'Translating Disciplinary Knowledge: Model Coupling through Ontologies'

Authors: Sander Janssen[1], Irina Bezlepkina[1], Ignacio Pérez Domínguez[2], and Ioannis N. Athanasiadis[3]

[1] Business Economics Group, Wageningen University, Wageningen

[2] Institute for Food and Resource Economics, University of Bonn, Nussallee 21, Bonn

[3] Dalle Molle Institute for Artificial Intelligence (IDSIA), USI/SUPSI, Galleria 2, CH-6928 Manno, Lugano, Switzerland

Contact information:

Sander Janssen

Business Economics Group

Hollandseweg 1

6706 KN Wageningen

Tel: 0317 484057

E: sander.janssen@wur.nl

Start of PhD: March 2005

This paper is based on work in progress

Keywords: Knowledge, economic models, coupling

Abstract

In a multi-disciplinary environment a common understanding of concepts and their relationships is needed for successful cooperation between disciplines. To achieve a common understanding between models – that is a model provides inputs to other models in a coherent way – first the modellers should understand and translate the knowledge that they let their models to exchange. The aim of this paper is to illustrate the potential usefulness of knowledge bases and ontologies in making knowledge explicit and re-usable between different models, exchanging data with spatio-temporal, biophysical and economic dimensions. We will present a case study based on the SEAMLESS project, which applies ontologies to a set of economic models, based on different methodologies, e.g. empirical econometric estimation models versus a mechanistic optimization model operating across different scales and one biophysical model, e.g. a dynamic crop growth simulation model. An ontology in computer science is considered as a specification of a conceptualization. After several iterations during our collaborative approach in which a number of scientist participated, a common ontology was developed. Within this common ontology the ontologies of the individual models can be distinguished, just as the links between these ontologies through shared concepts. We thus demonstrated how models can be linked through meaningful inputs and outputs, which are stored as concepts in an ontology. It is concluded that ontologies help to rigorously link models of different structures from different disciplines in a meaningful way, and an ontology can be beneficial in further ensuring that scientific knowledge is salient, legitimate and credible.

1 Introduction

Integrated assessment studies are continously asking for knowledge bits from various disciplines in order to evaluate policies under continuous changes in economic, environmental and social conditions. As noted by Meinke, Howden, & Nelson (2006) '*due to the multiple dimensions of policy, many integrated assessments are limited in their usefulness and often do not provide insights into the emergent properties of complex system*' and that '*the limits of our present knowledge mean that scientific knowledge could be described as islands of understanding in oceans of ignorance … an enduring problem is that islands of scientific understanding have been seen as separate entities which are not connected* (Lowe, 2002).'

Based on the above remarks we consider that in a multi-disciplinary environment a common understanding of concepts and their relationships is needed for successful cooperation between disciplines. This is the case in SEAMLESS (http://seamless-ip.org), an integrated project which aims at developing a computerized, integrated and working framework (SEAMLESS-IF) to assess and compare, ex-ante, alternative agricultural and environmental policy options, allowing analysis across different scales (from field, farm to region and EU), dimensions (economic, social, environmental and institutional) and a broad range of issues (Ewert *et al.*, 2005). In SEAMLESS-IF a large set of tools and models is available for analysing a broad range of agricultural policy questions, focusing on economic, social and environmental aspects. These models and tools are developed by scientists from different disciplines, e.g. economics, agronomy, land use planning, farming and livestock agriculture, and computer science. The big challenge of SEAMLESS-IF is to link models across scales and dimensions for providing integrated solutions.

One important scientific challenge is to connect Lowe's fragmented islands of knowledge, next to recognizing the wide range of outlooks regarding what makes knowledge usable within both science and society (Kates *et al.*, 2001). 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). In this research we propose to use ontologies in facilitating model linking in a model chain. To achieve a common understanding between models – that is a model provides inputs to other models in a coherent way – first the modellers should understand and translate the knowledge that they let their models to exchange.

Next to solving technical issue of model linking, there are many theoretical issues to consider. In economics, linking micro and market level models is not straightforward. This has been done for quite some time applied to the estimation of consumer demand (see Deaton and Muellbauer (1980), for an early review). Blundell and Stoker (2005) state that *'aggregation problems are among the most difficult problems faced in either the theoretical or empirical study of economics.'* Market model uses partially exogenous supply-price elasticities in its supply module. However, if we assume endogenously triggered technological changes at the farm level, exogenously set elasticities may no longer reflect what actually happens in reality. Finding ways of transmitting the endogenous farm level supply changes from farm level model (which are linked to technology and farm management) to market model is imperative in order to reach the desired level of consistency between the farm and market level layers.

The aim of this paper is to illustrate the potential usefulness of knowledge bases and ontologies in making knowledge explicit and re-usable between different models, exchanging data with spatio-temporal, biophysical and economic dimensions. Also, this paper investigates the use of knowledge bases for model integration. We will present a case study based on the SEAMLESS project that applies a knowledge base to a set of economic models, based on different methodologies, e.g. empirical econometric estimation models versus a mechanistic optimization model operating across different scales and one biophysical model, e.g. a dynamic crop growth simulation model.

This paper is structured as follows: Section 2 briefly reviews literature on use of ontologies in computer sciences and modelling. In Section 3 we will focus on methodological considerations of this study by firstly presenting the economic and biophysical models within the case study of SEAMLESS, followed by a joint ontology development. Then, the developed ontology for the case study will be presented, followed by a discussion on experiences we had during the use of ontologies and knowledge bases in our case study. Finally, some recommendations will be given.

2 Related work

Integrated assessment is an analytical approach to balance the different aspects (biophysical, institutional, social and economic) of integrated systems (Harris, 2002) that are highly complex. A way to address the complexity associated with integrated assessment is to structure the knowledge by means of ontologies and knowledge bases, which is the approach presented in this paper.

Knowledge Bases provide through ontologies a way to define rigorously conceptual models that can be easily shared between various disciplines. In the context of integrated modelling, ontologies could be useful for defining data structures describing model inputs and outputs (see Athanasiadis, Rizzoli, Donatelli and Carlini, 2006). The term ontology originates from philosophy, originally coined by classical philosophers Plato and Aristotle (Aristotle, 336-332 BC) in the study of types of being and their relationships (metaphysics). An ontology in computer science is considered as a specification of a conceptualization (Gruber, 1993), where a conceptualization is '*an abstract, simplified view of the world* e.g. systems under study (addition by author) *that we wish to represent for some purpose*' (Gruber, 1993). Such a formalization could be expressed in a machine readable format, i.e. as the Web Ontology Language (McGuinness & van Harmelen, 2004). An ontology consists of a finite list of concepts and the relationships between these concepts (Antoniou & van Harmelen, 2004). In integrated modelling research, scientists from various disciplines can define a common conceptual schema that their models share. A common ontology, i.e. ontology which is shared by all models to-be-integrated , serves as a knowledge-level specification of the joint conceptualization of the participating models and each model must adhere to the semantics of the concepts in the common ontology, including restrictions on the concepts and relationships between the concepts, but the internal specification of the knowledge in the model does not have to adhere to the common ontology (Gruber, 1993). This ontology captures scientists' disciplinary knowledge in a declarative fashion, that can be ultimately translated in a machine understandable format and be available for subsequent research. Finally, a knowledge base is the result of expressing the information related to a discipline in line with a given discipline ontology or in a common ontology. A knowledge base can thus be understood as a structure containing both the ontologies and the instances that populate the ontology (Villa et al., 2006).

Applications of ontologies are known in the field of medical research (for examples, see (Flanagan et al., 2005; Musen, 1992)) for lexicon or taxonomy-like descriptions of diseases or the genome, and computer science (for examples, see Antoniou & Van Harmelen (2004)) for information retrieval and search methodologies. So far, only one applications of ontologies is known in the field of economics, which is about their use for meta-data specification by Brilhante, Ferreira, Marinho, & Pereira (2006) Their use for model linking as described in this paper is advocated by Rizzoli *et al.* (2005) and Athanasiadis, Rizzoli, Donatelli, & Carlini (2006).

Ontologies help to formalize the knowledge captured in and/or between models, in order to subsequently facilitate model knowledge 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). A set of models with common ontology supports portability (Gruber, 1993; Villa et al., 2006) and working in a multi-disciplinary environment, as (i) different disciplines can more easily share each others models as the knowledge is not hidden in program code (Athanasiadis, Rizzoli, Donatelli, & Carlini, 2006; Villa et al., 2006), as (ii) singular model of the set of models can easily replaced by other models with the same functionality (Donatelli, Bellocchi, & Carlini, 2006; Szyperski, Gruntz, & Murer, 2002) and (iii) as models can easily linked to other models developed by other disciplines or third parties (Rizzoli et al., 2005; Szyperski, Gruntz, & Murer, 2002). These benefits can only be enjoyed if the scientific challenge is overcome of adopting tight, well-reasoned and shared conceptualizations among a group of modellers or one individual modeller.

