available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/envsci

A database for integrated assessment of European agricultural systems

Sander Janssen [a,b,*], *Erling Andersen* [c], *Ioannis N. Athanasiadis* [d], *Martin K. van Ittersum* [a]

[a] *Wageningen University, Plant Production Systems Group, P.O. Box 430, 6700 AK Wageningen, The Netherlands*
[b] *Wageningen University, Business Economics Group, P.O. Box 8130, 6700 EW Wageningen, The Netherlands*
[c] *Forest and Landscape, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Copenhagen, Denmark*
[d] *Dalle Molle Institute for Artificial Intelligence (IDSIA), Galleria 2, 6928 Manno-Lugano, Switzerland*

ARTICLE INFO

Published on line 6 March 2009

Keywords:
Policy making
Data and model integration
Database integration

ABSTRACT

A major bottleneck for data-based policy making is that data sources are collected, managed, and distributed by different institutions, residing in different locations, resulting in conceptual and practical problems. The use of dispersed data for agricultural systems research requires the integration of data sources, which means to ensure consistency in data interpretations, units, spatial and temporal scales, to respect legal regulations of privacy, ownership and copyright, and to enable easy dissemination of data. This paper describes the SEAMLESS integrated database on European agricultural systems. It contains data on cropping patterns, production, farm structural data, soil and climate conditions, current agricultural management and policy information. To arrive at one integrated database, a shared ontology was developed according to a collaborative process, which facilitates interdisciplinary research. The paper details this process, which can be re-used in other research projects for integrating data sources.

1. Introduction

1.1. Problem definition

Statistics and indicators based on data are essential to inform policy (Niemeijer, 2002; AbouZahr et al., 2007). Governments benefit from specialized statistical agencies for data collection, such as FAOSTAT (FAO, 2008), EUROSTAT (Eurostat, 2008) and national bureaus of statistics. Effectiveness and efficiency of policies can be evaluated through processing data on potential impacts, either after a policy is implemented (ex-post), or before a policy is implemented (ex-ante). For this purpose, different methodologies can be used, for example indicators and typologies derived from primary data, or indicators derived from quantitative modeling.

Indicators and typologies are means that can be used to process datasets to provide new insights. Both provide summarized information about complex issues (Andersen et al., 2007a). Indicators synthesize relevant data and indicate the change or define the status of something (Gallopin, 1997), while a typology is a stratification of data that is homogeneous according to specific criteria relevant to policy, such as environmental and economic performance (Andersen et al., 2007a). Relying directly on available data, indicators and typologies may be used to (a) identify or justify needs for policy intervention, and (b) assess ex-post the impact of previous and current policies. Indicators are indeed established for achieving both uses, as for example in the IRENA initiative on agri-environmental indicators (EEA, 2005), and in the assessment of the impact of the rural development programs of the

* *Corresponding author at*: Wageningen University, Plant Production Systems Group, P.O. Box 430, 6700 AK Wageningen, The Netherlands. Tel.: +31 317 484769.
E-mail address: Sander.Janssen@wur.nl (S. Janssen).
1462-9011/$ – see front matter
doi:10.1016/j.envsci.2009.01.007

European Union (EC, 2006). These uses are less acknowledged in relation to typologies. Recently Andersen et al. (2007b) argued that the criteria for a European farm typology may influence the assessment of policy changes.

Another technique to process data and inform policy making is Integrated Assessment and Modelling (IAM), which is used to assess the impacts of policies, technologies or societal trends on the environmental, economic and social sustainability of a system (Parker et al., 2002). IAM is a methodology that combines quantitative models representing different aspects of sub-systems and scales into an overall framework for Integrated Assessment (Parker et al., 2002). Quantitative models used in an IAM study originate from a different disciplines, operate on different spatial and temporal scales, and require diverse (and sometimes, overlapping) data-sources. Model integration within an IAM project requires that all input and output data of each model have to be integrated. Prominent examples of IAM relate to the assessment of climate change impacts (Weyant et al., 1996; Cohen, 1997) or water quality in catchment areas (Turner et al., 2001).

There are technical, conceptual and institutional barriers to the effective use of data for policy making (AbouZahr et al., 2007). Examples of technical barriers are missing data, i.e. missing values in a time series (Britz et al., 2007), uncertain data, i.e. noisy data (Refsgaard et al., 2005), and non-available data, i.e. no data sources available (Niemeijer, 2002). Conceptual barriers refer to different interpretations of data, while institutional bottlenecks include issues related to data management policies and conflict of interests between the hosting institutions. The use of dispersed data in IAM studies requires the integration of data sources, both in conceptual and technical terms. Here integration means to define shared concepts, to ensure consistency in data interpretation, units, spatial and temporal scales and to respect legal regulations of privacy, ownership and copyrights.

1.2. *Integrated database*

There have been several efforts in different application domains to bring various data sources together. For example, in the field of medical research, Ali et al. (2007) made an inventory of data sources available to assess the environmental conditions that could affect the frequency of chronic diseases in Pittsburgh, Pennsylvania. In the field of environmental sciences, Gobin et al. (2004) connected different data sources to assess indicators on the European scale relevant to soil erosion, while Refsgaard et al. (2005) integrated data on the Water Framework Directive of the European Union. Herrero et al. (2007) developed a generic household-level database to store data on crop-livestock systems in developing countries. Villa et al. (2007) have demonstrated how artificial intelligence tools can be used for developing next-generation ‘‘intelligent databases’’ for the transparent and sound valuation of ecosystem services. The INSPIRE initiative (INSPIRE, 2008) of the European Commission targets the creation of a European spatial information infrastructure that improves the interoperability and the availability of spatial data across the EU. Refsgaard et al. (2005), Herrero et al. (2007) and Villa et al. (2007) reported on the availability of an integrated database to store the datasets, so that these datasets can be re-used easily for policy assessments.

As data sources on agricultural systems are distributed across institutions, scientists, who are required to integrate data, typically extract data from the original data sources in an ad-hoc manner. This practice is certainly prone to errors and a paradigm shift is needed to overcome technical, conceptual and institutional problems. To support policy evaluation and policy impact assessment through indicators, typologies and models, there is a need for an integrated database on agricultural systems, which consistently combines data from different sources and which ensures easy availability of data. SEAMLESS (System for Environmental and Agricultural Modelling; Linking European Science and Society) is an IAM research project (Van Ittersum et al., 2008), which aims to provide a computerized framework to assess the impact of policies on the sustainability of agricultural systems in the European Union at multiple scales. This aim is achieved by combining micro- and macro-level analysis, addressing economic, environmental and social issues, facilitating the re-use of models and providing methods to conceptually and technically link different models (Van Ittersum et al., 2008). SEAMLESS provides a framework for policy assessment in agriculture by integrating relationships and processes across disciplines and scales and combining quantitative analysis with qualitative judgments and experiences (Ewert et al., 2009).

In SEAMLESS, models of different kinds, designed for specific purposes and scales, are integrated for achieving the overall project objectives. Part of the integration activity is related to the extensive data requirements of the models. Data need to be collected and made consistent and available for serving dynamic biophysical models, static bio-economic farm models and partial equilibrium market models, with the ultimate goal to provide multi-scale assessment capability as to agricultural systems (Fig. 1 and Section 2.2). To achieve this goal, it is required to integrate several data-sources related to European agriculture, including economic, biophysical, climatic data, model simulation input and output data, scientific workflow configurations and calculation of indicators into a single relational database schema. By *data integration* in this paper

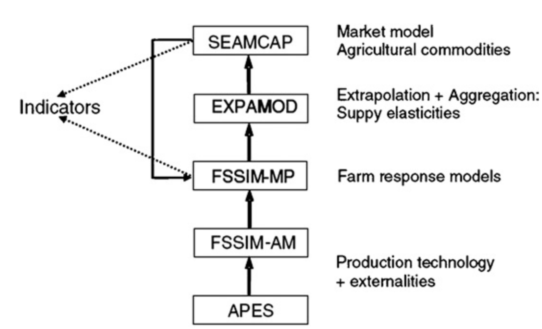

Fig. 1 – The models in SEAMLESS (after Van Ittersum et al., 2008). APES: Agricultural Production and Externalities Simulator; FSSIM-AM: Farm Systems SIMulator-Agricultural Management; FSSIM-MP: FSSIM-Mathematical Programming; EXPAMOD: EXtraPolation and Aggregation MODel and SEAMCAP: SEAMLESS version of the Common Agricultural Policy Regional Impact Analysis model.

we mean both *data alignment* across different sources, so that a unified schema is defined with references to shared concepts and scaled data structures, and *data homogenization*, by creating one single database that can simultaneously hold data from different sources.

The present paper describes the SEAMLESS integrated database on European agricultural systems and demonstrates the use of the data in the database for calculating indicators and for model inputs in IAM. The paper also describes the process of development of the SEAMLESS database, and the human factors involved in the process of reaching consensus across peers with clashing requirements and needs. To consistently define concepts across the different data sources, we adopted a structured process using an ontology as a means to arrive at one integrated database serving a set of models from different disciplines. We argue that this process is re-usable for other IAM projects, whereas we aim to make the end result (i.e. the database) freely available for non-commercial purposes in agricultural systems research and policy evaluations or assessments carried out in Europe. The paper illustrates how the development and use of a shared ontology facilitates interdisciplinary research through development of an integrated database.

