

# Ontology for Seamless Integration of Agricultural Data and Models

Ioannis N. Athanasiadis<sup>1</sup>, Andrea-Emilio Rizzoli<sup>1</sup>, Sander Janssen<sup>2</sup>,  
Erling Andersen<sup>3</sup>, and Ferdinando Villa<sup>4</sup>

<sup>1</sup> Dalle Molle Institute for Artificial Intelligence, USI-SUPSI, Lugano, Switzerland

<sup>2</sup> Alterra, Wageningen University and Research Centre,  
Wageningen, The Netherlands

<sup>3</sup> Danish Centre for Forest, Landscape and Planning, University of Copenhagen,  
Copenhagen, Denmark

<sup>4</sup> Ecoinformatics Collaboratory, University of Vermont, Burlington, VT, USA

**Abstract.** This paper presents a set of ontologies developed in order to facilitate the integration of a variety of combinatorial, simulation and optimization models related to agriculture. The developed ontologies have been exploited in the software lifecycle, by using them to specify data communication across the models, and with a relational database. The Seamless ontologies provide with definitions for crops and crop products, agricultural feasibility filters, agricultural management, and economic valuation of crop products, and agricultural and environmental policy, which are in principle the main types of data exchanged by the models. Issues related to translating data structures between model programming languages have been successfully tackled by employing annotations in the ontology.

## 1 Introduction

The study of agricultural systems requires data spanning across several domains, including ecology, crop science, agronomy, meteorology, economy, policy and demographics. Any modelling framework that aims to integrate crop biophysical models and agro-economic models, at different scales of time and space, obviously needs to offer processes and tools for the seamless and sound management of data. Accessing data is just one side of the problem, as different sources need to be homogenized, documented and properly annotated, before been made available. The other side is persistent storage of simulation results, which again requires rich meta-data to ensure transparency and provide some degree of quality control. We faced such issues in the development of the Seamless-IF framework, where a community of more than one hundred scientists were in need to achieve consensus in their data and model conceptualizations.

This paper presents a remedy to tackle the complexity of agricultural data management issues, by developing and utilizing a set of ontologies for the development of knowledge bases related to agriculture. In the following section we discuss in short the Seamless-IP project and its supporting software infrastructure, *SeamFrame*, from the perspective of data integration and annotation. Next,

we document the empirical process followed in the development of the Seamless ontologies by a community of scientists. In section 4 the main constructs of the developed ontologies are presented along with their use to facilitate the integration of a set of constituent models with a relational database. We also report how the ontology development was integrated in the software lifecycle.

## 2 Ontologies in Integrated Assessment Studies

### 2.1 The Seamless Integrated Project

The *Seamless Integrated Project (Seamless-IP)*<sup>1</sup> develops an integrated framework for assessing and comparing, *ex-ante*, alternative agricultural and environmental policy options, allowing analysis across different scales, dimensions of sustainability and for a broad range of issues and agents of change [1].

A large community of more than a hundred scientists from different disciplines was involved in Seamless-IP to study the phenomena involved, develop new (or adopt existing) computer models to quantify them, discover and organize appropriate data required for model calibration and execution, and develop a computer-based integrated framework that is capable to execute the model chain and apply it to various regions of Europe. Certainly, the goals of the project are highly complex, as it is required to bring together an array of heterogeneous models, which are developed following different paradigms (continuous-time simulation models, combinatorial models, market and farm optimization models) accessing data provided by diverse sources. Agricultural, economic, meteorological and landscape data, at different temporal and spatial scales, are fed into the models.

The wide diversity of modeling paradigms and data sources underline the need for cross-disciplinary conceptual integration, by facing the challenge of scientific integration, while providing with practical solutions that can be applied in the software development process. The approach adopted in Seamless-IP was to employ Semantic Web techniques for specifying the domain of agriculture. Specifically, this was achieved by developing a set of domain ontologies in order to:

- build a shared view on the systems modeled, through identifying and resolving ambiguities in terms and data structures;
- facilitate model integration in a sound way, by overcoming scaling problems that are typically remain hidden in low levels (i.e. at the coding phase);
- contribute with added value to the model development, by targeting reusability, interoperability and extensibility of model components.

