SEAMLESS-IF covered five hierarchical systems levels (e.g. field, farm, region, country and continental) linking different types of models (i.e. simulation, optimization, partial equilibrium and upscaling) for the calculation of impacts on European agricultural systems through indicators. The integration of SEAMLESS-IF involves a large team of about hundred scientists from agronomy, economics, landscape ecology, information technology and environmental sciences.

A desired outcome of the integration effort is a common ontology, i.e. ontology, which is shared by all domains to-be-integrated and serves as a knowledge-level specification of the joint conceptualization of SEAMLESS-IF. Each scientist can refer to and should adhere to the semantics of the concepts in the common ontology, including restrictions on the concepts and relationships between the concepts. As an example of an ontology, the concept *crop* has different meanings across models (Fig. 1). In a crop growth simulation model (Van Ittersum and Donatelli, 2003) a *crop* refers typically to a variety that is grown somewhere, e.g. potatoes grown in the North of Europe are quite different from potatoes grown in the South of Europe. In contrast, in a market model (Britz, et al., 2007) a *crop*

typically refers to a group of similar crops throughout a large zone, for example wheat refers both durum wheat and soft wheat and then mostly to the grain as produced by growing wheat.

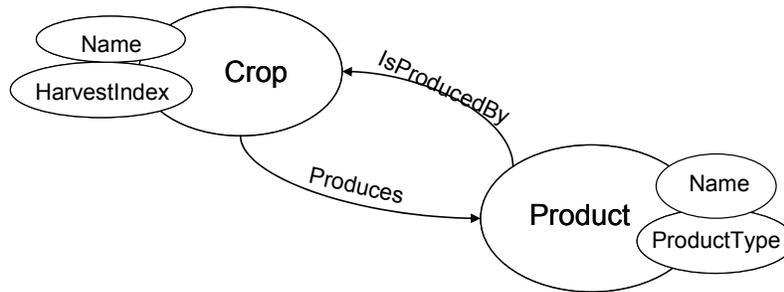


Figure 1. a part of an ontology showing two concepts (in ovals; *Crop* and *Product*), their relationships (uni-directional arrows; relationship as *Crop Produces Product* and relationship as *Product is-produced-by Crop*) and their data-properties (*Name* for Concept *Crop* and *Product*, *Producttype* only for Concept *Product* and *HarvestIndex* only for concept *Crop*).

A set of granular ontologies was developed for the SEAMLESS project according to the example provided in Figure 1 (Athanasiadis, et al., 2009). These eleven granular ontologies comprise 303 concepts, 303 object properties and 1069 data type properties and are available on <http://ontologies.seamless-if.org>. The ontologies were used directly in the development of SEAMLESS-IF, as source code could be generated for the application layer in the Java programming language and for the persistence layer in the Standary Query Language (SQL) (Athanasiadis, et al., 2007). The ontologies functioned as the upper knowledge level specification of the domain and changes in the ontology directly influenced the source code (i.e. relational database and classes for model development) used for development of SEAMLESS-IF (Figure 2).

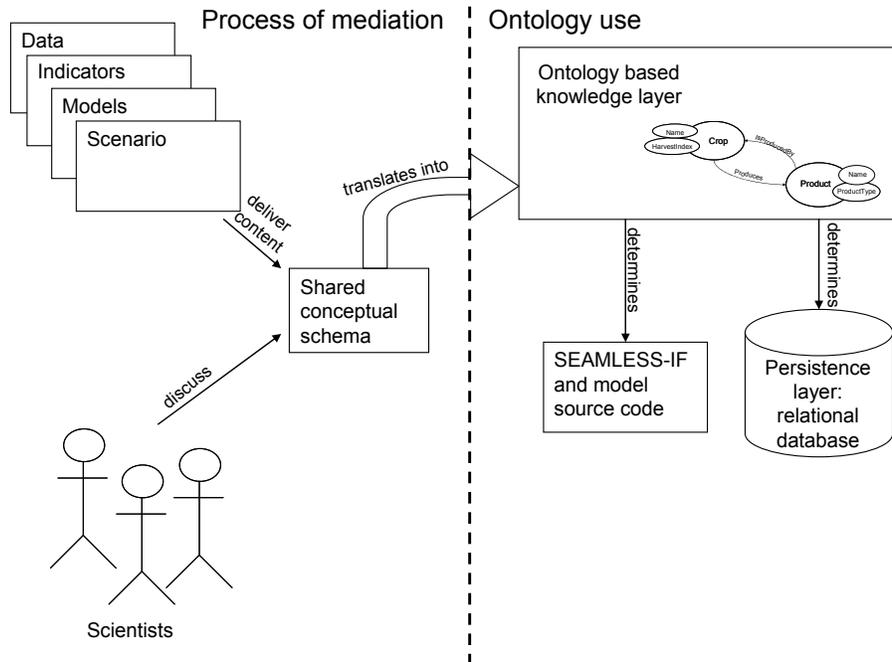


Figure 2. Process of ontology mediation (left side) and ontology use (right side).