Section 2 describes the relevant data sources and models of the SEAMLESS project. Section 3 presents the background and the process of ontology engineering. Subsequently, the results are presented in Section 4. The database on European agricultural systems is described, along with examples of the data present in the database and the process used to construct this database with a group of researchers. Section 5 offers a discussion of the database, the maintenance and support of the database and some reflections on the process of database development. Finally, conclusions and recommendations are provided.

2. Data sources and their use in models

2.1. Data sources

2.1.1. Farm Accountancy Data Network

The Farm Accountancy Data Network (FADN) (EC, 2008a) is an instrument for evaluating the income of agricultural holdings and the impacts of the Common Agricultural Policy. It consists of an annual survey carried out by the Member States of the European Union. Every year the Member States of the European Union collect accountancy data from a sample of the agricultural holdings (EC, 2008a). The sample only covers "professional" farms, which means that small, part-time, and hobby farms are poorly represented. Data collected per farm include physical and structural data, such as location, crop areas, livestock numbers, labour force, and economic and financial data, such as the value of production of different crops, sales and purchases, production costs, production quotas and subsidies. Data on farm management and externalities are not collected. Due to legal disclosure rules, data from FADN can only be displayed as averages of more than 15 sample farms, as data from individual farms should not be traceable for reasons of privacy.

2.1.2. European Soil Database

The European Soil Database (ESDB) (ESBN, 2008) provides a harmonised set of soil parameters, covering Europe (the enlarged EU) and bordering Mediterranean countries, to be used in agro-meteorological and environmental modelling at regional, national, and/or continental levels. It is 1 km $\times$ 1 km raster data and it contains the Soil Geographical Database of Eurasia, PedoTransfer Rules Database, Soil Profile Analytical Database of Europe and Database of Hydraulic Properties of European Soils (ESBN, 2008). These soil data have been supplemented with selected variables from the SINFO project (Baruth et al., 2006), which improved the soil parameters, pedo-transfer rules and the soil classification for use in a yield-forecasting tool. Finally, the map of organic carbon content in topsoils in Europe (Jones et al., 2005) was crucial for the development of the agri-environmental zones used in SEAMLESS (see Section 4.1.3).

2.1.3. European Interpolated Climate Data

The European Interpolated Climate Data (EICD) (JRC, 2008) provide interpolated daily data for a grid of 50 km $\times$ 50 km covering Europe and Maghreb (in most cases for the period 1975–today). The original observations originate from around 1500 meteorological stations across the European continent, Maghreb countries and Turkey. The observations at station level are not available in the dataset, only spatially interpolated data are (JRC, 2008). The interpolation is a simple two-step procedure in which the first step is the selection of up to 4 suitable meteorological stations for the determination of the representative meteorological conditions for a grid cell. The actual interpolation, the second step, is a simple average for the meteorological parameters, corrected for an altitude difference in the case of temperature and vapour pressure (Van der Goot, 1997).

2.1.4. Surveys on farm management

Farm management data have been collected through dedicated surveys as part of the SEAMLESS project (Borkowski et al., 2007). In the SEAMLESS project, a lack of European data on agricultural management was identified. Agricultural management data are the use of inputs (fertilizers, pesticides, irrigation) and the timing of input use at crop level, which are crucial to bio-economic farm models and biophysical crop growth models. FADN only provides aggregated farm level input data often expressed in monetary terms. Two different surveys (Borkowski et al., 2007) were developed as part of the SEAMLESS project: a detailed and a simple one. In the detailed survey, only data for arable systems were collected including timing and amounts of inputs, crop rotations, machinery, labour requirements and costs. It has been carried out in five regions in Europe (Brandenburg, Andalucia, Midi-Pyrenees, Flevoland and Zachodniopomorskie). The detailed survey was completed by regional experts, who in their day-to-day work provide advice to farmers or work regularly with farmers, and thus aims at describing an average farmer behavior. The detailed survey aims to meet the input requirements of biophysical crop growth models.

The simple survey was applied to a larger sample of 16 regions in Europe. It collects data on arable, livestock and perennial agricultural systems. These 16 sample regions aim to cover the range of biophysical conditions and farm types present in the European Union. The simple survey differs from the detailed survey as only a sub-set of the variables from the detailed survey is collected, including economic variables (e.g. costs, product prices), production, rotations and some aggregate variables describing input use (e.g. total nitrogen use for a crop or total medicine costs per livestock unit). The simple survey was completed by scientists working in the region with the help of farm management handbooks, which are used by farmers for advice.

2.1.5. COCO/CAPREG

The COCO/CAPREG dataset (Britz et al., 2007) is based on NewCronos (Eurostat, 2008) and FAOSTAT (FAO, 2008). Missing values, missing time series and incorrect values from New-Cronos and FAOSTAT were estimated and adjusted through statistical estimation procedures. COCO/CAPREG is the dataset linked to the SEAMCAP market model (a market equilibrium model detailed in Section 2.2.3). This dataset provides the data on agricultural policies and prices in the 27 Member States from 1985 to 2004, e.g. subsidies given to farmers for different regions, cuts of subsidies given to farmer, coupling degrees and prices per Member State, subsidized exports and tariff agreements between European Union and trading blocks.

2.1.6. Relevance of typologies

The datasets from the FADN, ESDB and EICD have been categorized into farm and regional typologies (Metzger et al., 2005; Andersen et al., 2007a; Hazeu et al., 2009) to enable modelling in homogenous spatial units and to allow for characterization of the variation in the environment, e.g. climate, soil and farms. This is useful for sampling purposes. For example, farm management data were not available and they cannot be easily collected for all regions across Europe due to budget and time restrictions. Instead, based on classification in typologies, representative regions were selected for the simple and detailed surveys (Section 2.1.4). Typologies are used to combine data, to provide a flexible and manageable data structure and to respect disclosure rules. Further regional typologies have been developed which characterize regions to provide contextual information for the assessments. Examples of regional typologies are livestock density, share of area in nitrate vulnerable zones and degree of rurality.

The data sources have been aligned with the existing administrative categorization of the EU territory, like the Nomenclature of Territorial Units for Statistics (NUTS) (EC, 2008b). In SEAMLESS, the NUTS-level of relevance is NUTS-2 (except for United Kingdom, where level 1 is used) and when reference is made to NUTS-regions in this paper, NUTS-2 regions are intended. EU25 has 270 NUTS-2 regions, which typically correspond to provinces, or constituent states/cantons.

2.2. Models using the data

2.2.1. APES: a dynamic crop growth simulation model

APES is a cropping system model estimating the biophysical processes of agricultural production systems, at point level, in response to weather, soils and different options of agro-technical management (Van Ittersum and Donatelli, 2003). APES is a modular simulation model targeted at estimating the biophysical behavior of agricultural production systems taking into account the interaction among weather, soil and crop characteristics and different options of agricultural management. Biophysical processes are simulated in APES with deterministic approaches which are mainly based on mechanistic representations of biophysical processes (Donatelli et al., 2009).

2.2.2. FSSIM: a bio-economic farm model

The Farm System SIMulator (FSSIM) is a bio-economic farm model developed to assess the economic and ecological impacts of agricultural and environmental policies and technological innovations. A bio-economic farm model is defined as a model that links formulations describing farmers' resource management decisions, to formulations that describe current and alternative production possibilities in terms of required inputs to achieve certain outputs (both yield and environmental effects) (Janssen and Van Ittersum, 2007). FSSIM consists of a mathematical programming model (FSSIM-MP), and an agricultural management module (FSSIM-AM) (Louhichi et al., 2009).

2.2.3. SEAMCAP: a market level model

SEAMCAP is a version of the model Common Agricultural Policy Regional Impact Analysis (CAPRI) (Heckelei and Britz, 2001) integrated in SEAMLESS. CAPRI is a partial equilibrium model for the agricultural sector. SEAMCAP makes use of non-linear mathematical programming tools to maximise regional agricultural income with explicit consideration of the Common Agricultural Policy 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 module, which interact iteratively.

2.2.4. EXPAMOD: a regional upscaling model

EXPAMOD is an econometric model describing price-quantity responses of farms given specific farm resources and biophysical characteristics that are available EU-wide (Pérez Domínguez et al., this issue). It provides an aggregation procedure to make the regional supply modules of CAPRI behave like the aggregate of the farm (FSSIM) models of the same region – apart from additional aspects entering the market supply such as regional land or political constraints (premium ceilings). All available FSSIM models run for ranges of exogenously fixed prices, computing multi-dimensional price-quantity response surfaces. Thus, the econometric model 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 (Pérez Domínguez et al., this issue).