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 [2].

<sup>1</sup> The Seamless-IP project website is: <http://www.seamless-ip.org>

## 2.2 A Platform for Agro-Environmental Impact Assessments: SeamFrame

*SeamFrame* is the software platform used to develop the Seamless integrated modelling framework. Seamframe aims to facilitate model integration through scientific workflows, transparent data access and storage, and end-user interaction. Seamframe's architecture and components are detailed in [3,4,5]. On the basis of SeamFrame, Seamless-IF has been built, following a layered, client-server architecture. The end user interacts with the server by means of a Graphical User Interface (GUI) that is executed as a web client. The server-client architecture of Seamless-IF allows for future applications to be developed and linked to the existing server, in order to cater for specific needs of different user groups.

The Seamless-IF GUI is structured from an end-user perspective in three phases. First comes the pre-modelling phase that involves the interaction with policy experts. At this stage, a Seamless assessment project and narrative experiments are defined and related indicators are selected. Then comes the modeling phase, that consists of the selection of model chains through scales, the detailed specification of experiments, and the execution of model chains. The SeamFrame server provides most of these functionalities, while the SeamFrame client supports remote invocation of model chain execution and retrieving of model results. The last step involves the post-modelling activities, which are supported through the visualization of model results and indicators.

## 3 Ontology Development for Scientific Workflows

### 3.1 Model Chains and Scales – Interoperability Issues

In Seamless there are several models to be integrated, which are of different types, follow different modeling paradigms, operate at different scales, and are implemented using different programming languages. More specifically, in Seamless we find [5]:

- combinatorial models, as those required for the generation of agricultural management alternatives;
- biophysical models for crop growth simulation;
- economical models dealing both with farmer income optimization and agricultural product market equilibrium;
- decision making models, including social, economic and environmental indicators;
- databases, providing with reference agro-economic, meteorological and landscape data, at various temporal and spatial scales.

As a result, interoperability issues played a major role in model integration, as different models were originally developed in different programming languages, platforms and operating systems. There is often a distinction between syntactic, structural and semantic interoperability [6]. *Syntactic interoperability* is the ability of two or more systems to exchange and share information by marking up

data in a similar fashion, in order to overcome technical heterogeneity (e.g. by using XML). *Structural interoperability* means that the systems share common data models to structure and exchange information. *Semantic interoperability* ensures that the communication between systems is sound, as data models are formally defined and using logical operations the peers can verify the content exchanged [7].

In Seamless, we opted for a solution that lies between the two latter options. For our developments, we adopted a common schema for all data that is exchanged by models. In this respect, model chain composition was based on binding contracts in terms of data structures. In order to enable future extensions to benefit from rich semantics, and enable semantic interoperability, the common data schema was defined in terms of ontologies. In this respect, ontologies were employed to document binding decisions that specify data types exchanged.

### 3.2 Community Ontology Development in Seamless-IP

A community process for knowledge elicitation and representation was deployed, with the goal to come up with ontologies that specify the data structures exchanged by the models. Following [8], at first researchers were asked to compile a list of concepts they considered relevant for model coupling, e.g. concepts that were shared between the models. These concepts were supplemented with examples and comments, in order to exemplify the meaning. In this first step, we captured an ad hominem response of the researchers about the models to be linked. Then, all list of concepts were merged into one full list of concepts, which served as a sort of lexicon [8]. In this full list of concepts, conflicts between concepts and unclear concepts were indicated through iterative discussions in smaller groups. In these group discussions also the relationships between concepts were discussed. After some iterative rounds of discussion the common ontology was created, which included concepts, properties of concepts and instances of concepts. As experienced by [9], the common ontology can rapidly increase in size through iterative rounds with additional specifications that might make the ontology over-comprehensive. As models were developed in parallel, the synchronisation of the development of the ontology is therefore a difficult task.