2.2 Ontology mediation

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 and Joshi (2002) suggest several methods of ontology mediation. In the inspirational method one developer sets up the ontology as required for the application and bases this on his own view and creativity. The inductive method means examining relevant cases from the domain to extract the ontology. In the deductive method general principles are declared, from which a specific domain ontology can be deduced. The synthetic method derives a domain ontology by coupling existing independent ontologies. In this coupling of ontologies, links need to be defined and aspects of the coupled ontology might need to be developed. For developing a shared ontology, the most suitable method seems a collaborative approach. A collaborative approach is based on 'development as a joint effort reflecting experiences and viewpoints of persons who intentionally cooperate to produce it' and it thus requires a consensus-building mechanism (Holsapple and Joshi, 2002). A collaborative approach has two advantages. The first advantage is researchers from different disciplines are diverse in their contributions, which reduces the chance of blind spots and which has more chances of getting a wide acceptance (Holsapple and Joshi, 2002). As a second advantage, it can incorporate approaches other than the collaborative approach (e.g. inductive, inspirational, deductive approaches) as required for development of parts of the ontology. A promising method, that confirms the methods as described by Holsapple and Joshi (2002) is METHONTOLOGY (Fernandez, et al., 1997), that builds on iterations and a strong specification phase and lacks the collaborative aspect important in our case.

3. RESULTS

This section explains the methods in semantic mediation used for the project as whole (3.1 Task force) and for the different cases (i.e. models, data, scenarios and indicators) relevant to the development of an integrated assessment tool.

3.1 Task force

Overall a collaborative approach was used to ontology mediation in a dedicated task force as part of the larger research project (e.g. SEAMLESS with 30 institutes and 120 scientists). For this collaborative process of ontology development, no useful off-the-shelf methodology was found, that could handle all the different cases relevant for the project. Therefore the task force was formed, which consisted of a small group of knowledge engineers (3-4 members) and of domain experts (12-15 members), which were involved upon request, from different parts of the project. The task force started with the drafting of a work plan, in which the role and the mission of the task force were formalized. This work plan was crucial to build a shared vision between participants and clearly define the desired products of the task force. The desired products were one or a set of ontologies to act as a joint conceptual schema for data, models, indicators and scenarios.

The membership of the task force was on a voluntary basis, which had the advantage that participants were motivated to participate and that more participants could easily join the task force, after initial promising actions of the task force generated interest. Disadvantages were that there was considerable time needed to involve participants and that there was a high turn-over rate of participants as many participants left the group due to other priorities on top of participants changing jobs. About half of the domain experts and one of the knowledge engineers left the group. The domain experts were replaced within the task force. The task force worked in an iterative process of developing prototypes and a final version. At the end of each iteration, a version of the ontology was delivered. The setting up of each prototype started with planning of the activities for that prototype and the interactions required.

3.2 Models

For the ontologies of the models, adaptations were made to the collaborative approach according to the models involved. For one model, i.e. the bio-economic farm model, which followed an architecture close to the SEAMLESS-IF architecture, the ontology was largely

built through an inductive approach of examining model inputs, outputs and equations. This inductive approach was supplemented with extensive discussions of the resulting ontologies between one domain scientist, one computer scientist and one knowledge engineer. In this case, one of the modellers developing the model had affinity with the development of the ontology, and could be engaged in an intensive process with the knowledge engineers. For the other models and for the model chain, workshops were organised with relevant domain members to develop parts of the ontology. These models were relatively more distant from the SEAMLESS-IF architecture, and followed their own architecture. Still the modellers could be engaged in the process of ontology mediation through workshops. A useful exercise during one of these workshops was to compare the different ways of capturing relationships between concepts and concepts themselves in different modelling paradigms (i.e. mathematical programming, object oriented programming, procedural programming) used for the models and framework. This exercise gave participants an example of the use of an ontology and the complexity involved in making an ontology across modelling paradigms.