The development of a common ontology by a group of researchers is a complex, challenging and time-consuming task (Farquhar, Fikes, Pratt, & Rice, 1995; Gruber, 1993; Holsapple & Joshi, 2002; Musen, 1992), that still remains a scientific challenge. Tools are available that help in ontology development (Farquhar, Fikes, Pratt, & Rice, 1995) and to store the ontology once it was developed (Knublauch, 2005). 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 is suggested to 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 blindspots 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 Methodology

3.1 Understanding the models

3.1.1 APES: a dynamic crop growth simulation model

APES is a modular simulation model estimating the biophysical processes of agricultural production systems, at point level, in response to weather and different options of agro-technical management (cf. Van Ittersum & Donatelli (2003)). The processes are simulated in APES with deterministic approaches mostly based on mechanistic representations of biophysical processes.

3.1.2 FSSIM: a bio-economic farm model

The Farm System Simulator (FSSIM) is an integrated modelling system developed to assess the economic and ecological impacts of agricultural and environmental policies and technological innovations. Based on the link of biophysical and micro-economic models, FSSIM seeks to describe the technical aspects at the farm level given specific biophysical conditions, using different sets of constraints to derive a set of feasible technological alternatives for each farm type. FSSIM is a modular system, which involves a mathematical programming model (FSSIM-MP), and an agricultural management module (FSSIM-AM). FSSIM-AM aims to describe and generate current and alternative activities and quantifies their input output coefficients (both yields and environmental effects), using agronomic and economic information coming respectively from the biophysical model APES and statistical data bases (costs, labour and machine requirement, etc). FSSIM-MP seeks, after including generated information from FSSIM-AM, to solve problems of maximising utility function under a limited number of explicit constraints representing the production opportunity set, resource constraints, and price parameters that farmers face.

FSSIM-MP is mathematical programming model that is i) static i.e. with a one period decision; ii) positive, which aims to reproduces as precisely as possible observed behavior of the farmers as opposed to normative models that respect the first-order optimality conditions and try to find the optimal solution to the problem of resource management and allocation (Flichman & Jacquet, 2003); iii) risk programming which at the moment uses a basic specification based on utility is dependent on the expected income and risk (Freund, 1956);

iv) a non-linear programming model as the objective function is non-linear and sometimes binary variables are included.

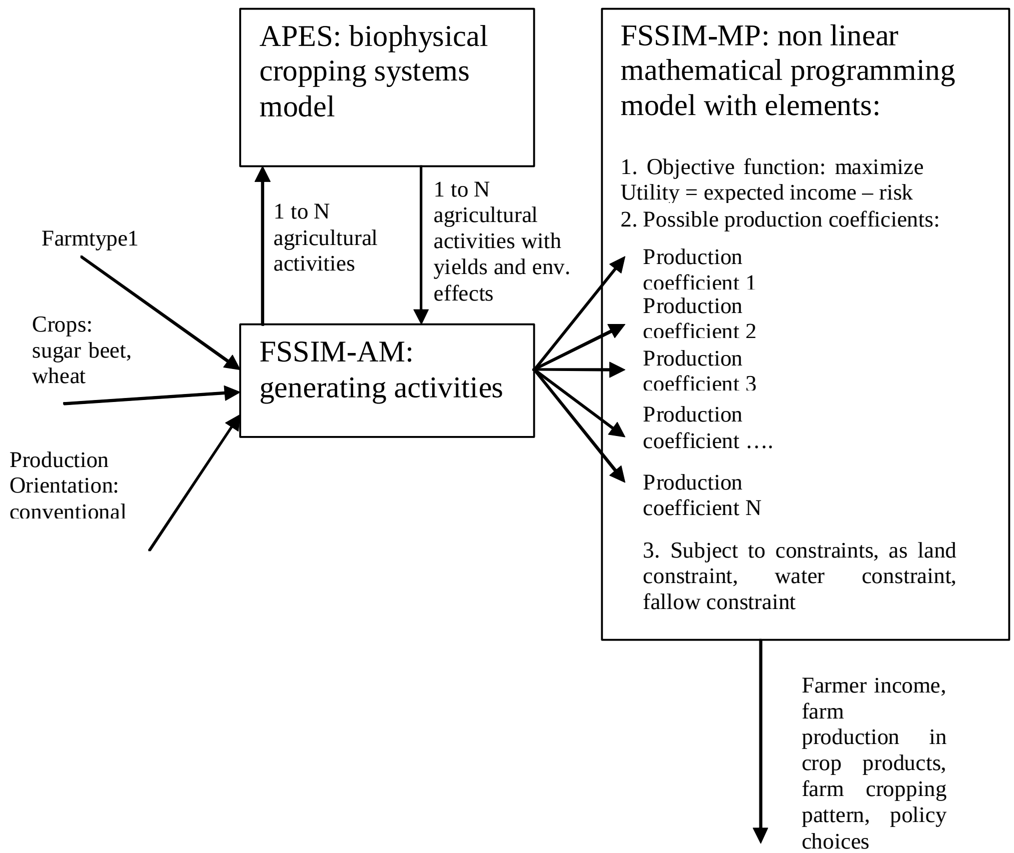

Figure 1 Overview of the FSSIM modeling system that displays the different modules and their inputs and outputs, for one farmtype

FSSIM-AM receives as inputs a farmtype, a set of crops and a production orientation, for which the farm production in terms of crop products and cropping pattern should be calculated by the FSSIM modeling framework (Fig. 1). A farmtype (Fig. 2) is an average 'farm' of group of farms based on the data found in the statistical database of FADN. As can be seen in Figure 2, a farm type is characterized by a large number of properties, for example average field size, a risk aversion coefficient, farm production by that farmtype, etc. A production orientation is a set of value driven aims and restrictions of the agricultural activity that direct the input and output levels (Van Ittersum & Rabbinge, 1997), for example 'integrated', 'organic', 'conventional' or 'low labour input.' With these inputs of farmtype, crops and production orientation, FSSIM-AM constructs agricultural activities which can be evaluated by the biophysical cropping systems model APES (Fig. 1), which consequently attaches yields and environmental effects to these agricultural activities. Finally, these agricultural activities are translated to production coefficients (Fig. 3) by FSSIM-AM, which can consequently be used by FSSIM-MP in an optimization to find the optimal cropping pattern, policy choices and farm production in terms of user defined goals. A production coefficient is characterized and defined by among others the references to a crop, a year, a crop management applied to the crop, a rotation, one or more products that have a certain yield e.g. productyield, some environmental effects, etc.

Farmtype			**Name of concept**
distanceToField	Float		Distance to field property of the farmtype
riskAversionCoefficient	Float		Risk-aversion property
irrigationWaterAvailability	Float		
hasFarmProduction	Instance*	FarmProduction	Reference to one or more farm production concepts
hasFarmHouseHoldDataPerSoil	Instance*	FHDperSoil	
hasInputsAndImplements	Instance*	InputsAndImplements	Reference to one or more inputs or implements for the farm
hasFarmHouseHoldDataperRefYear	Instance*	FHDperYear	
hasFamilyLabour	Instance	Labour	Reference to only one labour input of the farm
hasCropAgriPolicyData	Instance*	CropAgriPolicyData	
inCell	Instance	Cell	Reference to only one grid cell where the farm can be found

Figure 2 The concept of Farmtype, its properties and relationships to other concepts. Some explanation to some of the fields is included on the right of the concept. 'Instance*' in the Figure means that the each farmtype has a reference to one or more instances of another concept for example inputs and/or implements, while on the left of the 'Instance*' the name of the relationship can be found and on the right of the 'Instance*.' If it is written only 'Instance', then it means each farm type has reference to only one instance of another concept.

ProductionCoefficient		
fossilEnergyRequirement	Float	
variableCosts	Float	
nitrogenUse	Float	
hasLabourRequirement	Instance	Labour
hasFarm	Instance	Farmtype
hasRotation	Instance	Rotation
hasEnvironmentalEffects	Instance	EnvironmentalEffects
hasProductYield	Instance*	ProductYield
hasProductionOrientation	Instance	ProductionOrientation
hasIrrigationWater	Instance	IrrigationWater
hasSoil	Instance	Soil
hasCropManagement	Instance	CropManagement
hasCrop	Instance	Crop
hasYear	Instance	Year

Figure 3 The concept of Production Coefficient, its properties and its references to other concepts.

3.1.3 CAPRI: a market level model

CAPRI is a spatial economic model that makes use of non-linear mathematical programming tools to maximise regional agricultural income with explicit consideration of the CAP instruments of support in an open-economy where price interactions with other regions of the world are taken into account. It consists of a supply and market modules, which interact iteratively (see Fig. 4).