3. Database development, data consistency and integration

3.1. Process of database development

Data integration across the sources presented above requires to take into account complex conceptual problems, related to

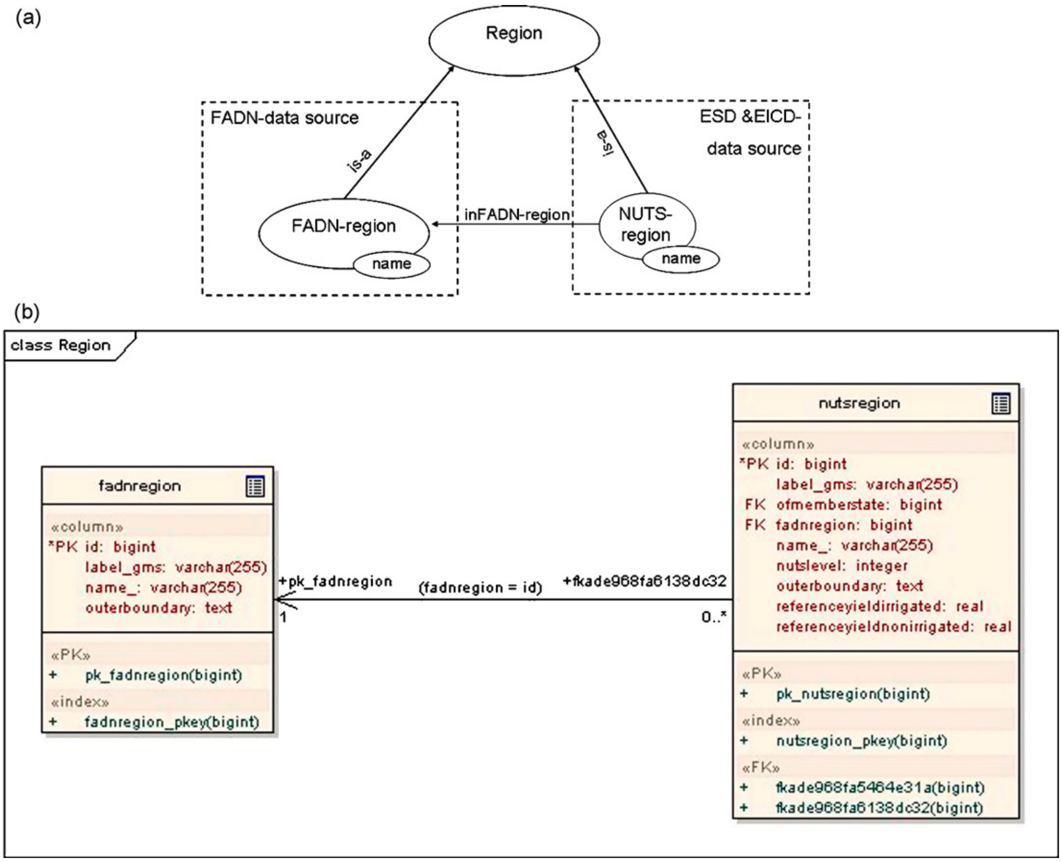

Fig. 2 – The different types of Regions in the integrated database in an ontology schema (a) and a relational database schema (b). The same relationship is represented in (a) and (b) between NUTS-region and FADN-region, with the difference that the relationship in the ontology schema (a) has a name ('inFADN-Region') and definition, while this is not the case in the relational database (b). ESDB = European Soil Database; EICD = European Interpolated Climate Data. (a) The different definitions of the concept Region between data sources. (b) The representation of the relationships between FADN-region and NUTS-region in a relational database.

the terminology adopted, the scale of information, and the heterogeneity of the original database schemas. For example, FADN, ESDB and EICD all refer to a ''Region'' entity. In the case of FADN, the definition of regions is different than those of ESDB and EICD. ESDB refer to soil mapping units and EICD refer to 50 km × 50 km grid, which were both linked to NUTS regions (Fig. 2a), when preparing the data for the database. FADN uses a delineation of regions that is specific to FADN, and these regions are referred to as FADN-regions in this paper. In integrating the data sources in one database schema, these data sources have to be adapted to shared concepts, to respect geographical entities and to be aligned in time, e.g. covering overlapping time periods. Integrating the data sources into one database is a time consuming and challenging task that requires collaboration of scientists from agricultural economy, environmental science, agronomy and computer science, with dissimilar education and research experience.

To tackle the heterogeneity of the constituent data schemas, we developed an overall ontology, covering the union of the constituent data sources and domains. An ontology is the appropriate tool for defining a shared conceptual schema, as ontologies consist of a finite list of concepts and the relationships between these concepts (Antoniou and Van Harmelen, 2004), and they are expressive enough for defining equivalent entities, hierarchies, complements, unions or intersections, based on description logics. This was particularly useful for marking and resolving ambiguities across the original schemas.

A shared ontology is an ontology that is jointly developed between a group of individuals, in this case researchers. A collaborative approach was adopted for developing a shared ontology about the different data sources in SEAMLESS. Our development was 'a joint effort reflecting experiences and viewpoints of persons who intentionally cooperate to produce it' and thus requires a consensus-building mechanism (Holsapple and Joshi, 2002). Part of our effort was based on an inductive approach (Holsapple and Joshi, 2002), where the shared ontology was developed by examining and analyzing the initial data sources and extracting relevant properties or discussing the relationships between concepts in these data sources.

Semantic modelling and ontologies are more powerful for domain modelling than conventional relational data schemas, and this is why we adopted ontologies for defining the integrated schema, First of all, ontologies are richer in their representation of relationships between concepts than relational database schemas (Fig. 2). In an Web Ontology Language

(OWL) (McGuinness and Van Harmelen, 2004) ontology relationships have a direction and can be shared across concepts, restricted with logical constraints, and form hierarchies. Also, ontologies have a strong inter-operability background, as they are in line with the Semantic Web initiative (Berners Lee et al., 2006). There are much more tools and techniques for ontology alignment and integration. For example, two ontologies developed in separate efforts can easily be linked to each other by investigating the semantic relationships between their concepts (El Gohary and El Diraby, 2005). Furthermore, OWL ontologies can be connected by a reasoner that is based on description logics and thus data can be validated against logical constraints. Finally, an ontology may be considered a distinct product for capturing knowledge, which can be re-used in the future for building other systems.

3.2. Technical implementation

The shared ontology was subsequently translated into a relational database schema. A relational database schema provides the structure of the database, in which the data from the different data sources can be entered. This translation from ontology to relational database schema was done based on the conventions of the Semantic-Rich Development Architecture (SeRiDA) (Athanasiadis et al., 2007a,b), which acts as a bridge between different programming paradigms, e.g. object-oriented programming, relational databases and ontologies (Athanasiadis et al., 2007a). Object-oriented programming is used in SEAMLESS for model and application development, relational databases for persistent storage of data and ontologies for defining and storing knowledge.

The integrated database is running on a PostgreSQL database server (PostgreSQL, 2008). The models are linked to the database through Hibernate (JBOSS, 2008) and the exchanged datatypes are implemented with JavaBeans™. The database is linked to a spatial database that provides geographical information, exploiting the PostGIS capabilities (PostGIS, 2008). The spatial information is also available through Web Mapping and Web Feature Services provided by a GeoServer (GeoServer, 2008). The entire database deployment and data management solution is based on Open Source software. A detailed description of the data management process and the technical integration of the models is discussed in Athanasiadis and Janssen (2008).

4. Results

4.1. Database on European agricultural systems

4.1.1. Full ontology

Fig. 3 provides a partial view of the ontology developed for the database on European agricultural systems as developed in the SEAMLESS project. Fig. 3 illustrates the part related to soil, farm and climate data. It includes typology concepts, such as Farm Specialization and Farm Size, concepts that facilitate spatial links, such as NUTS regions and Climate Zones, and concepts that hold the actual data, such as Representative Farm, Soil Characteristics and Daily Climate entities. The current version (October 2008) of the database consists of 379 tables including 2379 fields and with 487 relations between the tables. The database exceeds 12 million records.

4.1.2. Representative Farms

A central concept of the ontology is the concept of Representative Farm, which defines a Farm Type in an FADN region in Europe for a specific year. A Farm Type is specified according to the dimensions of farm size, farm intensity and farm specialization and land use (Andersen et al., 2007a) (Fig. 4). As an example of a classifying concept, Farm Intensity classifies farms according to their total monetary output of agricultural produce per hectare (Andersen et al., 2007a). If the total output is below 500 euros per hectare, then the farm falls in the class of low intensity, if it is between 500 and 3000 euros, then it is medium intensity and if is more than 3000 euros, then it is high intensity. The threshold values are adjusted with yearly producer price indices. The values presented above refer to 2003. While a Farm Type is not linked to a specific region or year, a Representative Farm is associated to a region and a year (Fig. 4).

4.1.3. Climate and soil data

Another central concept is that of the agri-environmental zone, that links soil and climate data. An AgriEnvironmentalZone is a unique combination of an EnvironmentalZone, the SoilType and NUTS-region. An AgriEnvironmentalZone is the smallest homogenous area in terms of climate and soil data. Environmental Zones are used to stratify the diverse European Union in zones with a similar climate (Metzger et al., 2005). Environmental Zones cover more than one administrative region. A Climate Zone is a unique combination of a NUTS-2 region and Environmental Zone and for each Climate Zone, a set of climate data are available. This set of climate data includes the daily climate data for a 25 years time period. Examples of daily climate data attributes are rainfall, minimum and maximum temperature and wind speed at 10 m.

