In Environmental Modelling and Software, **24** (**5**):577-587, 2009.

Environmental Modelling & Software 24 (2009) 577–587

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Review

Modelling with knowledge: A review of emerging semantic approaches to environmental modelling

Ferdinando Villa[a,*], Ioannis N. Athanasiadis[b], Andrea Emilio Rizzoli[b]

[a] Ecoinformatics Collaboratory, Gund Institute for Ecological Economics and Department of Plant Biology, University of Vermont, 617 Main Street, Burlington, VT, USA
[b] Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Lugano, Switzerland

ARTICLE INFO

Article history:
Received 24 August 2007
Received in revised form 2 September 2008
Accepted 6 September 2008
Available online 11 December 2008

Keywords:
Semantic modelling
Ontologies
Semantic annotation
Model and data integration
Conceptual design
Model-based query

ABSTRACT

Models, and to a lesser extent datasets, embody sophisticated statements of environmental knowledge. Yet, the knowledge they incorporate is rarely self-contained enough for them to be understood and used – by humans or machines – without the modeller's mediation. This severely limits the options in reusing environmental models and connecting them to datasets or other models. The notion of "declarative modelling" has been suggested as a remedy to help design, communicate, share and integrate models. Yet, not all these objectives have been achieved by declarative modelling in its current implementations.

Semantically aware environmental modelling is a way of designing, implementing and deploying environmental datasets and models based on the independent, standardized formalization of the underlying environmental science. It can be seen as the result of merging the rationale of declarative modelling with modern knowledge representation theory, through the mediation of the integrative vision of a Semantic Web. In this paper, we review the present and preview the future of semantic modelling in environmental science: from the mediation approach, where formal knowledge is the key to automatic integration of datasets, models and analytical pipelines, to the knowledge-driven approach, where the knowledge is the key not only to integration, but also to overcoming scale and paradigm differences and to novel potentials for model design and automated knowledge discovery.

1. Introduction

Many environmental models relate to the structure of the systems they represent only in partial, hardly understandable ways. In fact, very few models can serve as an explanation of the modeled processes: the understanding of the system is usually implicit and typically resides outside the model specification and implementation. As a result, the purpose of models is typically restricted to the specific application they have been developed for; the potential for reuse, communication and integration with data and other models is limited.

Inspired by the declarative programming approaches that became popular in computer science in the 1980s, *declarative modelling* has been suggested as a remedy for the "black box" nature of self-contained models (Robertson et al., 1991; Wenzel, 1992; Keller and Dungan, 1999; Muetzelfeldt, 2004). Declarative models describe environmental processes in the form of concise mathematical statements that bear a closer resemblance to the way environmental scientists conceptualize systems. Declarative modelling approaches have been incorporated in graphical languages that can produce readable and fairly self-explanatory model statements, e.g. Simile (Muetzelfeldt and Massheder, 2003) and STELLA (Richmond, 2001), or in specialized conventional languages such as Modelica (Tiller, 2001). Such approaches have become popular enough with environmental scientists to make libraries of reusable model components conceivable (Salles et al., 2006). Yet, truly modular model repositories, capable of meeting large-scale integration goals, have largely remained wishful thinking. One reason is that declarative approaches to modelling, while greatly enhancing readability of model components, have mostly focused on syntactic aspects rather than semantics: declarative models use convenient abstractions that are relevant to the modelling process (such as "stocks" and "flows") but differentiate processes and environmental entities merely by means of variable names, and remain incapable of incorporating any formal statement about what those names mean. It is close to impossible to ensure that the requirements of two models match with sufficient precision to allow their merging without "conceptual" error, if the semantics is not fully explicit. Model integration in software terms does not guarantee sound integration of the model logics (Athanasiadis et al., in press), and declarative modelling is no exception on this principle. While declarative modelling makes

* Corresponding author.
E-mail addresses: ferdinando.villa@uvm.edu (F. Villa), ioannis@athanasiadis.info (I.N. Athanasiadis), andrea@idsia.ch (A.E. Rizzoli).

1364-8152/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2008.09.009

equations readable and easy to simulate within appropriate computing environments, the environmental knowledge itself has not been made explicit to the full extent.

In recent years, large-scale visions such as the Semantic Web (Lee et al., 2001) have started to investigate the problem of communicating shared resources (such as web documents) in ways that make meanings explicit and meaningful, automatic associations possible. Fueled by such developments, attention to the semantic nature of shared data and resources has begun to percolate down to the natural sciences. *Semantic annotation* of datasets and models (Kiryakov et al., 2003; Athanasiadis, 2006, 2007; Parr et al., 2006; Khatri et al., 2006; Rizzoli et al., 2008; Villa, 2007; Villa et al., 2007; Chen et al., 2007; Lee et al., 2007; Madin et al., 2007) is now a recognized topic in modelling in general, and in environmental modelling in particular. The rationale of such developments is easing integration and reusability of environmental datasets, models and processing dataflows. Yet, many of the potential benefits of a semantically explicit environmental modelling remain unexplored.