## 4 The Seamless Ontologies

### 4.1 The Seamless Model Chain

The collaboration of scientists resulted to a shared ontology, covering scales, models, indicators and dimensions relevant to the Seamless project. Instead of making one large ontology, spanning across different sub-domains of the project, we developed eleven small ontologies, each one of which refers to a distinct aspect of the project. In this respect, common concepts and relationships are shared across granular ontologies. In Figure 1, a simplified view of the scientific workflow is presented, along with the ontologies that specify model communications.

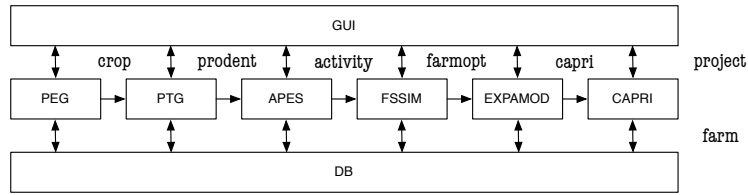


Fig. 1. Seamless model chain and ontologies involved

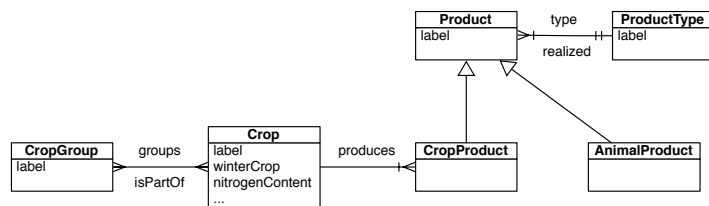


Fig. 2. Partial view of the crop ontology

First come two combinatorial models: Production Enterprise Generator (PEG) and the Production Technique Generator (PTG), which generate alternative arable activities and are coded in Java. Then follows APES<sup>2</sup>, which is a bio-physical crop growth model, and is written in C#. The workflow concludes with two optimization models (FSSIM and CAPRI<sup>3</sup>), and an extrapolation model (EXPAMOD) written in GAMS.

#### 4.2 Crops and Products Ontology

The crop ontology provides with a conceptualization of crops and crop products and their relationships, as they have been defined in Seamless. The concept of a crop is central for the project, as many components rely on it. We consider industrial crops, which are grown and harvested for producing one or more products, that are of different types. Also, crops form groups according to several criteria. Crop groups are used for abstracting crop production to higher levels in optimization models. Key concepts of the crop ontology are depicted in Fig. 2. As an example consider ‘winter soft wheat’, ‘spring soft wheat’, ‘winter durum wheat’, ‘spring durum wheat’ as crops grouped in the ‘wheat’ crop group. ‘Winter soft wheat’ produces two products: ‘winter soft wheat grain’ and ‘winter soft wheat straw’, of type ‘Grain’ and ‘Straw’ respectively.

<sup>2</sup> For more on the Agricultural Production and Externalities Simulator (APES) see: <http://www.apesimulator.it>

<sup>3</sup> CAPRI stands for Common Agricultural Policy Regionalized Impact Modeling System. See also: <http://www.capri-model.org>

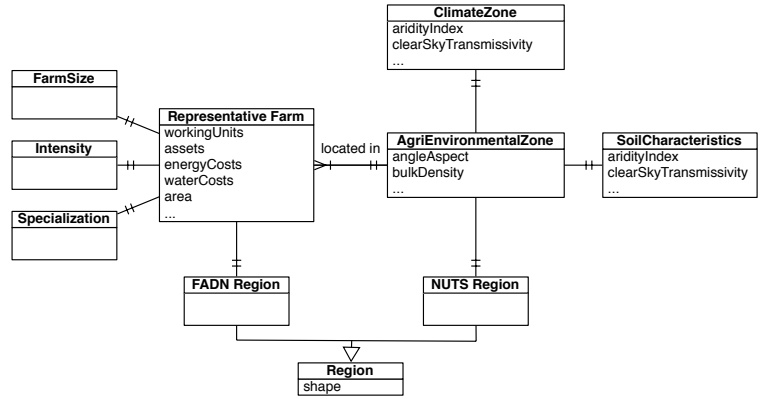


Fig. 3. Partial view of the farm ontology

### 4.3 Farm Ontology

The farm ontology provides with concepts about farms, geographical regions, in which farms are located, and soil and climate information. A representative farm corresponds to an average farm associated with a FADN region (and agro-economical data)<sup>4</sup> and with a unique specialization, intensity and size classification. As an example consider a large farm specialized in arable crops, of medium intensity production, in the region of Flevoland, The Netherlands. Each representative farm is located in one agri-environmental zone, through which is associated to climate, soil and administrative information. An agri-environmental zone identifies a location within Europe as a unique combination of a soil type, an environmental zone and a NUTS region<sup>5</sup>. Key concepts are depicted in Figure 3. Farm ontology and its use for developing an common database for European agricultural data is further discussed in [10].