Each AgriEnvironmental Zone is linked to a set of soil data, which are classified in Soil Types. Six different Soil Types were defined according to topsoil organic carbon levels (Hazeu et al., 2009). For each unique combination of a Soil Type and a Climate Zone a set of soil data is available as stored in the concept of Soil Characteristics. Examples of properties of the soil characteristics are thickness of soil layers, textural class and maximum usable moisture reserve.

The link between AgriEnvironmental Zones and Representative Farms is made through statistically allocating an area of an AgriEnvironmental Zone to each Representative Farm (Elbersen et al., 2006). This implies that the farmed area within each AgriEnvironmental Zone is allocated to one or more Representative Farms and each Representative Farm manages farmed areas in one or more AgriEnvironmental Zones. As can be seen from Fig. 3, Representative Farms and AgriEnvironmental Zones are based on different administrative regions. AgriEnvironmental Zones refer to NUTS-regions and Representative Farms refer to FADN-regions. Usually the borders of the NUTS and FADN-regions coincide, however some FADN-regions consist of several NUTS-regions. Through allocating the area of Agri-Environmental Zones to Representative Farms, this mismatch between the borders of FADN-regions and NUTS-regions has been resolved.

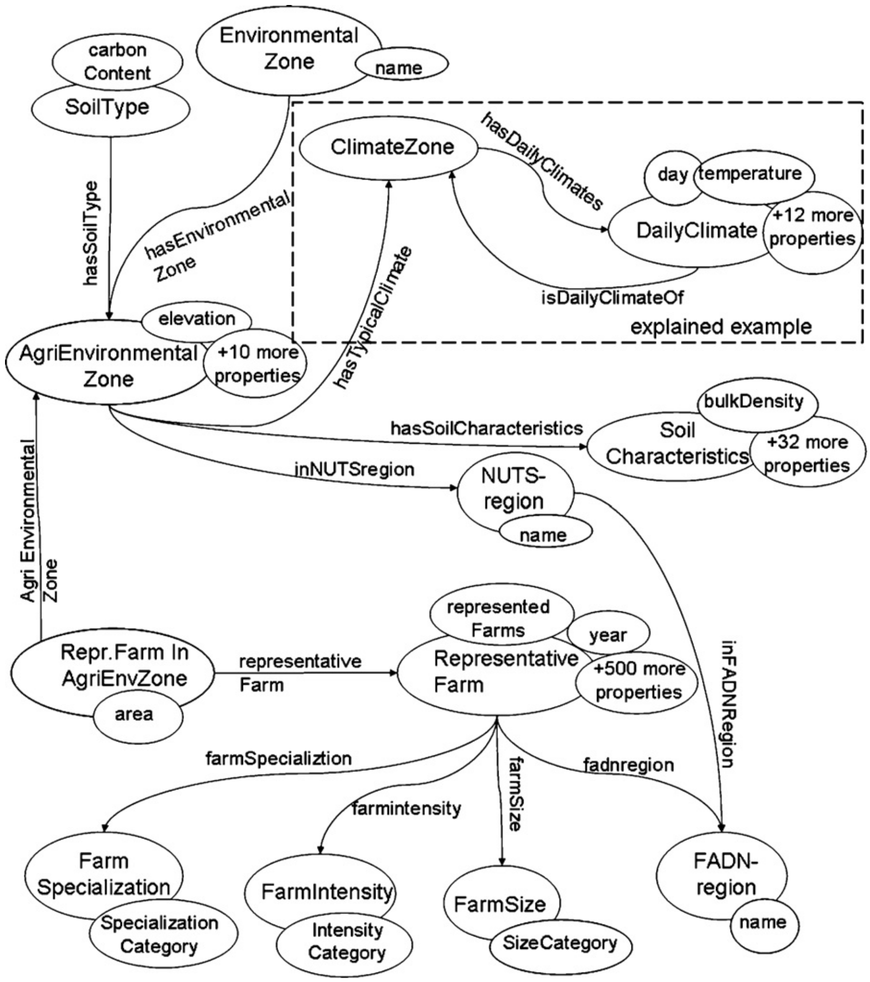

Fig. 3 – An ontology-schema of the database on European agricultural systems showing the parts on farms, soils, climate and their links. Two concepts (ClimateZone and DailyClimate: large ellipses), their relationships (hasDailyClimates and isDailyClimateOf: uni-directional arrows) and the properties of the concepts (name, temperature, and 12 more properties: small ellipses) can be found in the explained example (dashed box). The figure can be read by following the direction of the arrows, for example ClimateZone.hasDailyClimates (DailyClimate). A daily climate is characterized by a day, a temperature of that day and twelve more properties.

4.1.4. *Agricultural management*

As the agricultural management differs between and within regions, Regional Agricultural Management Zones were created. A Regional Agricultural Management Zone can be linked to distinct sets of agricultural management data and each Regional Agricultural Management Zone refers to one or more AgriEnvironmental Zones (Fig. 5). The central concept in Fig. 5 is the RotationElement, which signifies 1 year of crop

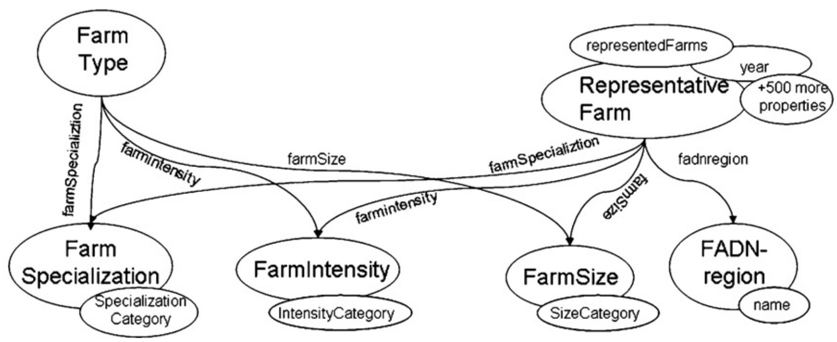

Fig. 4 – The concepts Farm Type and Representative Farm and the relationships to their classifying concepts (for an explanation of how to read this figure, see Fig. 3).

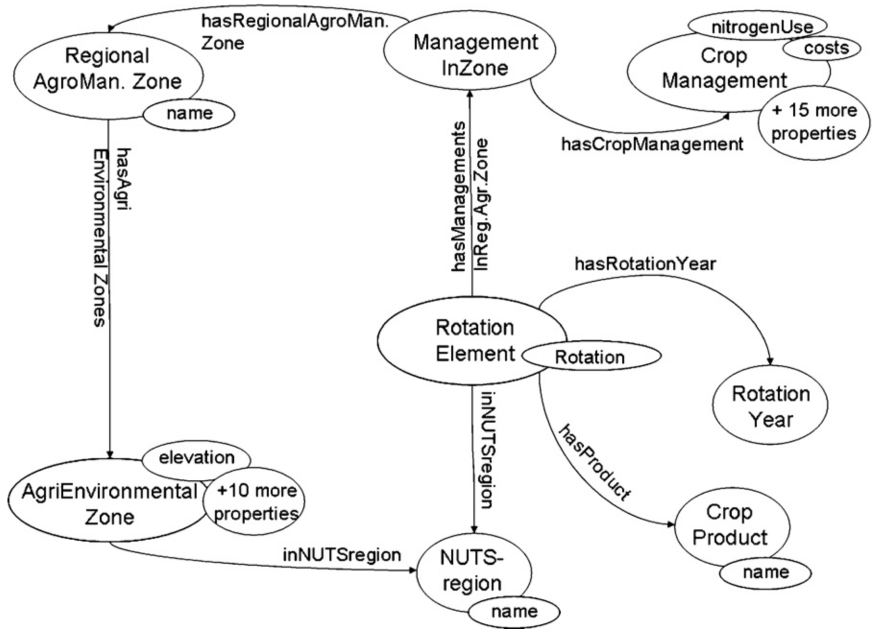

Fig. 5 – Agricultural management data and their links to NUTS-regions and AgriEnvironmental Zones (for an explanation of how to read this figure, see Fig. 3).

rotation as found in a region. A rotation is defined as a sequence of crops in time and space, where the last crop is the predecessor of the first crop (creating a loop) and rotations are widely practiced in agriculture for pest control, soil fertility management and risk diversification. The RotationElement links to one or more ManagementInZones, which means that crop management applied to the crops in a rotation is different across Regional Agricultural Management Zones. For example, in a Regional Agricultural Management Zone with high yield potential (favourable climate and soils) more nitrogen may be applied to the crop than in a Regional Agricultural Management Zone with lower yield potential. This assumes that rotations are the same throughout the NUTS-region, and only the management of the crops in the region can differ. Note that for explanatory purposes we focused here on arable crop management data. The database also holds data on livestock and perennial systems.