In the supply component of CAPRI, regional agricultural supply of crops and animal outputs[1] is modelled by an aggregated profit function approach under a limited number of constraints. The underlying methodology assumes a two-stage decision process. In the ***first stage***, producers determine optimal variable input coefficients per hectare or head for given yields. Nutrient requirements enter the supply models as constraints and all other variable inputs, together with their prices, define the accounting cost matrix. In the ***second stage***, the profit-maximising mix of crop and animal activities is determined simultaneously with cost-minimisation of feed and fertiliser use in the supply models. The supply module follows a 'template approach', where the optimisation models can be seen as representative farms maximising their profit by choosing the optimal composition of outputs and inputs at given prices for the final products and variable inputs. Is at this stage where the link between CAPRI and FSSIM takes place: representative FSSIM farm type models and regional CAPRI supply models.

The market module, a constrained equation system, comprises of a spatial world trade model based on the Armington assumption (Armington, 1969). Technically, the composition of demand from domestic sales and different import origins depends on price relationships resulting from bilateral trade streams.[2] This allows the model to reflect trade preferences for certain regions (e.g. Parma or Manchego cheese) that cannot be observed in a net-trade model. The market module breaks down the world into several country aggregates or trading blocks[3], each one featuring systems of ***supply***, ***human consumption***, ***feed***, and ***processing*** functions. The parameters of these functions are derived from elasticities borrowed from other studies and modelling systems and are calibrated to projected quantities and prices in the simulation year.

---

[1] Some examples of activities in CAPRI: soft wheat, durum wheat, pulses, potatoes, sugar beet, flax & hemp, tobacco, wine, nurseries, flowers, other crops, fodder maize, fallow land, obligatory set-aside, grass & grazings intensive, suckler cows, yield, dairy cows high yield, and heifers fattening high final weight.

[2] In the market model there is a connection between domestic prices in the EU and in the rest of the World (trade blocks). This means that world prices react to changes in the CAP, since the EU is an important trade partner.

[3] Trade blocks in CAPRI are: EU15, EU10, Bulgaria & Romania, Rest of Europe, USA, Canada, Mexico, MERCOSUR countries, Rest of South America, India, China, Japan, Rest of Asia, Australia & New Zealand, Mediterranean countries, Least Developed Countries, ACP countries and Rest of the World. The EU15, EU10, MERCOSUR and Mediterranean countries feature behavioural equations at single country level.

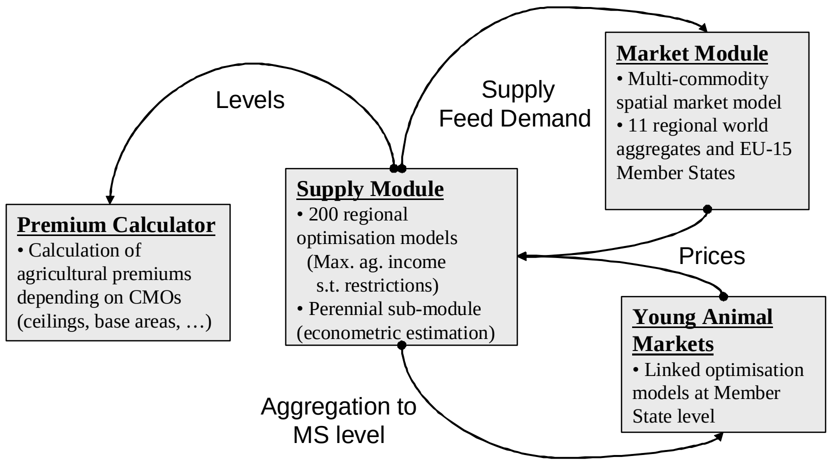

Figure 4 The different modules of CAPRI (Source: CAPRI modelling system)

These modules perform an important exchange of knowledge with each other and with other models e.g. the CAPRI market module calculates prices at Member State level and delivers them to FSSIM through the CAPRI supply module.

3.1.4 EXPAMOD: a regional upscaling model

EXPAMOD can be defined as an econometric metamodel[4] describing price-quantity responses of farms given specific farm resources and biophysical characteristics that are available EU-wide. The principle idea of such aggregation procedure is to make the regional supply modules of CAPRI behave like the aggregate of the FSSIM models of the same region – apart from additional aspects entering the market supply such as regional land or political constraints (premium ceilings). To do this, all available FSSIM models are run for a small set of exogenously fixed prices. This results in multidimensional price-quantity response surfaces, possibly one for each scenario if the policy considered affects the supply behavior at farm level. In most cases this set of product prices will vary from one scenario to another. Thus, the metamodel is estimated using simulated price-response data for farm types in regions for which farm type models exist and then applied to project supply responses of other farm types and regions.

EXPAMOD serves not only for expanding FSSIM supply to out of sample regions, but also to introduce FSSIM responses into CAPRI and, through the market module, obtain endogenous prices. EXPAMOD operates with products and prices obtained from FSSIM, farm type characteristics and regional biophysical characteristics obtained from databases. The price elasticities, i.e. EXPAMOD output, are conveyed to the CAPRI model.

---

[4] A metamodel is an approximation of the Input/Output behaviour of the underlying simulation model. “Metamodels”, are also called response surfaces, emulators, surrogates, repromodels, auxiliary models, etc.(from: (Kleijnen, 2006))

3.2 *Research design*

This research employs a collaborative approach in designing ontologies (see also Holsapple & Joshi, 2002). As a case study, the linkage of the models within the SEAMLESS project was studied. These models have been developed by different teams (having dissimilar educational backgrounds and research intentions) and are based on methodologies from different disciplines. One of the models is the agricultural sector model CAPRI that simulates supply-demand relationships in the EU25 for agricultural commodities. CAPRI derives information on price-supply relationships from two other models, e.g. bio-economic farm models (FSSIM) and a regional up-scaling model (EXPAMOD). The bio-economic farm models in turn simulate farm behaviour and use agricultural activities (i.e. crop rotations, cultivation management options) assessed through a simulation model for agricultural production and externalities (APES). The knowledge base should enable the links between models, as indicated in Figure 5. In Figure 5 the links indicate linkages between economic models, for example from a partial equilibrium model at the regional scale (CAPRI) to a bio-economic model at the farm scale (FSSIM). The dashed link indicates the linking to models from other disciplines, such as for example biophysical models at the field scale.

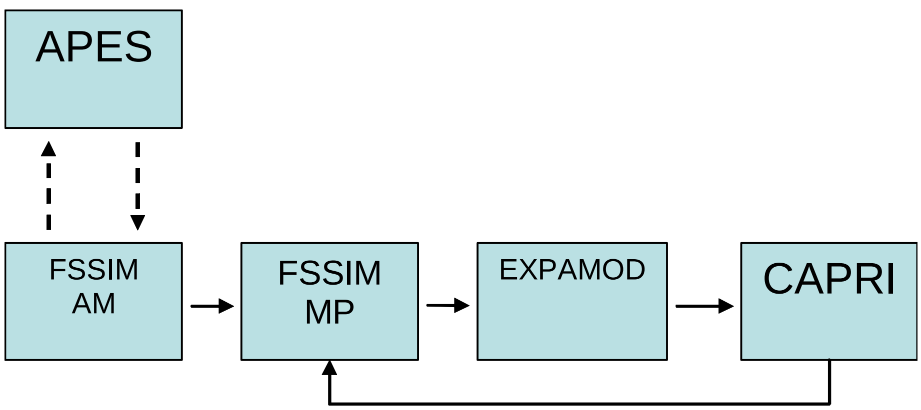

Figure 5 links between the models used in the case study.

3.3 *Common ontology mediation process*

The development of a common ontology between a group of researchers is a complex, challenging task. To facilitate the ontology development in this research, Protégé-OWL (Knublauch, 2005) was used that is specifically targeted at ontology development, that is an open source program, and that has export-facilities to export ontologies to Java source code or advanced ontology visualizations.

Following Musen (1992), at first researchers are asked to compile a list of concepts they consider relevant for coupling models, e.g. concepts that were shared between the models. These concepts should be supplemented with some examples of the concepts and additional comments. This captures some aspects of the knowledge about the models to be linked. This is further addressed in Section 4. These separate list of concepts were then merged into one full list of concepts, which serves as a sort of lexicon (Musen, 1992). 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 have been discussed. After some iterative rounds of discussion the common ontology was created, which covered concepts, properties of concepts and instances of concepts. As experienced by Holsapple & Joshi (2002), the common ontology can rapidly increase in size across iterative rounds with additional specifications which might make the ontology over-comprehensive and ontologies might present something of a moving target as models tend to develop simultaneously.

4 The developed ontology

4.1 Common Ontology

After several iterations during our collaborative approach a common ontology was developed, which is still subject to further development (Fig. 6). During the fist steps researchers produced a number of simple text-files and spreadsheets to store the ontologies as list of concepts. After some iterative rounds of discussion with the involved scientist a switch was made to a tool like Protégé (Knublauch, 2005).