This paper discusses and previews the road ahead of conventional declarative modelling. We discuss the current approaches to environmental modelling that exploit the formalized semantics of natural systems to unify representations of data and metadata, improve their usability in scientific workflows, and ease the definition of dynamic models. The approaches we discuss link the notions of declarative modelling, metadata, scientific workflow, data integration and environmental databases in a synthesis of the most recent advances in ecoinformatics and knowledge representation.

Because the key tool to allow this unification is the use of structured knowledge (organized in *ontologies*) to inform data and metadata compilation, model conceptualization, and simulation, we start with a brief introduction to ontologies geared to environmental modelling applications. In Section 3 we shall discuss the ways in which formal knowledge is being incorporated in environmental data and model design, and how environmental sciences are moving towards more formal statements of meaning as an integral part of the activity of modelling. In Section 4 two approaches to semantic modelling are presented. The "mediation approach" where semantically enriched datasets and declarative models facilitate integration and reuse will be discussed first. The knowledge-driven modelling approach, discussed next, makes the knowledge explicit and the model logics implicit, only requiring a declaration of the modeled system; we discuss how a simulation strategy can be inferred from it by means of machine reasoning, and how the declaration of a model is merely an extension of that of a dataset. In the last section (Section 5), we shall discuss novel scenarios and scientific approaches that are made possible by the adoption of knowledge-based perspectives in environmental modelling, and the corresponding remaining challenges and uncertainties.

2. Formalizing knowledge: ontologies

The term *ontology* originates in philosophy, and refers to the study of being, but nowadays it is also used more widely and with eminently practical meanings. In artificial intelligence, an ontology is any formal description of a conceptualization of a domain of interest (Gruber, 1993, 1995). In computer science, an ontology is a formal data structure that describes a conceptual domain, usually consisting of a set of statements (axioms) that define *concepts* and *relationships* between concepts (Wand et al., 1999). Languages, formalisms, and tools to create, store and communicate ontologies have proliferated in recent years (e.g. OWL (Guinness and van Harmelen, 2004)). It is now common to identify an ontology with a web-accessible document that can be created, edited and validated using ad hoc tools. Ontologies are used as references to annotate resources with concepts in standardized ways, e.g. in RDF (Beckett, 2004).

In ontologies, a *concept* (or *class*) is the statement (definition) of an entity, usually including at least a textual description and a short label. Concept names (IDs) and descriptions do not define meanings of an entity formally: the actual meaning of concepts is always defined through properties, which relate concepts to one another. For example, the identifier "person" means little by itself; what makes a "person" identifiable as such is having a name, an address, relatives and other personal data. By making these properties explicit, ontologies allow an automated process to recognize the formal description of a person from that of a different object by checking its properties and the logical constraints that accompany them.

A *property* is the statement of an association between concepts. The need for mathematical and computational tractability of the resulting conceptualizations constrains most ontology frameworks to handling binary properties, where a concept is associated to one other only. *Relationships* are statements that link an instance to the value of a specified property, which can be a concept, an instance or a literal value such as a numeric or textual value. Ternary statement such as "John birth date 10–10–1972" can be used to state relationships, which must comply with the model specified by the concept and its properties. Properties may define a cardinality constraint, which limits the number of allowed values (e.g. a population must have at least one individual and only one species).

Properties allow building the bones of knowledge representation structures: notably, the subsumption property (which can be stated as "is-a", "kind-of", or "is-type-of") allows constructing generalization–specialization (classification) hierarchies. Such structures are often called *taxonomies* and allow specializing meanings into less and less general concepts. In a very simple example, the concept "Ecological assemblage" may be specialized into "Population" and "Community"; the latter may in turn be specialized into "Floral" and "Faunal" and so on. Note that the subsumption of concepts requires the inheritance of relationships: all properties of an ecological assemblage (such as being composed of one or more individual organisms) must also be properties of populations and communities, although their meaning may be restricted in specialized concepts to define narrower logical models (e.g. floral communities may only include individuals of plant species). Another common relationship in natural system ontologies is "part-of", used for example to guide collection and classification of plant specimens (Paterson et al., 2004). Properties can be generalized or specialized just like concepts: e.g. "*economic-dependency*" is-a "*dependency*".

Instances (also called Individuals or Objects) are the statement of a real-world entity and ultimately the incarnation of a concept defined in an ontology. Compared to a concept, an instance cannot be further specialized and may not have additional or restricted properties compared to those of the concept it incarnates. A collection of instances that conform to an explicit ontology is often called a *knowledge base*, although the term is not rigorous. The records of a database with a formally specified relational schema can always be considered a set of instances of the correspondent ontology; the database paradigm provides an interface and optimized methods for storage, search and retrieval of instances based on matching property values to user-defined constraints.