3.3 Data

For the development of the integrated database, the collaborative approach was arranged in long iterations resulting in prototypes of the database. The collaborative approach was carried out mainly by the three knowledge engineers, as one of them could act as a domain expert for the domain. This participant consulted a small group of domain scientists when required. An important step in validating the ontology in each iteration was to insert the data in the database schema derived from the ontology, as then it could be checked whether the data fitted the schema of the database. The data relevant for the integrated database was comprehensive and consisted of some large data sets, which meant that developing the database schema and filling it with data was a laborious process. This process required a meticulous way of working to ensure the integrity, consistency and completeness of the data.

3.4 Indicators

Once the ontology for models and data were defined and clarified, the ontology for indicators was clarified. In SEAMLESS indicators were primarily based on model outputs, and could only be clarified when the models and their ontology were clear. This implied that there was relatively less time to develop an elaborate and advanced ontology for indicators (Table 1). The ontology for indicators was developed during the last two prototypes, with many smaller iterations occurring especially during the final prototype development. Two one-day workshops were organised, in which relevant scientists were brought together with knowledge engineers to develop the ontology-content. After the workshops the agreed ontology had to be revised, as one scientist not participating in the workshop gave his view and ideas. This incident demonstrates the importance of getting the right balance of participants in the workshop.

Table 1. A summary of some key characteristics in ontology mediation used for the different parts of models, data, indicators and scenarios. (* Lead time here refers to the time between start and finish of the activity, in which one does not work full time on this activity)

Case	Method	Lead time*	Number of researchers involved	Prototypes and iterations
Models	Collaborative, parts inductive and deductive	2.5 years	10	3 prototypes
Data	Collaborative, with parts inductive	2.5 years	3	3 prototypes with iterations
Indicators	Collaborative	6 months	7	2 prototypes with several iterations
Scenarios	Collaborative	6 months	19	1 prototype with

The background of the domain scientists working on indicators was a different from the background of domain scientists working on models and data. The domain scientists working on indicators were much less used to think in terms of schemas for variables, parameters and data. During the workshops the scientists had to get used to thinking along these lines and clearly formulating their ideas on relationships between concepts.

3.5 Scenarios

To define an ontology for scenarios a large group of scientists was involved in comparison to the other ontology mediation efforts (Table 1), because the definition of scenarios affected many parts and members of the project (i.e. scientists working on models, data, applications, participatory processes.). The ontology was developed in only one prototype and involved ten rapid iterations, in which scientists jointly edited a word-document, containing a verbal description of the ontology (Janssen, et al., 2009b). The document started with a set of instructions and only very precise formulations were allowed in the document. The role of the knowledge engineers was to ask questions and pro-actively discuss confusing statements from the participants to reach consensus and a detailed description of relevant concepts and relationships. The ten iterations ended with a final version, which was approved by the management board of the project. After this approval, the ontology content was disseminated to the scientist in the rest of the project (Janssen, et al., 2009b).

4 DISCUSSION

4.1 Dialects and metaphors

During the collaborative method of ontology mediation, different types of confusion on the meaning of concepts and terms (Wien, et al., 2009) have been experienced due to dialects and metaphors. Dialects (Wear, 1999, Bracken and Oughton, 2006) are the specialized languages used by each of the disciplines. One consequence of dialects is that the same concept is used with different meanings. Another consequence of dialects is that different concepts might be used, which have the same meaning. Metaphors are abstract notions used within a context or discipline to illuminate an argument, develop thinking in a new direction or refer to the unknown and these metaphors might become so entrenched that they seem true or real (Wear, 1999, Bracken and Oughton, 2006). An example of a metaphor is the *concept scenario*. Another example are the concepts exogenous and endogenous in models from economics versus parameter and variable in biophysical models. In the collaborative approach, one other type of confusion was experienced both with dialects and metaphors. The type of confusion concerns relationships between concepts, that are understood in a different way across or even within disciplines. An example is the complex relationships between *outlooks, policies and technology innovations*, relevant to the *scenario* case.

Dialects are easiest to solve, because a different understanding of concepts is relatively easy to identify. Metaphors require time and effort, as meaning of a concept is vague and abstract or many meanings exist due to the large number of participants involved. Clarifying metaphors entails defining new concepts and relationships, and researchers might not be willing to give up the relative freedom of the vagueness. These metaphors typically occur if researchers are not sure about something or need a container term to hide poorly defined concepts. Confusion on relationships is the most challenging to identify, because the differences in understanding only become apparent through detailed discussion or inspection of data sources when there is already agreement on the meaning of concepts.