The common ontology in Figure 6 is used by the four models: APES, FSSIM, EXPAMOD and CAPRI. This Figure 6 is created with Jambalaya (Storey *et al.*, 2001) and can be read as follows: the bullets and squares are associated with a concept, which has a name, for example crop in one of the Eclipses (Fig. 6). The relationships between the concepts are shown as arrows with different colours, and a relationship in this case means that a concepts contains a reference to another concept, for example the relationship between crop, which refers to CropGroup, as a group of crops. This implies that the concepts relevant to a model can be found in the common ontology. For example, the CAPRI ontology with the relevant concepts to the CAPRI model forms a clear part of the common ontology, while the link between this CAPRI ontology (see also Fig. 7-9) and the rest of the ontology is made through the relationship between the concepts named CropProductCapri[5] and the CropProduct (see Table 1 below). A part of the owl-file containing the common ontology is provided in Appendix 1. In the common ontology (Fig. 6) several concepts are central, e.g. crop, farm type and production coefficient (Fig. 6, Eclipses), as these concepts have a lot of relationships to other concepts and thus appear as a central node.

[5] The name of the concepts was constructed by merging separate words to one and starting every word with a capital letter. This is done to ensure computer readability in the OWL-file.

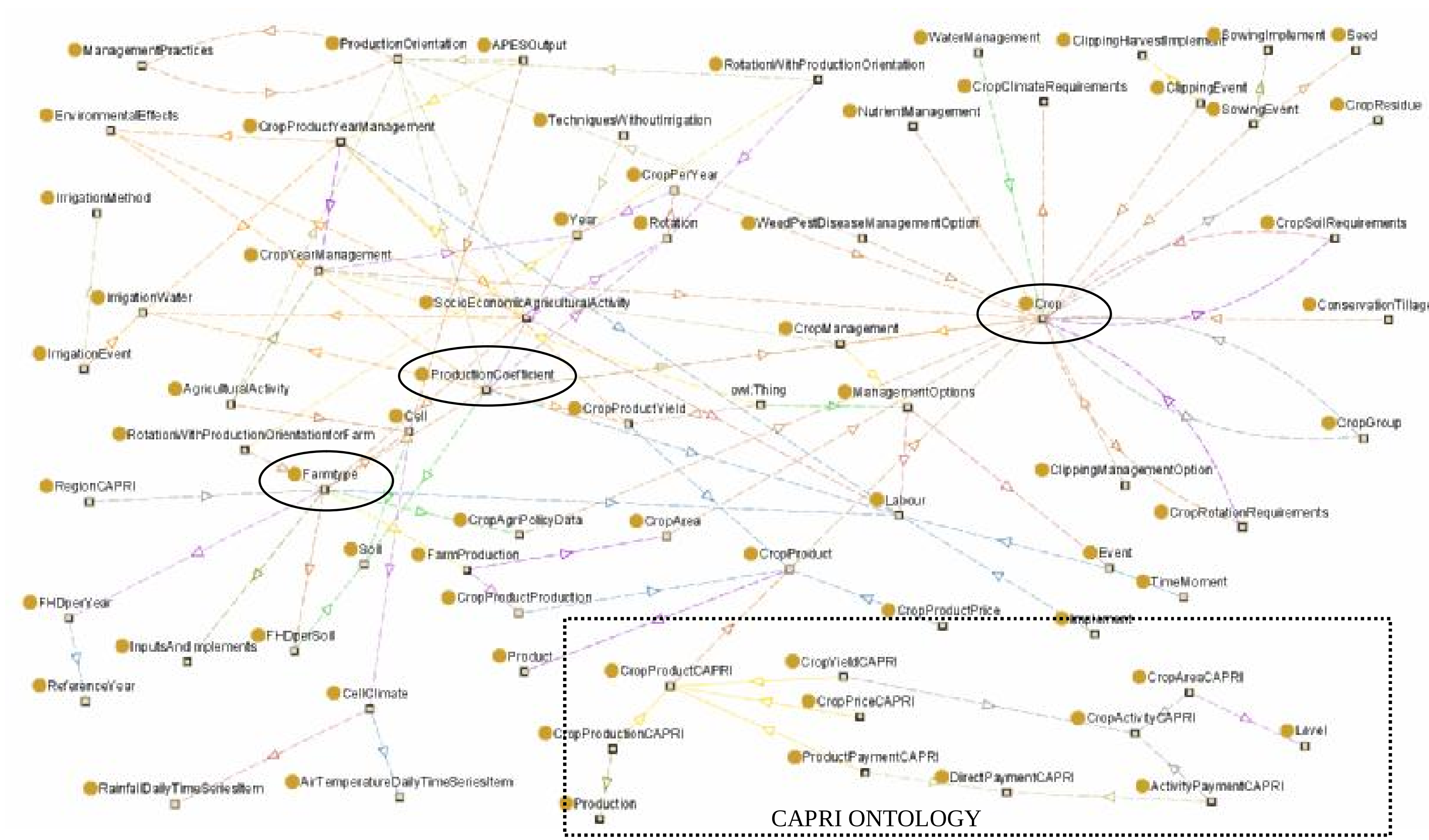

Figure 6 the common ontology as developed through the collaborative approach, with crop, farm type and production coefficient as central concepts (see eclipses) and the CAPRI ontology in the down right corner

4.2 CAPRI ontology as a single model ontology

The ontology for CAPRI can easily be found in the common ontology (see box titled CAPRI ontology in Fig. 6) and therefore will be explored as a single model ontology. For the other models within SEAMLESS, also single model ontologies are available, but these are much more complex to discern in Figure 6. A feasible conceptualisation of CAPRI within an ontology is an ambitious goal, since economic models contain very complex relationships between variables and no examples of ontologies for economic models have been found in the literature.

In the CAPRI *supply module*, the basic information needed for calculations can be structured in atomic and composite concepts. Whereas atomic concepts define the basic knowledge entity and might be related to concepts in other modelling systems, composite concepts are designed as combinations of concepts. Both contain a detailed set of characteristics and might be populated with data (i.e. instances (Fig. 9))[6]. In the following two figures, a selected number of atomic and composite concepts are shown as classes and subclasses of CAPRI ontology.

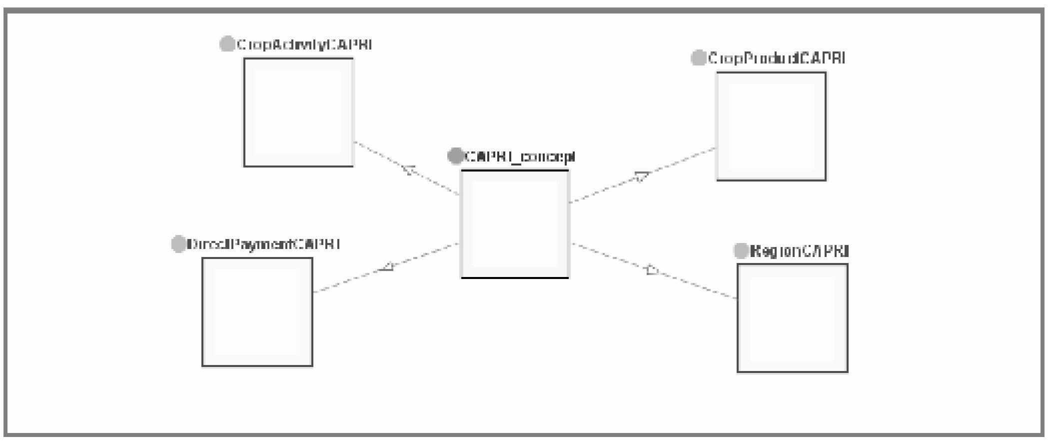

Figure 7 Structure of atomic concepts

[6] Composite concepts relate basically to two concepts and are so desgined, so that the classical data matrix behind agricultural programming models (input-output tables, market balances, …) can be stored from a conceptual perspective (e.g. a yield is a combination of an activity and a product).