### 4.4 Production Enterprise

The production enterprise ontology specifies concepts related to crop rotations and cultivation choices available to farmers. A rotation is a crop succession scheme. Typically, artificially generated crop rotation schemes start from all possible rearrangements of the available crops that are subsequently filtered with respect to cyclic equivalence and crop succession suitability requirements. Crop-specific cultivation restrictions are defined in a generic concept **CropRequirements**, which define conditions that have to be met for a crop to be able to grow. Crop requirements

<sup>4</sup> FADN stands for Farm Accountancy Data Network, which is an instrument of the European Commission.

<sup>5</sup> NUTS stands for Nomenclature of Territorial Units for Statistics. It provides with geographical references of the administrative divisions of European countries for statistical purposes.

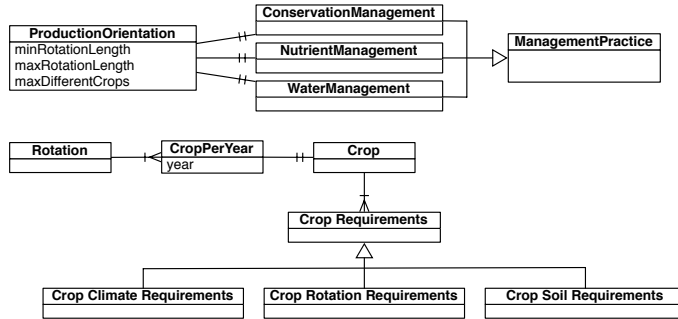


Fig. 4. Key concepts of the production enterprise ontology

can be specified with respect to climate, soil or rotation. Other types of restrictions are defined with respect to production orientations, which quantify stakeholders’ preferences on agricultural production. Examples of production orientations include integrated, organic, or conventional farming, which are quantified in terms of management practices available in the farm and restrictions related to the rotation size and structure. Key concepts of the production enterprise ontology are depicted in Figure 4.

### 4.5 Agricultural Activities

Another ontology is devoted to agricultural activities. An agricultural activity is a coherent set of crops (or animals or grass or trees) with operations and associated inputs, which (when applied on a farm) result in the delivery of a marketable

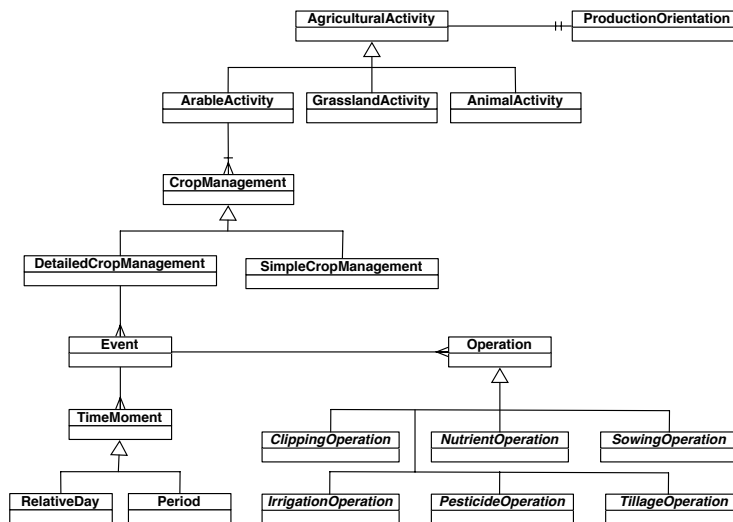
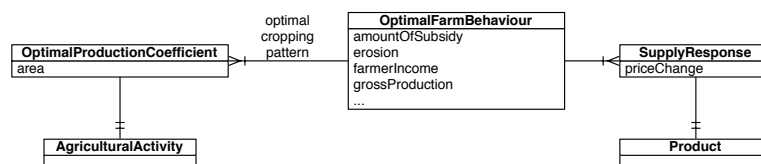