4.1.5. *Policy data*

Finally, data on agricultural and environmental policies are linked to Member States and NUTS-regions. Each Member State consists of one or more NUTS-regions. Fig. 6 shows part of the database schema for policy data, which are related to the premiums the European Union pays to farmers and the decoupling of these premiums as part of the Common Agricultural Policy 2003 reform (EC, 2003). Decoupling means that financial support is not related to production anymore, but farmers receive income support instead. The European Union has established premium amounts per premium groups (e.g. a group of crops or animals for which the same premium is provided), the Basic Premiums. Individual Member States can, with some restrictions, decide on the percentage of decoupling of these Basic Premiums for different Premium Groups. These relationships are shown in Fig. 6, as the Basic Premiums are not linked to a NUTS Region and as the relationship ‘hasMemberstate’ between Coupling Degree and Premium Group in Fig. 6. The database on European agricultural systems contains policy data for 14 more policy measures.

4.2. *Method to develop the integrated database*

To develop a shared ontology for all data-sources, three scientists (a computer scientist, an expert on agri-environmental policies, and a systems analyst) engaged in an integration process. These three scientists involved other domain experts in the integration process, when additional knowledge was required.

The integration process was an iterative procedure, with four milestones: three intermediate ‘‘prototypes’’ each concluded with a stable version of the database schema used for running the models, and one final version. Every prototype began with a phase of ontology building and review in several iterations. The ontology was developed using Protégé (Knublauch, 2005), an ontology editor. Once the shared ontology was fixed, it was exported to a relational database schema using the SeRiDA-framework (Section 3.2). Subsequently, the data from the original sources were entered into the database, which led to the identification of obstacles and further issues. These issues were discussed again in a new iteration among the domain scientists involved, and the solutions were reflected in the ontology, resulting in an updated stable version of the database schema, which was then released as a version and linked to the models. As a final step in each prototype, the relational database schema and shared ontology were reviewed by the three scientists involved and lists of improvements were made. During the review of the database schema of the three prototypes, scientists tried to simplify the shared ontology and relational database schema

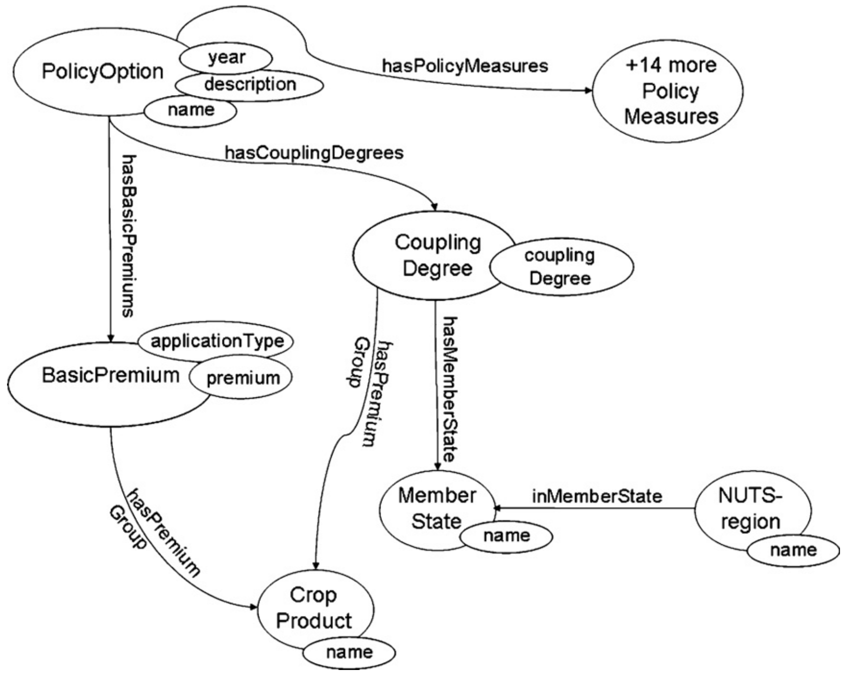

Fig. 6 – Part of the policy data related to premiums for farmers and coupling degree of these premiums to production (for an explanation of how to read this figure, see Fig. 3).

as much as possible, as the shared ontology had the tendency to grow in detail and complexity.

As part of the fourth and final version of the database schema, metadata have been included as part of the ontology in accordance with ISO (ISO, 2008) and the INSPIRE (INSPIRE, 2008) standard. The metadata document the original data sources, textual descriptions, units and contact persons for the original data source.

4.3. *Examples of the use of data from the database*

4.3.1. *Examples of data extractions (use of typologies and indicators)*

The following section gives some examples of data that can be extracted directly from the SEAMLESS database providing novel ways of aggregating or combining the original data. The first example provides an overview of the trends in intensity of farming in EU-15 since 1994 (Table 1). The example is based on FADN data aggregated using the typology of farm types according to intensity. Clear trends can be identified in Table 1: the share of the agricultural area managed by low intensity has declined continuously over the period from almost 30% to close to 20%. At the same time the agricultural area managed by high intensity farms has increased with 5.1% from 7.9% of the area to 13% of the area.

The second example shows the relationship between livestock density and the Nitrate Vulnerable Zones designated according to the Nitrates Directive (Table 2). The Nitrate Vulnerable zones include catchments that drain to surface and groundwater where the NO_3 content exceeds 50 mg/l. As can be seen in the table, there is a tendency towards higher livestock densities in the regions with the largest share of the area as Nitrate Vulnerable Zones. In the 68 regions where the entire area is designated as NVZ, 63% of the regions have an average livestock density above 1, and 22% have a livestock density above 2. Compared to this, livestock density only exceeds 1 in 27%, and 2 in 4% of the regions, where no areas are designated as Nitrate Vulnerable Zones. The analysis can be used to identify the hot spots, where a high livestock density is found in regions with a large share of the area designated.

Table 1 – The share of agricultural area in EU15 managed by different farm types according to intensity in the period from 1995 to 2004[a].

	Share of agricultural area managed by		
	Low intensity farms	Medium intensity farms	High intensity farms
1995	29.5	62.6	7.9
1996	28.6	63.0	8.3
1997	27.9	63.3	8.8
1998	27.3	63.5	9.2
1999	25.8	63.6	10.6
2000	24.4	64.2	11.4
2001	26.3	62.2	11.5
2002	23.8	64.6	11.7
2003	23.3	64.7	12.0
2004	21.0	66.0	13.0

[a] Source: EU FADN-DG AGRI/G-3; SEAMLESS adaptation. The Farm Intensity is defined by the total monetary output per ha. The threshold values vary over time according to the price index on agricultural products. In 2003 the threshold values were: Low intensity <500 €, medium intensity ≥500 and <3000 € and high intensity ≥3000 € (Andersen et al., 2007a).

Table 2 – Number of NUTS2 regions in EU15 according to livestock density (Livestock units per ha) in 2003 and share of area designated as Nitrate Vulnerable Zones (NVZ)[a].

NVZ area in % of total area	Livestock density LU/ha				
	<0.5	≥0.5 and <1	≥1 and <2	≥2	Total
0	17	18	11	2	48
>0 and <33	7	21	12	10	50
≥33 and <66	2	6	5	1	14
≥66 and <100	3	7	3	1	14
100	8	17	28	15	68
Total	37	69	59	29	194

[a] Source: EU FADN-DG AGRI/G-3; SEAMLESS adaptation.

The last example explores whether there is a relationship between soil types, farm family income and livestock densities (Table 3). Generally it is assumed that the agronomic potential of the soils increases with increased carbon content in topsoil except for the soils with a very high content, which normally are related to other restrictions on agricultural production. This high agronomic potential might lead to a higher family farm income. However, looking at the family farm income in Table 3, no correlation exists: The lowest income per hectare is found on soil type 5 and the highest on soil type 1 with no trend in between. Taking under account the livestock density, on the one hand we observe that the livestock density seems to be correlated with the soil types with an increasing density following increasing carbon content. On the other hand, this cannot be used to explain the variation in family farm income. The SEAMLESS database provides data that can be used to explore these relations also at regional and local levels or to seek for other variables to examine relationships.

4.3.2. *Data as model inputs for FSSIM simulations*

FSSIM, a bio-economic farm model (Section 2.2.2), has been applied to Flevoland, a NUTS2-region in the Netherlands. Flevoland has been reclaimed from the sea in the 1960s and its young and fertile soils are very productive. It is very homogenous in terms of soil and climate as can be seen in Table 4. For FSSIM, the data of relevance from the database are the Representative Farms found in Flevoland, as for each of these Representative Farms, FSSIM executes to obtain a simulated cropping pattern. FSSIM execution requires a set of possible farming rotations to choose from, and data on the associated crop management for each rotation. This is illustrated for one sample rotation in Table 4. The database provides all these data, and by using hibernate querying facilities, the model is able to retrieve them easily. The results of the model are verified by comparing the simulated cropping pattern with the observed cropping pattern as found in 2003 for each of the representative farms in Flevoland. In Table 4 only a limited subset of the input data for FSSIM have been provided, e.g. the database contains more rotations, representative farms for more years and many more properties to describe the soil, climate and representative farms found in Flevoland.