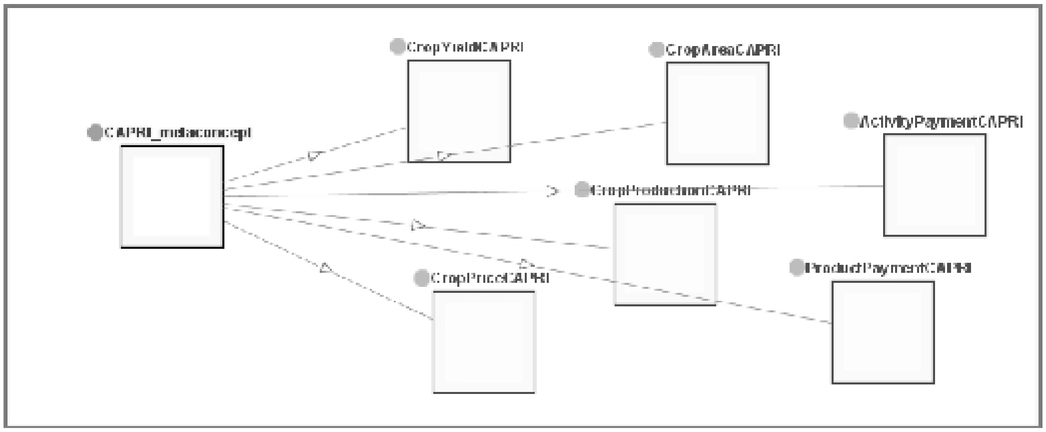

Figure 8 Composite concepts

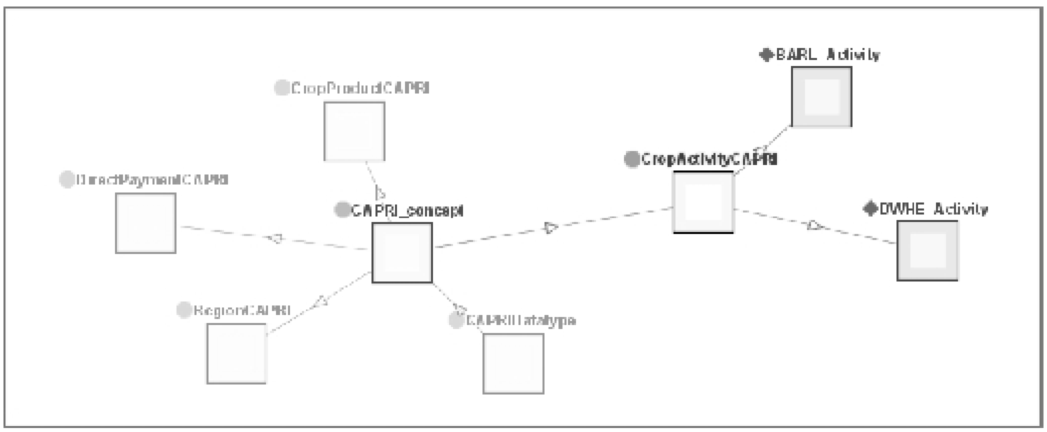

Figure 9 Some examples of crop activities in CAPRI e.g. a barley (BARL_Activity) and durum wheat activity (DWHE_Activity)

The market model mimics the approach set by the supply component of CAPRI in terms of activities and products, but introduces a new layer of information at a superior level. This means that new concepts are needed, such as scenarios and a different regional dimension. In the following table, the list of non-EU regions in CAPRI is presented. Whereas the supply model is defined at the Nuts2 level within the EU25, the market model has a different set of regions covering the world (see Fig. 10). All these regions have simplified supply and demand systems with respect to European regional supply models, but exchange information in the same way. Extending the common ontology of Fig. 6 to cover new concepts as regions and scenarios in consistent and coherent way in all the models will be part of further development.

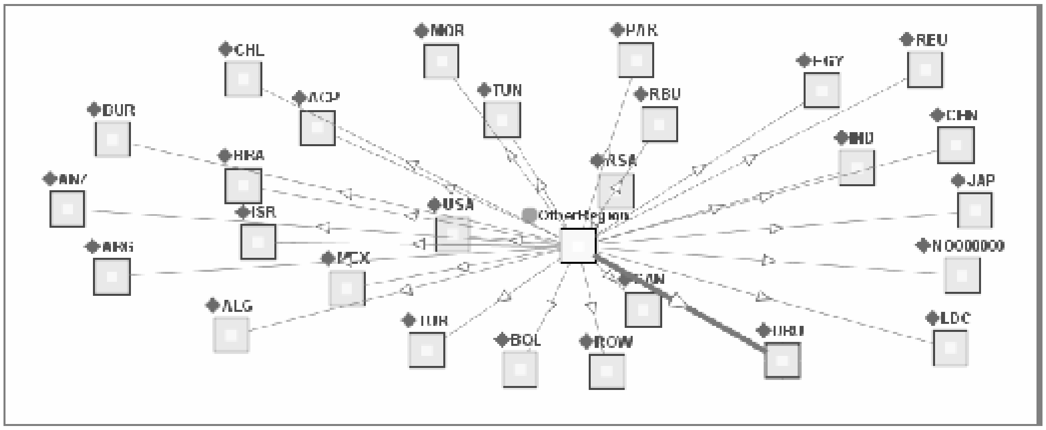

Figure 10 Regional mapping between regions in CAPRI

4.3 Linking the different economic models

EXPAMOD does not introduce a new knowledge structure in the system, since it serves just as a link between information found in other models: FSSIM e.g. production and managerial data and CAPRI e.g. prices. Nevertheless, it requires specific procedures to transfer knowledge from one scale to the other in order to be operational, i.e. it has to belong to the common ontology. In order to guarantee the consistency in the link between FSSIM and CAPRI, EXPAMOD needs to map products and prices in both models at different levels.

Thus, the linking between the economic models, e.g. bio-economic farm model FSSIM, the econometric metamodel EXPAMOD and the market model CAPRI is done through the relationship between the CropProduct and the CropProductCAPRI (see Fig. 6 and Fig. 11).While both within FSSIM and CAPRI a crop product are considered, there are differences between the crop products of CAPRI and FSSIM, which is reflected by the different names CropProduct and CropProductCAPRI. One CropProductCAPRI has as a property a reference to one or more CropProducts as used in FSSIM (Fig. 11), in other words, one crop product in CAPRI is an aggregation of several crop products from FSSIM. For example, while FSSIM makes a distinction between the crop products grain from spring soft wheat and grain from winter soft wheat, these are aggregated into the CropProductCAPRI called soft wheat (Table 1). Similarly the CropProductCAPRI Straw is an aggregation of the crop products straw from spring soft wheat, straw from spring barley, straw from oats, straw from winter soft wheat, etc in FSSIM (Table 1).

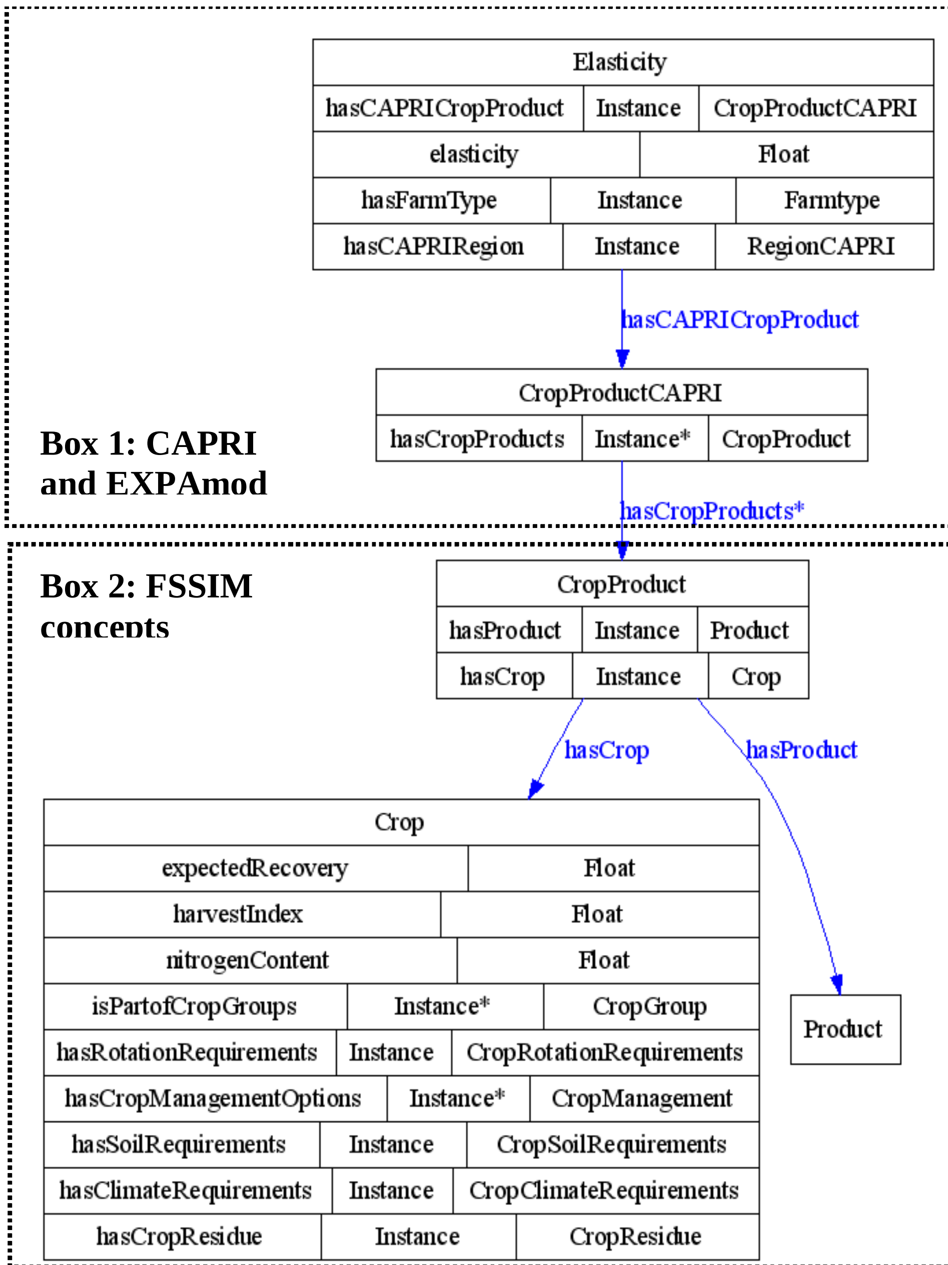

Figure 11 The concepts Crop, CropProduct, Product, CropProductCAPRI and Elasticity that are used by the models CAPRI, FSSIM and EXPAMOD. Box 1 indicates which of the concepts are part of CAPRI and EXPAMOD, while Box 2 indicates which concepts are part of FSSIM.