Fig. 5. Key concepts of the activity ontology



**Fig. 6.** Key concepts of the farm optimization ontology

product. An arable activity is a type of agricultural activity, and refers to a set of crop management entities. A crop management concept is further specified as a simple crop management entity that realizes a compact, though simplified view of crop management alternatives. A detailed crop management concept provides with a more detailed definition, as management operations and their timing are defined as events are defined as events. Operations can be of different types (e.g. irrigation, clipping, sowing, tillage, etc). Each operation is associated with the necessary inputs, which include among others fertilizers, water, and seeds, and implements which include sowing implements, irrigation methods and tools and fertilizer application methods.

Agricultural activities specifications span across three ontology files in Seamless: `activity.owl` provides with the generic framework, while `agrirule.owl` and `livestock.owl` specify respectively arable and livestock activities in more detail. Key concepts of the activity ontology are depicted in Figure 5.

#### 4.6 Economic Valuation of Agricultural Activities and Optimal Farmer Behaviour

In the Seamless scientific workflow, agricultural activities are used as inputs to a biophysical simulation model, namely APES, that results to yields and environmental effects, which are associated to agricultural activities, constructing the production coefficient concept. The production coefficients are in turn the input of a farm optimization model that takes under account economic, environmental and policy constraints for allocating the optimal farm area to each activity. The Optimal Farm Behavior concept aggregates optimal farm behaviors and provides with an economic and environmental valuation of the optimal production pattern. It also links to a set of supply-response values that quantify changes in production levels due to price changes. Supply-response values are associated with products of the crops and products ontology.

#### 4.7 Agricultural Policy Assessment

Agricultural policy assessment has been modeled in Seamless using the CAPRI model. CAPRI is a spatial economic model that makes use of non linear mathematical programming tools to maximize regional agricultural income with explicit consideration of the Common Agricultural Policy instruments of support



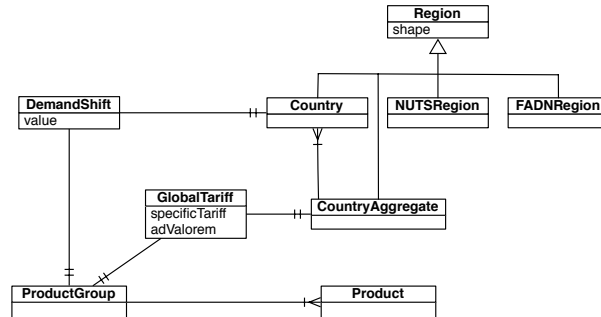


Fig. 7. Key concepts of the CAPRI ontology

in an open economy where price interactions with other regions of the world are taken into account. The corresponding ontology specifies a subset of the CAPRI parameters that are relevant for SEAMLESS assessments, including tariffs, energy prices, price elasticities, basic premiums, inflation rate and exchange rates. Each one of those has a geographical reference to a country or a country-aggregate. As an example in Figure 7 ‘demand shift’ and ‘global tariff’ are illustrated along their geographical references. Note that CAPRI follows a coarser definition of crops and products therefore ‘product groups’ have been introduced as an aggregation of ‘products’.

#### 4.8 Assessment Project Ontology

An integrated assessment project refers to the process of assessing policy or technological innovations impact on the sustainability of agricultural systems as was adopted in Seamless (see discussion in [11]). The project concept encompasses several textual information fields documenting the process, and most importantly, it is associated with the problem at hand. The scientific problem defines its spatial and temporal scale, a set of associated models that can be used for its solution, a set of experiments to be evaluated and a set of indicators that are appropriate to measure the phenomena involved. Temporal and spatial scale define problem boundaries in terms of extent and resolution in time and space. An experiment is one of the alternative configurations of the scientific workflow to be evaluated, and it is composed of two configurations: one for biophysical simulation and one for policy assessment. The biophysical configuration is composed of a single context and a single outlook. The context specifies the boundaries of the biophysical simulation in terms of products of interest and agricultural management options available, including production orientations. The outlook is defines foreseen changes to the system that are not modeled endogenously. This includes climate, economic or societal trends (i.e. atmospheric  $CO_2$  concentration, energy prices). The policy assessment concept gathers together the attributes required for the ex-ante the impacts of a policy on agricultural