4.2 Process organization

Common aspects in the semantic mediation of models, data, indicators and scenarios were that i. it was organized in several prototypes, which were gradually improved; ii. a

collaborative approach was used to discuss ontology content between domain experts and knowledge engineers and iii. the editing of the ontology itself was done by a knowledge engineer and not by a whole community of researchers. First, the organization in prototypes helped to organize and test the ontology during development. It helped to gradually improve the ontology content, and to review versions of the ontology before finalizing them as a product. Second, only the knowledge engineers edited the ontology in a dedicated ontology editor, which was on the one hand necessary as ontology files tend to easily get corrupted when relationships are not correctly set. On the other hand, it was useful as scientists could focus on reaching agreement and a shared understanding. With agreement and a shared understanding, editing the ontology in an ontology editor was a minor task. Third, logically knowledge engineers cannot build an ontology on their own by just discussing with domain scientists individually or studying their materials. Having domain scientist work and discuss collaboratively on the ontology content was crucial to build a shared understanding, i.e. domain scientists learned from each other by finding out what was meant with concepts and relationships across domains. The knowledge engineers had to facilitate this process and proactively engage domain scientists, when required. Flexibility to adjust the process of ontology mediation to its participants, content and project planning was crucial to achieve functional ontologies. As can be seen from Table 1, models, data, indicators and scenarios each required different ontology mediation methods (i.e. collaborative, inductive, deductive) and had different lead times and number of scientists involved. For *models and data*, a lot of quantitative information is often around, allowing easily for an inductive approach, supplemented with a collaborative approach to build agreement across models or data sets. For *scenarios and indicators*, many different and often conflicting views existed in the project on their meaning and a collaborative approach was crucial to find out what each of the views implied. Once clarified, these ontologies could stay rather small and focused, while the ontologies for the models and data increased in size and complexity with each prototype. For these ontologies, critical review was crucial to maintain an overview and deprecate overly complex or unnecessary parts. Adopting the collaborative method of ontology mediation to the integration problem might limit the possibilities of using or developing computerized tools. Also, editing or creating the ontology itself is an easy task compared to reaching agreement between multiple scientists involved. Wiki-like tools to facilitate the discussions between scientists might be much more useful than advanced ontology editors to build the ontology itself. Arguably our usage of the ontology in the software development of databases, models and the integrated assessment tool could restrain the possibilities of using automated tools to construct the ontology, as the ontology needs to have a specific set-up to be usable in development.

5 CONCLUSIONS AND RECOMMENDATIONS

From our experiences in semantically annotating an integrated assessment tool for agricultural assessments, we learned that the method can be based on general principles (i.e. prototyping, collaborative approach and separation between knowledge engineers and domain experts). To succeed in building coherent and comprehensive ontologies, these general principles need to be adapted to the case (i.e. researchers, available information, mediation question). With flexible methods to semantic mediation, ontologies have proven to be a suitable technology to define the integrated conceptual schema of an integrated assessment tool. An important advantage was that all scientists involved could contribute to the ontology development, and not just developers or modellers. A shared understanding was built in a large group of scientists, which helped them to work together in an interdisciplinary way.

REFERENCES

- Antoniou, G. and F. van Harmelen, *A semantic web primer*, The MIT Press, 238 pp., Cambridge, Massachusetts; London, England, 2004.
- Athanasiadis, I. N., A. E. Rizzoli, S. Janssen, E. Andersen and F. Villa, *Ontology for Seamless Integration of Agricultural Data and Models*, paper presented at 3rd Intl Conf on Metadata and Semantics Research (MTSR'09), 2009.