The implementation of this relationship between the CropProduct of FSSIM and the CropProductCAPRI as an aggregation is the responsibility of EXPAMOD. How this aggregation is implemented within EXPAMOD is internal to EXPAMOD, and does not need to be included in the common ontology, as long as EXPAMOD receives from FSSIM the CropProducts with associated data, e.g. yield levels, region and farm type as an input, and supplies to CAPRI a CropProductCAPRI with associated data, e.g. elasticity, region and farm type. It is thus the responsibility of EXPAMOD (a) to aggregate more detailed crop products from the farm scale (e.g. from winter soft wheat, spring soft wheat, winter durum wheat, spring durum wheat) to regional scale (winter wheat, spring wheat) and (b) to disaggregate the crop products at the regional scale back to the farm scale. Although each of the models FSSIM, EXPAMOD and CAPRI on their own are quite complex, the linking through concepts is straightforward, as only one relationship is considered yet.

Table 1 the aggregation of some CropProducts as used in FSSIM into CropProductCAPRI as used by CAPRI and EXPAMOD

Crop	**CropProduct**	**CropProductCAPRI**
SpringSoftWheat	SpringSoftWheatGrain	SoftWheat
SpringSoftWheat	SpringSoftWheatSeed	SoftWheat
WinterSoftWheat	WinterSoftWheatGrain	SoftWheat
WinterSoftWheat	WinterSoftWheatSeed	SoftWheat
Backwheat	BackwheatStraw	Straw
Dinkel	DinkelStraw	Straw
GrassSeed	GrassSeedStraw	Straw
MaizePopCorn	MaizePopCornStraw	Straw
Maize	MaizeStraw	Straw
Meslin	MeslinStraw	Straw
Millet	MilletStraw	Straw
Oats	OatsStrawn	Straw
OtherCereals	OtherCerealsStraw	Straw
Rice	RiceStraw	Straw
Rye	RyeStraw	Straw
Sorghum	SorghumStraw	Straw
SpringBarley	SpringBarleyStraw	Straw
SpringDurumWheat	SpringDurumWheatStraw	Straw
SpringSoftWheat	SpringSoftWheatStraw	Straw
SweetMaize	SweetMaizeStraw	Straw
Triticale	TriticaleStraw	Straw
WhiteSorghum	WhiteSorghumStraw	Straw
WinterBarley	WinterBarleyStraw	Straw
WinterSoftWheat	WinterSoftWheatStraw	Straw

4.4 Linking economic and biophysical models

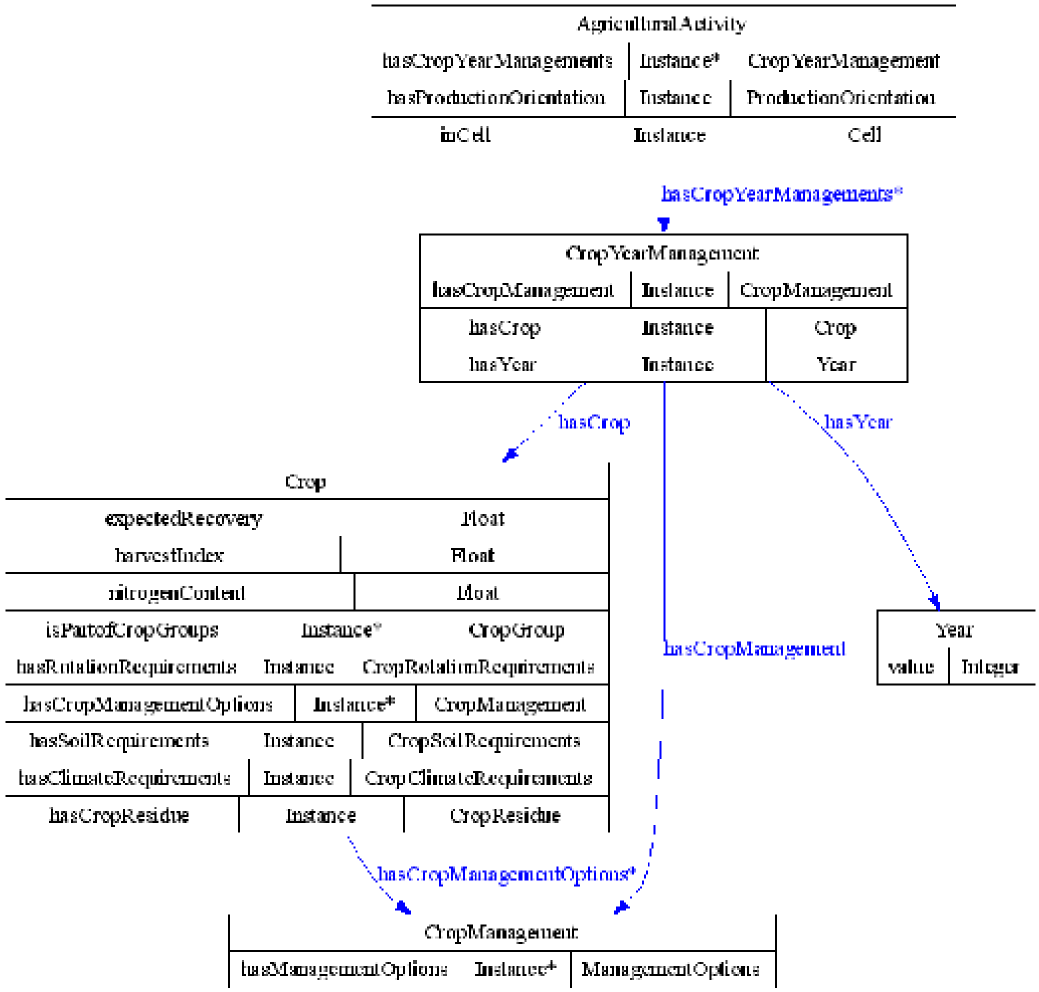

Figure 12 The concept of an Agricultural Activity and its links to other concepts, among others CropManagement

The crop growth simulation model APES receives from the bio-economic farm model FSSIM an Agricultural Activity for which yields and environmental effects should be calculated (Fig. 12). Figure 12 contains a specification of the concepts presented in the common ontology in Figure 6 by focusing on some of the concepts in Figure 6 and by defining both the relationships between concepts and the properties of the individual concepts. An agricultural activity holds information on the crop, the year of the rotation in which the crop is sown and the crop management applied to the crop (Fig. 12). The concept of Crop Management (Fig. 13) is associated to a set of events through the concept of management options. An event is one operation that takes place during the growing season of the crop, for example sowing, fertilization, irrigation, harvesting, field inspection, etc. The crop management of a crop in an agricultural activity exists out of several events, for example two nutrient events, one irrigation event, a sowing event and a harvesting event. Each of these events has properties that are required by APES, for example a mean tillage depth, the irrigation implement used for irrigation, or the amount of organic nitrogen. The properties of the different events are given in Figure 14.