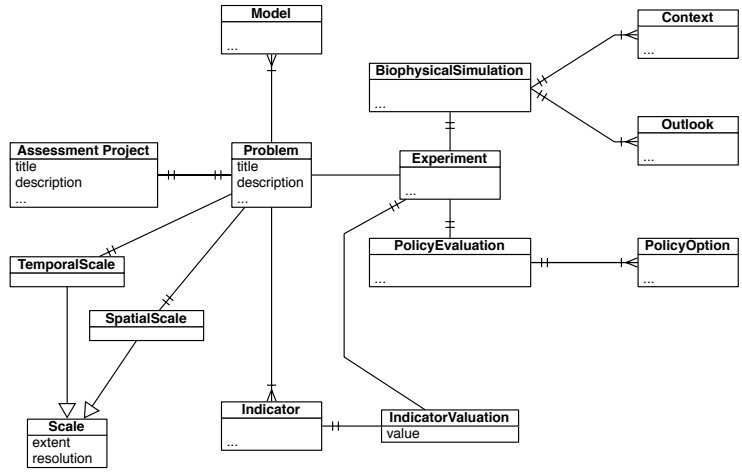


Fig. 8. Partial view of the project ontology

sustainability. It consists of policy parameters within a given timeframe, that include quotas, tariffs, set-aside regulations, subsidies and premiums.

## 5 Contribution of Ontology Development in Software Engineering Tasks

### 5.1 Cross-Programming Language Transliterations

One of the technical challenges that we addressed with the adoption of ontologies was to enable data sharing across different simulation and optimization models, each one of which is implemented in a different programming language. This diversity in programming paradigms is typical in environmental applications for various reasons, as prior developments and legacy, expertise of the developers, availability of supporting libraries and tools, and most importantly performance issues. The ‘right tools’ are needed to be used for solving each particular problem, thus we ended up having pleiad of different programming paradigms. In the Seamless project, the scientific workflow included models implemented in three programming languages: GAMS, C# and Java, plus the data that were stored in a PostgreSQL relational database. Through the community process followed, we achieved a consensus on the common data structures that specify the data exchanged by the model components of the workflow. The agreement was specified as a set of ontologies described above (in section 4). The challenge that remained open was to develop wrappers for the workflow components to enable data exchange across models, ensure the bi-directional communication with the database, and facilitate the execution process of the scientific workflow. This process is described in detail in [5].

In the Seamless ontologies, we annotated shared concepts with their names in different programming languages. For example, the Crop concept is annotated with the corresponding variable name (`Crop`) in C# for APES model and set name (`C`) in GAMS used by FSSIM, as:

```
<owl:Class rdf:ID="Crop">
  <rdfs:label xml:lang="en">Crop</rdfs:label>
  <rdfs:label xml:lang="gms">C</rdfs:label>
  <rdfs:label xml:lang="aps">Crop</rdfs:label>
</owl:Class>
```

These annotations have been further exploited by tools generating code for accessing data in different programming languages, as discussed in [12].

## 5.2 Persistent Storage of Simulation Results

Data schema definitions in Seamless ontologies have been exploited for the persistent storage of model results. Using the SeRIDA framework [13] we have transliterated OWL ontologies into relational schemas and corresponding data access classes. A software layer for transparent reading and writing data in the database was derived from the ontology definitions, and upon which the enterprise software was built. The technical solution is based on JavaBeans for structuring data and Hibernate for object-relational mapping, which was favored for the development of model wrappers and the GUI.

## 5.3 Documentation of Source Code and Database Schemas

As data access classes and database schema was generated from the ontology specifications, the comments of the ontology have been included into both JavaBeans and Hibernate mappings. In this respect, the ontology served as a single entry point not only for specifying data structures, but also for their documentation, which is propagated in the object oriented and the relational design.