Table 3 – Family farm income and livestock density for seven soil types in the EU[a].

Soil type	Carbon content in topsoil	Family farm income (€/ha)	Livestock density LU per ha
1	<1.23	866	0.5
2	≥1.23 and <2.46	594	0.8
3	≥2.46 and <3.94	416	0.9
4	≥3.94 and <5.66	671	1.0
5	≥5.66 and <8.86	338	1.1
6	>8.86	435	1.2
7	No soil information	463	1.2

[a] Source: EU FADN-DG AGRI/G-3; SEAMLESS adaptation. In SEAMLESS the soil types are defined by the carbon content in the topsoil (Hazeu et al., 2009).

5. Discussion

5.1. *Use of the database*

The database on European agricultural systems holds data on different aspects of the agricultural systems, e.g. cropping patterns, production, farm structural data, soil and climate conditions, current agricultural management and policy information. As demonstrated in Section 4.3, the database can be used to directly compute indicators related to agricultural and environmental policies in Europe or for policy assessments through the use of one or a set of models. The database in its current form is used by the models APES, FSSIM, SEAMCAP and EXPAMOD (Section 2.2). New models and indicators with similar data needs can easily be linked to the database, for example, the database could be useful for computing indicators on soil erosion (Gobin et al., 2004), energy use (Pervanchon et al., 2002), crop diversity (Dramstad and Sogge, 2003), pesticide usage and leaching (Reus et al., 2002) and marginalization based on farm income and employment (EEA, 2005).

The data in the database are organized according to typologies (Section 2.1.6), which implies that it is based on aggregated data (e.g. farm typology (Andersen et al., 2007a)), interpolated data (e.g. EICD (JRC, 2008)) or categorized data (e.g. ESDB (ESBN, 2008)). The database does not contain the original data on which these averages, interpolations and categorizations are based, which is required to respect disclosure rules and to avoid data pre-processing for each model and indicator computation.

Table 4 – A sample of data for Flevoland region in the Netherlands for 2003 (source among others: EU-FADN-DG AGRI-G3, Meteorological data Source JRC/AGRIFISH Data Base – EC – JRC).

Representative farms

Farm Specialization/Land Use	Farm Intensity	Farm Size	Usable Farm Area (ha)	Area in potatoes (ha)	Area in sugar beet (ha)	Area in wheat (ha)	Percentage area per AgriEnvironmentalZone	
							2993	1317
Arable/specialized crops	High intensity	Medium scale	16.8	4.8	3.1	2.7	100%	
Arable/specialized crops	Medium intensity	Large scale	67.2	17.9	11.2	10.4	100%	
Arable/specialized crops	High intensity	Large scale	69.0	24.8	9.1	11.5	90.3%	9.7%
Arable/others	High intensity	Large scale	35.2	3.5	1.3	2.0	100%	

Climate-Soil data per AgriEnvironmental Zone

AgriEnvironmental Zone	Environmental Zone[a]	Average rainfall (mm/day)	Average minimum temperature (°C)	Average maximum temperature (°C)	SoilType: soil carbon content (%)	Clay content (%)	Sand content (%)	Silt content (%)
2993	Atlantic Central	1.6	6.2	14.6	3.9–5.6%	20	45	35
1317	Atlantic North	1.6	6.2	14.4	3.9–5.6%	20	45	35

Sample rotation for Flevoland

Crop	Year	Costs NPK fertilizer (€/ha)	Sowing date (week of year)	Labour (h/ha)	Nitrogen use (kg N/ha)	Yield (tonnes/ha)
Soft wheat	1	113	42	10.7	205	8.2
Potato	2	269	11	27.5	255	53.4
Soft wheat	3	113	42	10.7	205	8.2
Sugar beet	4	147	13	19.6	150	65.5

Policy data for direct payments in Flevoland in 2013

Premium group	Premium (€/tonnes)	Decoupling degree (%)	Regional reference yield (tonnes/ha)
Energy crops	45	100	4.9
Cereals, oilseeds and pulses	63	0	6.6
Obligatory setaside	63	0	4.9

[a] Flevoland is on the border of two environmental zones.

The database aims to achieve a full coverage of the European Union, but this is not feasible for all data sources. For example, the FADN (EC, 2008a) contains data about the 10 new Member States only for 2004 onwards, while for 12 'old' Member States data are available from 1990 and onwards and for Austria, Finland and Sweden data are available from 1995 and onwards, as these joined the EU in 1995. There was no European-wide data source available on agricultural management, so a first effort was made in the SEAMLESS project to collect this type of data for a sample (Section 2.1.4) out of the 270 NUTS-regions in the EU25. Obviously work is required to add more regions and to obtain time series in order to increase the representativity of the agricultural management data. Still the database holds the most complete set of data available on agricultural systems in Europe, and data gaps are due to the original data sources on which the SEAMLESS database depends.

5.2. Availability, extension, support and maintenance

The database will be made available for non-commercial use in other projects requiring data on agricultural systems in the European Union (Information on access can be found on www.seamlessassociation.org and additional documentation can be found in Andersen et al. (2007c)). Using the SEAMLESS integrated database instead of using original data sources has the advantages that (a) several data sources are available on one server instead of on several locations in different formats, (b) difficult questions of data integration and consistency have been solved by specialists familiar with the original data sources and (c) the pre-processing of the original data sources is already done.

A plan for the maintenance of the database beyond the lifetime of the SEAMLESS project is available that ensures the database will be available in an updated version for at least 3 years and hopefully longer. The maintenance plan provides full documentation of how to update the database with data from the different data sources and ensures that new versions of relevant data sources in the database will be included as they come available, for example a dataset for 2005 for FADN data (EC, 2008a). Not all the original data sources are frequently updated in their structure and content, although for some data sources (e.g. FADN (EC, 2008a) and EICD (JRC, 2008)) new data become available annually. The introduction of new versions of the data sources can be automated, although this is dependent on the stability of the original data sources in their variables and structure. New models and indicators might require new data, that is not currently in the database, for which the database needs to be extended. Extension of the database with new data sources is encouraged and the methods described in Sections 3.1 and 4.2 for conceptual and technical integration are recommended. For this, the shared ontology would need to be extended for the new dataset and links to the concepts already in the ontology need to be made. Second, the database schema can be made and finally, this can be filled with the new data.

5.3. Reflection on development and technical implementation

The integration of multiple data sources into one shared ontology following an iterative process was successful, as it led to one database schema in which all the data from different sources could be stored. The iterative process with different versions was required to step-wise improve the shared ontology. During the review of the first and second version of the shared ontology it was concluded that the shared ontology was too complex and that some relationships between concepts were ambiguous and therefore difficult to understand. The use of shared ontologies can highlight such complexities and ambiguities as scientists are forced to clearly define the concepts in the ontology and as the concepts have to be consistently and coherently related to other concepts in the ontology. An important test for any shared ontology is whether the data from the data sources can be inserted in the relational database schema based on the shared ontology. Critical success factors in our approach of ontology development are the commitment of participants to the process and the presence of one or more knowledge engineers. Knowledge engineers are impartial scientists who can pro-actively identify and discuss open issues to find agreement, and who do not push their own opinion on the content of the shared ontology.

The database holds data that are spatially and temporally consistent and this difficult task of integration of different data sources has been done by specialists instead of scientists working on indicators or models with poor knowledge on the different data sources, which is an important advantage of the integrated database. Also, users of the data only have to retrieve data from one source instead of different sources. A disadvantage of having one integrated database from the data provider point-of-view is that the data provider has to maintain and oversee a large database with data from different domains instead of a small database requiring knowledge from one domain. This implies that data management needs to be done by more than one person and different data-providers need to interact closely for maintenance, support and extension.

The use of Semantic-Rich Development Architecture (SeRiDA; Section 3.2) for traversing across programming paradigms (relational databases, object-oriented programming and ontologies) allows the programmers to benefit from the strengths of each of programming paradigms, and not having to maintain the same conceptual schema in at least two places (the database schema and the data accessing codes). In SEAMLESS, we adopted an explicit process to specify an upper data structure (as an ontology), that was translated through SeRiDA into a database schema and the appropriate source code for retrieving and storing data. This allowed the domain scientists to focus on the actual challenge of domain modelling, instead of details of technical implementation in different programming paradigms. Finally, the database is running as a central repository that supports access rights, ensuring safety and consistency.

6. Conclusions

The integrated database on European agricultural systems can support policy evaluation and assessment through providing indicators and model inputs for integrated assessment. The integrated database contains data on cropping patterns, production, farm structural data, soil and climate conditions, current agricultural management and policy information and can be extended with more datasets. The database has been used by the models available in the SEAMLESS project, i.e., a dynamic cropping systems model, a bio-economic farm model, an econometric model and an agricultural sector model and can be linked to other models or indicators as required. Data on agricultural management throughout Europe are absent, but essential for the database and exploiting the modeling capabilities of SEAMLESS. The data on current agricultural management is only available for 16 regions in Europe due to time and budget constraints in the collection of data. A systematic and institutional arrangement at European level is needed to complete and to regularly maintain this data set.