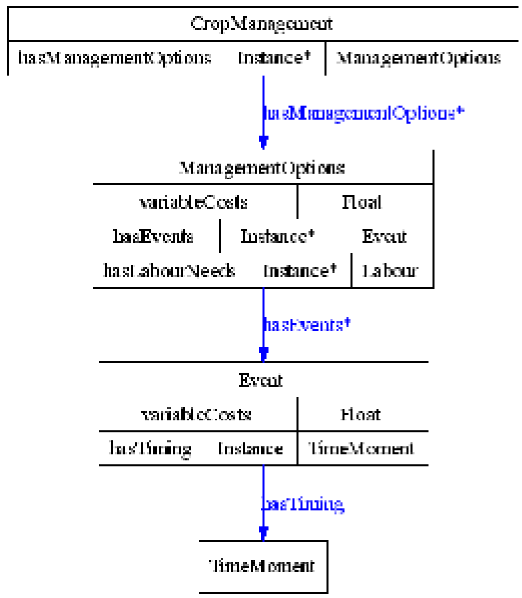

Figure 13 The concept CropManagement and its associated concepts ManagementOptions and Event

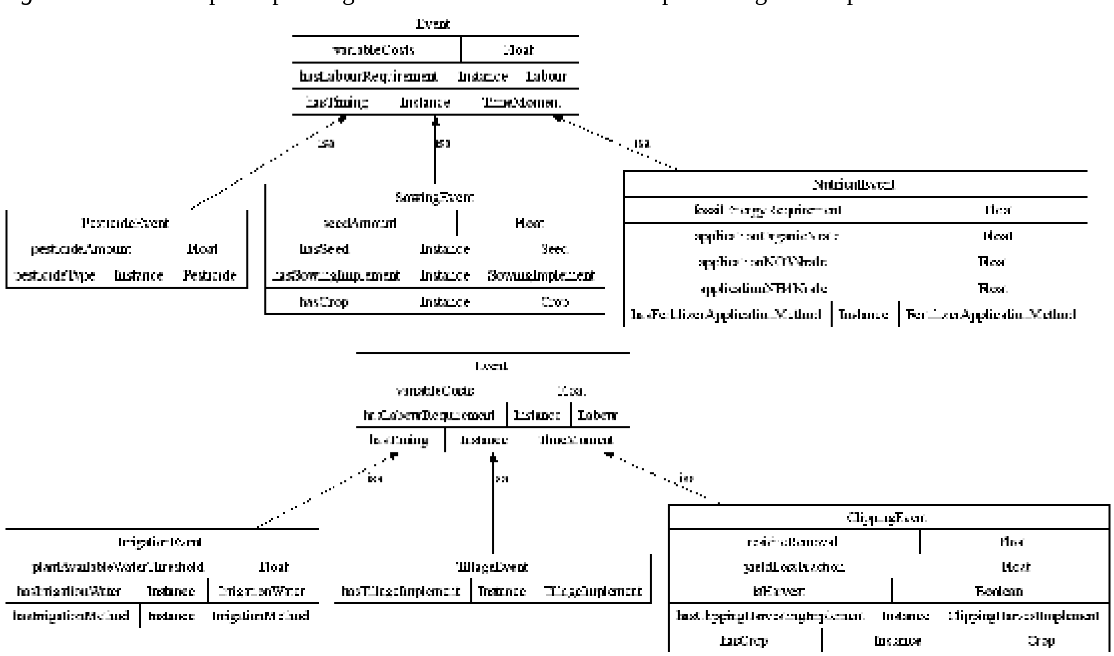

Figure 14 the different types of events and their properties.

APES uses the information it gets on agricultural activities with associated crops, rotations, and events to calculate for each of the agricultural activities an APESOutput, which holds the same information as the agricultural activity, but supplemented with yield and environmental effects. Instead of reference to a CropYearManagement (Fig. 12) as the agricultural activity has, the APESOutput (Fig. 15) has a reference to a CropProductYearManagement, which holds a reference to yields of crop products through the CropProductYield concept and which holds a reference to environmental effects through the EnvironmentalEffects concept as have both been calculated by APES. FSSIM receives this information on crop products with associated yields and uses this further in interaction also with EXPAMOD and CAPRI.

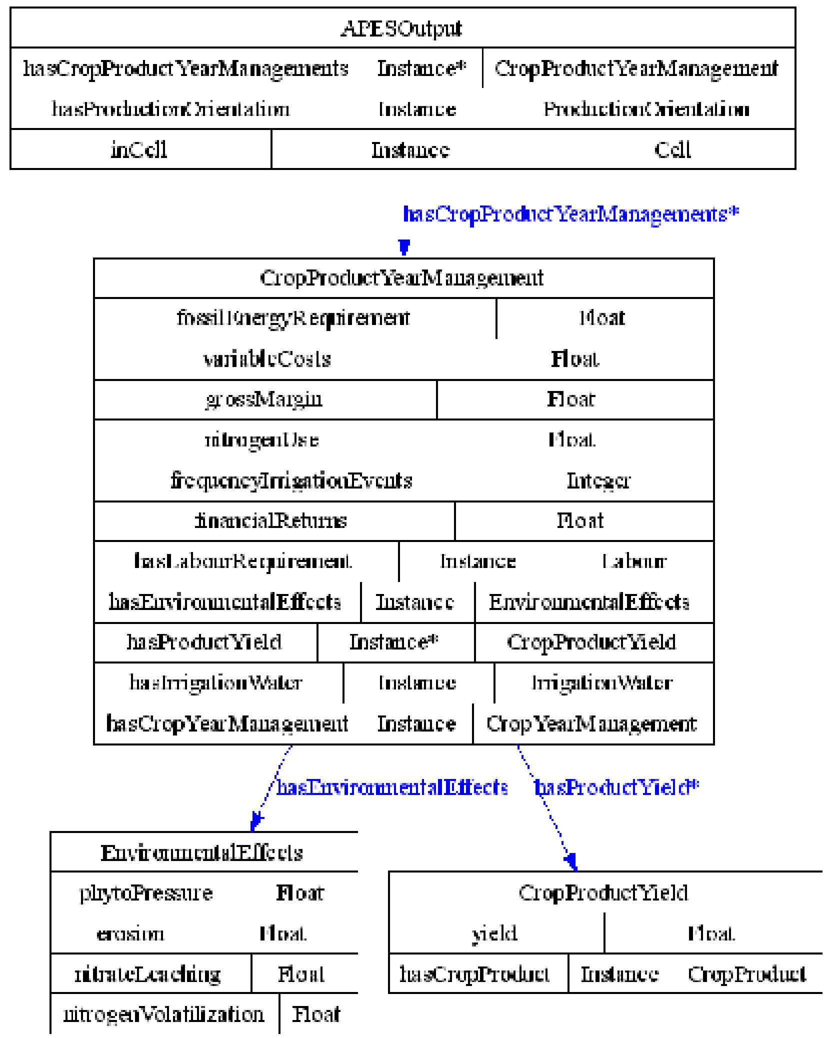

Figure 15 The APESOutput concept and its relationships with the concepts of Crop-ProductYearManagement, EnvironmentalEffects and the CropProductYield

4.5 Ontology processing by computers

As mentioned in Section 2 an ontology can be expressed in a machine readable format, i.e. as the Web Ontology Language (McGuinness & van Harmelen, 2004), which was used in our case study. An example of the developed ontology as an OWL-file can be found in Appendix 1. A computer can use an ontology i) for storing and retrieving data as done in databases or for linking between model code and databases, as an ontology is comparable with the conceptual schema of a database (Gruber, 1993); ii) automated generation of program code in Java or other software or modelling languages; iii) as a storage space and library that can be queried for relevant metadata of concepts (Brilhante, Ferreira, Marinho, & Pereira, 2006) and iv) for reasoning on the logical structure of the ontology (Horridge et al., 2004).

4.6 Discussion of developed ontologies

After an explanation of the models from different disciplines and methodologies, we demonstrated how these models can be linked through meaningful inputs and outputs, which are stored as concepts in an ontology. Between CAPRI and FSSIM the linking is currently done straightforward through crop products. The crop products in FSSIM are aggregated to crop products in CAPRI within EXPAMOD. Between APES and FSSIM the linking is more comprehensive as many more concepts are considered in the linking, like the agricultural activity, events, crop products, environmental effects and many more different data need to be provided in the linking. The use of the concepts within the models is not of interest to the common ontology developed. This allows modellers to implement their models in any appropriate modelling language, while the common ontology requires them to be explicit about the inputs they require from other models or databases and the outputs they provide to other models or databases. The ontology remains open to appending it with other models, for example with global trade model GTAP, or to new models within a chain of already integrated models.

The common ontology developed in this research could benefit from:

i) supplementing it with additional meta-data on the concepts it holds, like units, minimum and maximum value, source and references (Brilhante, Ferreira, Marinho, & Pereira, 2006; Rizzoli et al., 2005);

ii) enriching it with more concepts that are internal to the models and thus provide a more comprehensive overview;

iii) supplemented it with restrictions and axioms on the relationships between the concepts to support reasoning on the ontology;

iv) involving more stakeholders outside the research community, for example policy makers and farmers, to further clarify and expand the set of concepts in the ontology and to ensure the usefulness of the models for a diverse range of users. These developments will make the ontology more comprehensive and easier to understand as the current version, which would improve the portability and re-usability of the common ontology (Gruber, 1993; Rizzoli *et al.*, 2005), as it will also be made available online.