## 6 Discussion

In this paper we presented a set of ontologies related to agricultural models of different types and discussed their use for model integration and environmental software development. Ontologies proved a powerful medium for specifying the data structures involved in model integration. Ontologies were used as an abstract framework for conceptual modeling of data exchanged by models, and assisted in software and database development. As conceptual models with ontologies are more rich compared to object oriented and relational models, we used ontologies to generate code across different programming languages (namely C#, GAMS and Java) and a common underlying relational schema for data storage.

**Acknowledgement.** We thank the scientists involved in the Seamless-IP project, who contributed to the development of the shared ontology. This work has been carried out as part of the Seamless Integrated Project, EU 6th Framework Programme, Contract No. 010036-2.

## References

1. van Ittersum, M.K., Ewert, F., Heckelei, T., Wery, J., Olsson, J.A., Andersen, E., Bezlepkina, I., Brouwer, F., Donatelli, M., Flichman, G., Olsson, L., Rizzoli, A.E., van der Wal, T., Wien, J.E., Wolf, J.: Integrated assessment of agricultural systems - A component-based framework for the European Union (SEAMLESS). *Agricultural Systems* 96(1-3), 150–165 (2008)
2. Cash, D., Clark, W., Alcock, F., Dickson, N., Eckley, N., Guston, D., Jäger, J., Mitchell, R.: Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences* 100(14), 8086–8091 (2003)
3. Rizzoli, A.E., Donatelli, M., Athanasiadis, I.N., Villa, F., Huber, D.: Semantic links in integrated modelling frameworks. *Mathematics and Computers in Simulation* 78(2-3), 412–423 (2008)
4. Wien, J.J.F., Rizzoli, A.E., Knapen, M., Athanasiadis, I.N., Janssen, S.J.C., Ruinelli, L., Villa, F., Svensson, M., Wallman, P., Jonsson, B.: A web-based software system for model integration in agro-environmental impact assessments. In: *Environmental and agricultural modelling: integrated approaches for policy impact assessment*. Springer, Heidelberg (2009)
5. Athanasiadis, I.N., Janssen, S.: Semantic mediation for environmental model components integration. *Information Technologies in Environmental Engineering* 1, 3–11 (2008)
6. Ouksel, A.M., Sheth, A.: Semantic interoperability in global information systems. *SIGMOD Rec.* 28(1), 5–12 (1999)
7. Heiler, S.: Semantic interoperability. *ACM Comput. Surv.* 27(2), 271–273 (1995)
8. Musen, M.A.: Dimensions of knowledge sharing and reuse. *Comput. Biomed. Res.* 25(5), 435–467 (1992)
9. Holsapple, C.W., Joshi, K.D.: A collaborative approach to ontology design. *Communications of ACM* 45(2), 42–47 (2002)
10. Janssen, S., Andersen, E., Athanasiadis, I.N., van Ittersum, M.K.: A database for integrated assessment of European agricultural systems. *Environmental Science and Policy* (2009); article in press
11. Janssen, S., Wien, J., Li, H., Athanasiadis, I.N., Ewert, F., Knapen, M., Huber, D., Théron, O., Rizzoli, A., Belhouchette, H., Svensson, M., van Ittersum, M.: Defining projects and scenarios for integrated assessment modelling using ontology. In: *MODSIM 2007 Int'l Congress on Modelling and Simulation*, pp. 2055–2061 (2007)
12. Athanasiadis, I.N., Rizzoli, A.E., Donatelli, M., Carlini, L.: Enriching environmental software model interfaces through ontology-based tools. *Int. J. of Applied Systemic Studies* (accepted for publication)
13. Athanasiadis, I.N., Villa, F., Rizzoli, A.E.: Ontologies, JavaBeans and Relational Databases for enabling semantic programming. In: *Proc. of the 31st IEEE Annual International Computer Software and Applications Conference (COMPSAC)*, Beijing, China, vol. 2, pp. 341–346. IEEE, Los Alamitos (2007)