The database has the advantages that (i) several data sources are available on one server; (ii) difficult questions of data integration and consistency have been solved by specialists familiar with the original data sources and (iii) the pre-processing of the original data sources is already done. We aim to make the database available for non-commercial use.

The integration of different data sources into one database is a difficult and time consuming task (Gruber, 1993; Holsapple and Joshi, 2002), as we experienced in our collaborative process to derive one shared ontology. Such a collaborative and time-consuming process of ontology development is required to derive a schema that integrates a range of data sources from different domains specified at different spatial and temporal scales and to avoid inconsistencies and ambiguities in the meaning and definition of concepts across data sources. The explicit and iterative process of ontology development forced us to focus on the domain knowledge and the consistent and coherent linkage of the different data sources. This process could be potentially useful for extending the database on European agricultural systems with more data sources or to integrate other data sources.

We anticipate the database to be of interest for information specialists and systems analysts in the agri-environmental domain. They can derive or calculate policy relevant information. The paper also described the method to arrive at an integrated database, which we think can be transferred to attempts in other projects and domains.

Acknowledgements

We thank all scientists in the SEAMLESS project who contributed to development of the shared ontology. This work has been carried out as part of the SEAMLESS Integrated Project, EU sixth Framework Programme, Contract No. 010036-2. We gratefully acknowledge access to the multiple data sources by EU-FADN-DG AGRI-G3 for FADN data, JRC for JRC/AGRIFISH Data Base, European Soil Bureau for European Soil Database, University of Bonn for COCO/CAPREG and SEAMLESS partners for agricultural management data.

References

AbouZahr, C., Adjei, S., Kanchanachitra, C., 2007. From data to policy: good practices and cautionary tales. The Lancet 369 (9566), 1039–1046.

Ali, R., Wheitner, D., Talbott, E., Zborowski, J., 2007. Connecting environmental health data to people and policy: integrating information and mobilizing communities for environmental public health tracking. Journal of Community Health 32 (5), 357–374.

Andersen, E., Elbersen, B., Godeschalk, F., Verhoog, D., 2007a. Farm management indicators and farm typologies as a basis for assessments in a changing policy environment. Journal of Environmental Management 82 (3), 353.

Andersen, E., Elbersen, B., Godeschalk, F., Verhoog, D., 2007b. New policy objectives—new farm typologies. In: Poppe, K., Boone, K., Teeuwen-Vogelaar, C. (Eds.), Changes in Farming and the Effects on FADNs. Agricultural Economic Institute (LEI), The Hague, pp. 111–121.

Andersen, E., Elbersen, B.S., Hazeu, G., Van Diepen, C., Baruth, B., Verhoog, D., Terluin, I., Borkowski, N., Janssen, S., 2007c. The environmental component, the farming systems component and the socio-economic component of the SEAMLESS database for the Prototype 2. D4.3.5-D4.4-D4.5.4., SEAMLESS integrated project, EU 6th Framework Programme, (contract no. 010036-2), www.seamlessassociation.org, p. 195.

Antoniou, G., Van Harmelen, F., 2004. A Semantic Web Primer. The MIT Press, Cambridge, Massachusetts; London, England, p. 238.

Athanasiadis, I.N., Janssen, S., 2008. Semantic mediation for environmental model components integration. Information Technologies in Environmental Engineering 1, 3–11.

Athanasiadis, I.N., Villa, F., Rizzoli, A.E., 2007a. Enabling knowledge-based software engineering through semantic-object-relational mappings. In: Stojanovic, L., Sabbouh, M. (Eds.), 3rd International Workshop on Semantic Web Enabled Software Engineering, 4th European Semantic Web Conference, Innsbruck, Austria, p. 15.

Athanasiadis, I.N., Villa, F., Rizzoli, A.E., 2007b. Ontologies, JavaBeans and relational databases for enabling semantic programming. In: 31st IEEE Annual International Computer Software and Applications Conference (COMPSAC), Beijing, China, pp. 341–346.

Baruth, B., Genovese, G., Montanarella, L., 2006. New soil information for the MARS Crop Yield Forecasting System. Joint Research Centre of European Commission, Ispra, http://eusoils.jrc.it/projects/sinfo/pdf/mars_crop_yield_forecasting_system.pdf, p. 95.

Berners Lee, T., Hall, W., Hendler, J., Shadbolt, N., Weitzner, D.J., 2006. Creating a science of the web. Science 313, 769–771.

Borkowski, N., Zander, P., Stokstad, G., Uthes, S., Reinhardt, F.-J., Bellocchi, G., Hecker, M., 2007. Conceptual approach to identify and assess current activities. PD3.3.9, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.seamlessassociation.org, Wageningen, p. 23.

Britz, W., Pérez, I., Zimmermann, A., Heckelei, T., 2007. Definition of the CAPRI core modelling system and interfaces with other components of SEAMLESS-IF. SEAMLESS Report No. 26, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.seamlessassociation.org, p. 116.

Cohen, S.J., 1997. Scientist–stakeholder collaboration in integrated assessment of climate change: lessons from a case study of Northwest Canada. Environmental Modelling & Assessment 2 (4), 281–293.

Donatelli, M., Russell, G., Rizzoli, A.E., Acutis, M., Adam, M., Athanasiadis, I., Balderacchi, M., Bechini, L., Belhouchette, H., Bellocchi, G., Bergez, J.E., Botta, M., Braudeau, E., Bregaglio, S., Carlini, L., Casellas, E., Celette, F., Ceotto, E., Charron-Moirez, M.E., Confalonieri, R., Corbeels, M., Criscuolo, L., Cruz, P., Guardo, A.D., Ditto, D., Dupraz, C., Duru, M., Fiorani, D., Gentile, A., Ewert, F., Gary, C., Habyarimana, E., Jouany, C., Kansou, K., Knapen, M.J.R., Filippi, G.L., Leffelaar, P., Manici, L., Martin, G., Martin, P., Meuter, E.C., Mugueta, N., Mulia, R., Noordwijk, M.V., Oomen, R., Rosenmund, A., Rossi, V., Salinari, F., Serrano, A., Sorce, A., Vincent, G., Theau, J.P., Therond, O., Trevisan, M., Trevisiol, P., Van Evert, F.K., Wallach, D., Wery, J., Zerourou, A., 2009. APES: The Agricultural Production and Externalities Simulator. In: Brouwer, F., Van Ittersum, M.K. (Eds.), Environmental and agricultural modelling: integrated approaches for policy impact assessment. Springer Academic Publishing.

Dramstad, W., Sogge, C., 2003. Agricultural impacts on landscapes: developing indicators for policy analysis. In: Proceedings from an OECD Expert Meeting, Oslo, Norway, October 2002. NIJOS/OECD, Oslo, p. 343.

EC, 2003. Council regulation (EC) No 1782/2003 of 29 September 2003. European Commission, No1782/2003, p. 111.

EC, 2006. Rural Development 2007–2013, Handbook on common monitoring and evaluation framework, Guidance document. European Commission, Directorate General for Agriculture and Rural Development, Brussels, p. 15.

EC, 2008a. Farm Accountancy Data Network (FADN) Source: EU-FADN-DG AGRI-G3. http://ec.europa.eu/agriculture/rica/.

EC, 2008b. Nomenclature of territorial units for statistics—NUTS Statistical Regions of Europe. http://ec.europa.eu/comm/eurostat/ramon/nuts/home_regions_en.html.

EEA, 2005. Agriculture and environment in EU-15—the IRENA indicator report. European Environment Agency, Copenhagen, p. 122.

El Gohary, N., El Diraby, T., 2005. Achieving ontology interoperability using formal concept analysis: an approach to inter-organizational collaboration. In: Scherer, R.J., Katranuschkov, P., Schapke, S.-E. (Eds.), Proceedings of the 22nd CIB-W78 Conference on Information Technology in Construction, Dresden, pp. 1–8.

Elbersen, B., Kempen, M., Van Diepen, K., Hazeu, G., Verhoog, D., Andersen, E., 2006. Protocols for spatial allocation of farm types. SEAMLESS project deliverable (PD 4.7.1). SEAMLESS Integrated Project EU FP6 (contract nr. 010036), Wageningen.

ESBN, 2008. European Soil Database. http://eusoils.jrc.it/esbn/Esbn_overview.html.

Eurostat, 2008. European Dissemination Database (Formerly New Cronos). http://epp.eurostat.ec.europa.eu/.

Ewert, F., Van Ittersum, M.K., Bezlepkina, I., Therond, O., Andersen, E., Belhouchette, H., Bockstaller, C., Brouwer, F., Heckelei, T., Janssen, S., Knapen, R., Kuiper, M., Louhichi, K., Olsson, J.A., Turpin, N., Wery, J., Wien, J.E., Wolf, J., 2009. A methodology for integrated assessment of policy impacts in agriculture. Environmental Science & Policy 12 (5), 546–561.

Gallopin, G.C., 1997. Indicators and their use: information for decision-making. In: Billhartz, S.B., Matravers, R. (Eds.), Sustainability Indicators: A Report on the Project on Indicators of Sustainable Development. John Wiley and Sons, Chichester, pp. 13–27.