The collaborative approach was successful in the development of the ontology, although some attrition of participants occurred during the process, as also mentioned by Holsapple & Joshi (2002). The number of iterations required to arrive to the first version of a working common ontology was relatively low with three iterations. Probably in the further development of the ontology more iterations will be required as the common ontology will become more complex and more information needs to be added. As model development was done simultaneously with ontology development, the ontology development provided for participants a discussion platform to clarify what models were supposed to be doing and how. The simultaneous development of an ontology and of a model has a disadvantage as the ontology requires some updating to the developments in the models. It could be beneficial in the future to first develop the common ontology before to develop models committing to this common ontology. This would help to separate the knowledge itself from the implementation of this knowledge in a modelling language and thereby enforce the linking of the models with clearly defined concepts with useful meta-data.

5 General conclusion and recommendations

As demonstrated through a case study in which an ontology was developed for the multi-disciplinary project SEAMLESS, ontologies help to rigorously link models of different structures from different disciplines in a meaningful way, while forcing researchers to clarify the assumptions of their model interfaces and to set forth parts of their modeling knowledge, typically kept within their models. Thus, the islands of scientific understanding (Lowe, 2002) have been connected through the development of a common set of concepts with salient definitions. This common set of concepts and the connection between the islands of understanding could be extended with the incorporation of more stakeholders with different backgrounds leading to the development of a more comprehensive ontology. Although our case study was based on scientific project with scientists as participants, the collaborative approach could be easily used in other complex problems that involve experts with different jargon and stakeholders with different interests and outlooks. The development of a common ontology was beneficial in improving the understanding of the complex natural and economic system that is considered within SEAMLESS. An ontology can be beneficial in further ensuring that scientific knowledge is salient, legitimate and credible, which increases the changes of scientific knowledge being effective in influencing societal processes and environmental assessments, as noted by Cash et al. (2003) too.

Acknowledgements

The work presented in this publication is funded by 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).

References

Antoniou, G., & van Harmelen, F. (2004). *A Semantic Web Primer*. Cambridge, Massachusetts; London, England: The MIT Press.

Aristotle. (336-332 BC). *Metaphysics*.

Armington, P. S. (1969). *A theory of demand for products distinguished by place of production*: IMF.

Athanasiadis, I., Rizzoli, A. E., Donatelli, M., & Carlini, L. (2006, July 2006). *Enriching software model interfaces using ontology-based tools.* Paper presented at the iEMSs Third Biannual Meeting "Summit on Environmental Modelling and Software Third Biennial Meeting", Burlington, USA.

Blundell, R., & Stoker, T. M. (2005). Heterogeneity and aggregation. *Journal of economic literature 43*(2), 347-391.

Brilhante, V., Ferreira, A., Marinho, J., & Pereira, J. S. (2006). *Information Integration through Ontology and Metadata for Sustainability Analysis.* Paper presented at the iEMSs Third Biannual Meeting "Summit on Environmental Modelling and Software Third Biennial Meeting", Burlington, USA.

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.

Deaton, A., & Muellbauer, J. (1980). *Economics and Consumer Behavior*. Cambridge, UK: Cambridge University Press.

Donatelli, M., Bellocchi, G., & Carlini, L. (2006). Sharing knowledge via software components: Models on reference evapotranspiration. *European Journal of Agronomy, 24*(2), 186-192.

Ewert, F., Van Ittersum, M. K., Bezlepkina, I., Oude Lansink, A., Brouwer, F., Andersen, E., et al. (2005). *'Development of a conceptual framework for integrated analysis and assessment of agricultural systems in SEAMLESS-IF'*. Wageningen: SEAMLESS-IP.

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.

Flanagan, K., Stevens, R., Pocock, M., Lee, P., & Wipat, A. (2005). Ontology for genome comparison and genomic rearrangements. *Comparative and Functional Genomics, 5*(6-7), 537-544.

Flichman, G., & Jacquet, F. (2003). Couplage des modèles agronomiques et économiques: intérêt pour l'analyse des politiques. *Cahiers d'économie et sociologie rurales, 67*, 1-20.

Freund, R. J. (1956). The Introduction of Risk into a Programming Model. *Econometrica, 24*(3), 253-263.

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

Harris, G. (2002). Integrated assessment and modeling - science for sustainability. . In R. Costanza, Joergensen, S.E. (Ed.), *Understanding and Solving Environmental Problems in the 21st Century*. (pp. 5-17.): Elsevier.

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

Horridge, M., Knublauch, H., Rector, A., Stevens, R., & Wroe, C. (2004). A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools, Online Tutorial. from http://protege.stanford.edu/plugins/owl/

Kates, R. W., Clark, W. C., Corell, R., Hall, J. M., Jaeger, C. C., Lowe, I., et al. (2001). ENVIRONMENT AND DEVELOPMENT: Sustainability Science. *Science, 292*(5517), 641-642.

Kleijnen, J. P. C. (2006). *DASE: Design and Analysis of Simulation Experiments.* . Tilburg: Unpublished book.

Knublauch, H. (2005). Protege OWL. Stanford: Stanford Medical Informatics.

Lowe, I. (2002). The Need for Enviroment Literacy. from http://www.staff.vu.edu.au/alnarc/onlineforum/AL_pap_lowe.htm

McGuinness, D., & van Harmelen, F. (2004). *OWL Web Ontology Language Overview*: WWW Consortium.

Meinke, H., Howden, M., & Nelson, R. (2006, Burlington, July 2006). *Integrated assessements of climate variability and change for Australian agriculture - connecting the islands of knowledge.* Paper presented at the iEMSs Third Biannual Meeting "Summit on Environmental Modelling and Software Third Biennial Meeting", Burlington, USA.

Musen, M. A. (1992). Dimensions of knowledge sharing and reuse. *Computers and Biomedical Research, 25*(5), 435-467.

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

Storey, M.-A. D., Musen, M. A., Silva, J., Best, C., Ernst, N., Fergerson, R., et al. (2001). *Jambalaya: Interactive visualization to enhance ontology authoring and knowledge acquisition in Protege.* Paper presented at the Workshop on Interactive Tools for Knowledge Capture, K-CAP-2001, Victoria, B.C. Canada.

Szyperski, C., Gruntz, D., & Murer, S. (2002). *Component Software - Beyond Object-Oriented Programming* (Second Edition ed.). Ñew York, NY: ACM Press.

Van Ittersum, M. K., & Donatelli, M. (2003). Special Issue of the European Journal of Agronomy: Modelling Cropping Systems. *European Journal of Agronomy, 18*(3-4), 187-394.

Van Ittersum, M. K., & Rabbinge, R. (1997). Concepts in production ecology for analysis and quantification of agricultural input-output combinations. *Field Crops Research, 52*(3), 197-208.

Villa, F., Donatelli, M., Rizzoli, A., Krause, P., Kralisch, S., & Van Evert, F. K. (2006, July 2006). *Declarative modelling for architecture independence and data/model integration: a case study.* Paper presented at the iEMSs Third Biannual Meeting "Summit on Environmental Modelling and Software Third Biennial Meeting", Burlington, USA.

Zander, P., & Kächele, H. (1999). Modelling multiple objectives of land use for sustainable development. *Agricultural Systems, 59*, 311-325.

Appendix 1: a sample of the developed ontology as an owl-file which is machine readable.


```xml
<owl:Class rdf:ID="CropProduct">
   <rdfs:subClassOf>
     <owl:Class rdf:about="#TCGDatatype"/>
   </rdfs:subClassOf>
   <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
   >This gives the crop product, which is a combination between a crop and a product. Each
        crop can have one or more crop products, for example wheat has grain and straw,
        while potatoes only has potato</rdfs:comment>
   <rdfs:label xml:lang="gms">PRD</rdfs:label>
 </owl:Class>
 <owl:Class rdf:ID="NutrientEvent">
   <rdfs:subClassOf rdf:resource="#Event"/>
   <rdfs:label xml:lang="aps">fertilization</rdfs:label>
 </owl:Class>
 <owl:Class rdf:ID="IrrigationWater">
   <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
   >specifies the characteristics of water used for irrigation.</rdfs:comment>
   <rdfs:subClassOf rdf:resource="#Input"/>
   <rdfs:label xml:lang="gms">FSSIM: watc</rdfs:label>
 </owl:Class>
 <owl:Class rdf:about="#ClippingEvent">
   <rdfs:label xml:lang="aps">clipping</rdfs:label>
   <rdfs:comment xml:lang="en">Clipping means cutting the crop, which can mean harvest
        (so it has an isHarvest-property)</rdfs:comment>
   <rdfs:subClassOf rdf:resource="#Event"/>
 </owl:Class>
 <owl:Class rdf:about="#AgriculturalActivity">
   <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
   >This is APES Input</rdfs:comment>
   <rdfs:subClassOf rdf:resource="#PTGDatatype"/>
 </owl:Class>
 <owl:Class rdf:ID="CropRotationRequirements">
   <rdfs:subClassOf>
     <owl:Class rdf:about="#PEGDatatype"/>
   </rdfs:subClassOf>
 </owl:Class>
 <owl:Class rdf:ID="Production">
   <rdfs:subClassOf rdf:resource="#CAPRI_concept"/>
 </owl:Class>
```