- Athanasiadis, I. N., F. Villa and A. E. Rizzoli, Enabling knowledge-based software engineering through semantic-object-relational mappings, paper presented at 3rd International Workshop on Semantic Web Enabled Software Engineering, 4th European Semantic Web Conference, Innsbruck, Austria, 2007.
- Bracken, L. J. and E. A. Oughton, 'What do you mean?' The importance of language in developing interdisciplinary research, *Transactions of the Institute of British Geographers*, 31(3), 371-382, 2006.
- Britz, W., T. Heckelei and M. Kempen, Description of the CAPRI Modeling System, Final report of the CAPRI-Dynaspat Project. , Bonn, 2007.
- Fernandez, M., A. Gomez-Perez and N. Juristo, METHONTOLOGY: From Ontological Art Towards Ontological Engineering, AAI Technical Report SS-97-06, 1997.
- Gruber, T. R., A translation approach to portable ontology specifications, *Knowledge Acquisition*, 5, 199-220, 1993.
- Hinkel, J., *Transdisciplinary Knowledge Integration - Cases from Integrated Assessment and Vulnerability Assessment*, PhD Thesis, Wageningen University, 198 pp., Wageningen, 2008
- Holsapple, C. W. and K. D. Joshi, A collaborative approach to ontology design, *Communications of the ACM*, 45(2), 42-47, 2002.
- Jakobsen, C. H. and W. J. McLaughlin, Communication in Ecosystem Management: A Case Study of Cross-Disciplinary Integration in the Assessment Phase of the Interior Columbia Basin Ecosystem Management Project, *Environmental Management*, 33(5), 591-605, 2004.
- Janssen, S., E. Andersen, I. N. Athanasiadis and M. K. Van Ittersum, A database for integrated assessment of European agricultural systems, *Environmental Science & Policy*, 12(5), 573-587, 2009a.
- Janssen, S., I. N. Athanasiadis, I. Bezlepikina, R. Knapen, H. Li, I. P. Domínguez, A. E. Rizzoli and M. K. v. Ittersum, Linking Models for Assessing Agricultural Land Use Change, *Computers & Electronics In Agriculture*, In review.
- Janssen, S., F. Ewert, H. Li, I. N. Athanasiadis, J. J. F. Wien, O. Théron, M. J. R. Knapen, I. Bezlepikina, J. Alkan-Olsson, A. E. Rizzoli, H. Belhouchette, M. Svensson and M. K. Van Ittersum, Defining assessment projects and scenarios for policy support: use of ontology in Integrated Assessment and Modelling, *Environmental Modelling and Software*, 24(12), 1491-1500, 2009b.
- McGuinness, D. and F. Van Harmelen, OWL Web Ontology Language Overview, WWW Consortium. Retrieved on 24 January 2008 from www.w3.org/TR/owl-features/.
- Novak, J. D. and A. J. Cañas, The Theory Underlying Concept Maps and How to Construct Them Technical Report IHMC CmapTools 2006-01, 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(2-3), 412-423, 2008.
- Scholten, H., *Better Modelling Practice: an ontological perspective on multidisciplinary, model based problem solving*, PhD Thesis, Wageningen University, 313 pp., Wageningen, 2008
- Théron, O., N. Turpin, S. Janssen, I. N. Athanasiadis, M. J. R. Knapen, C. Bockstaller, J. A. Olsson, F. Ewert and I. Bezlepikina, From models to indicators: Ontology as a knowledge representation system, paper presented at Proceedings of Integrated Assessment of Agriculture and Sustainable Development: Setting the Agenda for Policy and Society (AgSap), Egmond aan Zee, 10-12 March 2009, 2009.
- Van Ittersum, M. K. and M. Donatelli, Special Issue of the European Journal of Agronomy: Modelling Cropping Systems, *European Journal of Agronomy*, 18(3-4), 187-394, 2003.
- Van Ittersum, M. K., F. Ewert, T. Heckelei, J. Wery, J. Alkan Olsson, E. Andersen, I. Bezlepikina, F. Brouwer, M. Donatelli, G. Flichman, L. Olsson, A. Rizzoli, T. van der Wal, J.-E. Wien and J. Wolf, Integrated assessment of agricultural systems- a component based framework for the European Union (SEAMLESS), *Agricultural Systems*, 96, 150-165, 2008.
- 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(5), 577-587, 2009.
- Wear, D. N., Challenges to interdisciplinary discourse, *Ecosystems*, 2, 299-301, 1999.

Wien, J. J. F., A. E. Rizzoli, M. J. R. Knapen, I. N. Athanasiadis, S. J. C. Janssen, L. Ruinelli, F. Villa, M. Svensson, P. Wallman, B. Jonsson and M. K. Van Ittersum, A web-based software system for model integration in agro-environmental impact assessments, In: Brouwer, F. and M. K. Van Ittersum (Ed.) *Environmental and agricultural modelling: integrated approaches for policy impact assessment*, Springer Academic Publishing, 24 pp., 2009.