GeoServer, 2008. GeoServer http://geoserver.org.

Gobin, A., Jones, R., Kirkby, M., Campling, P., Govers, G., Kosmas, C., Gentile, A.R., 2004. Indicators for pan-European assessment and monitoring of soil erosion by water. Environmental Science & Policy 7 (1), 25–38.

Gruber, T.R., 1993. A translation approach to portable ontology specifications. Knowledge Acquisition 5, 199–220.

Hazeu, G., Elbersen, B., Andersen, E., Baruth, B., Van Diepen, C.A., Metzger, M.J., 2009. A biophysical typology for a spatially-explicit agri-environmental modeling framework. In: Brouwer, F., Van Ittersum, M.K. (Eds.), Environmental and agricultural modelling: integrated approaches for policy impact assessment. Springer Academic Publishing.

Heckelei, T., Britz, W., 2001. Concept and explorative application of an EU-wide regional agricultural sector model (CAPRI-Project). In: Heckelei, T., Witzke, H.P., Henrichsmeyer, W. (Eds.), 65th EAAE Seminar Vauk Verlag, Kiel, Germany, pp. 281–290.

Herrero, M., Gonzalez-Estrada, E., Thornton, P.K., Quiros, C., Waithaka, M.M., Ruiz, R., Hoogenboom, G., 2007. IMPACT: generic household-level databases and diagnostics tools for integrated crop-livestock systems analysis. Agricultural Systems 92 (1–3), 240–265.

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

INSPIRE, 2008. INSPIRE. http://inspire.jrc.it.

ISO, 2008. ISO/tc211. http://www.isotc211.org.

Janssen, S., Van Ittersum, M.K., 2007. Assessing farm innovations and responses to policies: a review of bio-economic farm models. Agricultural Systems 94 (3), 622–636.

JBOSS, 2008. Hibernate. www.hibernate.org.

Jones, R.J.A., Hiederer, R., Rusco, E., Loveland, P.J., Montanarella, L., 2005. Estimating organic carbon in the soils of Europe for policy support. European Journal of Soil Science 56, 655–671.

JRC, 2008. Meteorological data Source JRC/AGRIFISH Data Base – EC – JRC. http://agrifish.jrc.it/marsstat/datadistribution/.

Knublauch, H., 2005. Protege OWL. http://protege.stanford.edu/.

Louhichi, K., Janssen, S., Li, H., Borkowski, N., Kanellopoulos, A., Ittersum, M.V., Flichman, G., Hengsdijk, H., Zander, P., Blanco, M., Stokstad, G., Athanasiadis, I., Rizzoli, A.E., Huber, D., 2009. A generic Farming System Simulator (FSSIM); an application for modelling European arable farming. In: Brouwer, F., Van Ittersum, M.K. (Eds.), Environmental and agricultural modelling: integrated approaches for policy impact assessment. Springer Academic Publishing.

McGuinness, D., Van Harmelen, F., 2004. OWL Web Ontology Language Overview. www.w3.org/TR/owl-features/.

Metzger, M.J., Bunce, R.G.H., Jongman, R.H.G., Mucher, C.A., Watkins, J.W., 2005. A climatic stratification of the environment of Europe. Global Ecology and Biogeography 14 (6), 549–563.

Niemeijer, D., 2002. Developing indicators for environmental policy: data-driven and theory-driven approaches examined by example. Environmental Science & Policy 5 (2), 91–103.

Parker, P., Letcher, R., Jakeman, A., Beck, M.B., Harris, G., Argent, R.M., Hare, M., Pahl-Wostl, C., Voinov, A., Janssen, M., Sullivan, P., Scoccimarro, M., Friend, A., Sonnenshein, M., Barker, D., Matejicek, L., Odulaja, D., Deadman, P., Lim, K., Larocque, G., Tarikhi, P., Fletcher, C., Put, A., Maxwell, T., Charles, A., Breeze, H., Nakatani, N., Mudgal, S., Naito, W., Osidele, O., Eriksson, I., Kautsky, U., Kautsky, E., Naeslund, B., Kumblad, L., Park, R., Maltagliati, S., Girardin, P., Rizzoli, A., Mauriello, D., Hoch, R., Pelletier, D., Reilly, J., Olafsdottir, R., Bin, S., 2002. Progress in integrated assessment and modelling. Environmental Modelling & Software 17 (3), 209–217.

Pérez Domínguez, I., Bezlepkina, I., Heckelei, T., Romstad, E., Oude Lansink, A., Kanellopoulos, A., this issue. Linking farm and market models by means of response functions. Environmental Science & Policy.

Pervanchon, F., Bockstaller, C., Girardin, P., 2002. Assessment of energy use in arable farming systems by means of an agro-ecological indicator: the energy indicator. Agricultural Systems 72 (2), 149–172.

PostGIS, 2008. PostGIS. http://postgis.refractions.net/.

PostgreSQL, 2008. PostgreSQL. www.postgreSQL.org.

Refsgaard, J.C., Nilsson, B., Brown, J., Klauer, B., Moore, R., Bech, T., Vurro, M., Blind, M., Castilla, G., Tsanis, I., Biza, P., 2005. Harmonised techniques and representative river basin data for assessment and use of uncertainty information in integrated water management (HarmoniRiB). Environmental Science & Policy 8 (3), 267–277.

Reus, J., Leendertse, P., Bockstaller, C., Fomsgaard, I., Gutsche, V., Lewis, K., Nilsson, C., Pussemier, L., Trevisan, M., Van der Werf, H., Alfarroba, F., Blumel, S., Isart, J., McGrath, D., Seppala, T., 2002. Comparison and evaluation of eight pesticide environmental risk indicators developed in Europe and recommendations for future use. Agriculture, Ecosystems & Environment 90 (2), 177–187.
Turner, R.K., Ledoux, L., Cave, R., 2001. The use of scenarios in integrated environmental assessment of coastal-catchment zones. Report for the EUROCAT project. EC-contractor. EVK1-CT-2000-00044, School of Environmental Sciences, University of East Anglia, Norwich, http://www.cs.iiia.cnr.it/EUROCAT/publications/EUROCAT%20WD01.pdf, p. 25.
Van der Goot, E., 1997. Technical Description of Interpolation and Processing of Meteorological Data in CGMS. Joint Research Centre of the European Commission, Ispra, Italy, p. 23.
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., Ewert, F., Heckelei, T., Wery, J., Alkan Olsson, J., Andersen, E., Bezlepkina, I., Brouwer, F., Donatelli, M., Flichman, G., Olsson, L., Rizzoli, A., Van der Wal, T., Wien, J.-E., Wolf, J., 2008. Integrated assessment of agricultural systems—a component based framework for the European Union (SEAMLESS). Agricultural Systems 96, 150–165.
Villa, F., Ceroni, M., Krivov, S., 2007. Intelligent databases assist transparent and sound economic valuation of ecosystem services. Environmental Management 39, 887–899.
Weyant, J., Davidson, H., Dowlatabadi, H., Edmonds, J., Grubb, M., Richels, R., Rotmans, J., Shukla, P., Tol, R.S.J., Cline, W., Fankhauser, S., 1996. Integrated Assessment of climate change: an overview and comparison of approaches and results. In: Bruce, J.P., Lee, H., Haites, E.F. (Eds.), Climate Change 1995—Economic and Social Dimensions. Cambridge University Press, Cambridge, pp. 367–396.

Sander Janssen after finishing his master degree at Wageningen University, Sander Janssen is currently undertaking his Ph.D. again at Wageningen University, as a joint project between the Business Economics group and the Plant Production Systems group. His Ph.D. concerns integrated assessments, integration between disciplines and bio-economic farm models. He has a particular interest for research that transcends disciplinary boundaries and thus achieves innovative outcomes or insights.

Erling Andersen holds a position as senior researcher in the field of agriculture and countryside planning at Forest & Landscape, University of Copenhagen. He has a Ph.D. in countryside planning and a master degree in Geography and Communications. His research fields are in agricultural policies and the environment, agri-environmental policies, agri-environmental indicators, countryside planning and environmental impact assessment. Most recent research is on farm typologies for integrated analyses and on a database integrating data on European farming, biophysical endowments and socio-economics.

Ioannis N. Athanasiadis is a research scientist with the Dalle Molle Institute for Artificial Intelligence (Istituto Dalle Molle di Studi sull'Intelligenza Artificiale), in Lugano, Switzerland. His areas of expertise include environmental informatics, software agents, decision support systems and machine learning. He hold a Dipl.Eng. and a Ph.D. in Electrical and Computer Engineering, both from the Aristotle University of Thessaloniki. During the period 2000–2005 he was a associate researcher with the Informatics and Telematics Institute, CERTH, in Thessaloniki, Greece and a founding member of the Intelligent Systems and Software Engineering Lab.

Martin K. van Ittersum is Associate Professor within the Plant Production Systems group at Wageningen University, The Netherlands. His research focus is on development of model-based methods for the assessment of current and future agricultural production systems, both at farm and regional scale. He has run a range of national and international research projects in Europe, Central America and South-east Asia.