Ontologies can be used at different degrees of internal complexity and expressive power (Fig. 1). The simplest use is as *controlled vocabularies*, consisting of a loosely structured set of concepts with no properties, whose names define allowed terms for purposes of harmonization of terminology. This usage is becoming common in environmental applications (Baker et al., 2000; AGROVOC, 2006; Batzias and Siontorou, 2006; Duepmeier

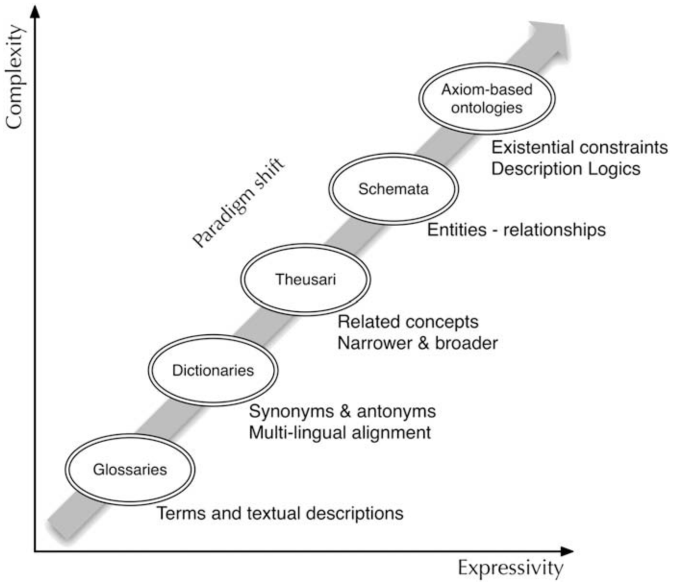

Fig. 1. A pictorial representation of common knowledge representation paradigms ordered by increasing expressivity and complexity.

and Geiger, 2006). Terms of a controlled vocabulary often include concept name translations in several languages. Yet, a useful ontology typically contains more than just a list of terms and their definitions. A step further is a taxonomy, where concepts are arranged in a generalization (is-a) hierarchy as seen above. Other common ontologies establish relationships like "broader/narrower/related" between concepts: such ontologies are often referred to as *thesauri*. The full power of ontologies is exploited when domain-specific properties are defined along with concept classes. Controlled vocabularies, thesauri and taxonomies can all be evolved into full ontologies by defining custom properties. Examples of full ontologies in the natural sciences are abundant. For example, the SWEET ontologies by NASA (Raskin and Pan, 2005; NASA, 2006) propose a formalization of common knowledge about the physical world and the biosphere, designed to facilitate coordination and communication of applications and agents in earth and space sciences.

These normative uses, where ontologies provide guidance for conceptualization, can be thought of as the generalization of formalisms that are commonly adopted in natural system information management and modelling, such as metadata standards (Jones et al., 2001; ISO, 2004), database schemata or object-oriented models. The main rationale for the existence of ontologies, however, is to enable *machine reasoning* on natural systems, either by human actors or by a computer program (reasoner). The mainstream ontology formalisms are carefully designed to only allow axioms that preserve decidability and therefore guarantee their usability for machine reasoning. The main operations in automated reasoning are subsumption (inferring that concept A is more specific that concept B) and classification (inferring that instance X is an incarnation of concept A). A common framework for reasoning, Description Logics (Baader et al., 2003) provides ways to state necessary and sufficient conditions for instances of an ontology to belong to concepts, so that logical inference (reasoning) can be performed on them to attribute concepts and instances to independently defined classes.

Reasoning is the key operation to enable sophisticated integration of datasets and models. For example, appropriately defined concepts may accompany variables in a model and columns in a tabular dataset. This in fact makes an instance out of the data. When that is done, datasets (or model outputs) can be deemed suitable to be used as the value of an input variable if and only if a reasoner can classify them as an instance of the same class as the variable's. In the following sections, we will describe applications, problems, and perspectives in applying ontologies and machine reasoning to different sectors of environmental modelling and management.

3. Knowledge models in ecology and environmental science

Ecological and environmental modelling activities are obviously related to knowledge representation and management. Natural systems knowledge is formalized in various ways, varying from scientific publications, taxonomic classifications, data collections, modelling theories, to simulation models and results. Ontologies are increasingly used to support natural system modelling and enhance its rigor and consistency. For example, Brilhante (2005) proposes *Ecolingua*, an ontology for ecological quantitative data that attempts to enable the synthesis of conceptual ecological models from data descriptions through the reuse of existing model structures; Ceccaroni et al. (2004) used an ontology to augment and improve the formalization of knowledge to be used in case-based and rule-based reasoning decision support system for wastewater management. Scholten et al. (2007) use ontologies to implement a methodology to support the development of multi-disciplinary models for water management, providing formal guidelines in the model development process. Currently, several projects are trying to address the need for a pool of shared concepts that natural system scientists can refer to for annotation of data and models. The SEEK project (SEEK, 2004) is eliciting ontologies from a large community of ecologists to enable semantic annotation of datasets and processing steps in scientific workflows (Williams et al., 2007). Similar efforts are underway in other disciplinary domains, e.g. in genomics (Ashburner et al., 2000), atmospheric physics (McGuinness et al., 2007), water quality (Chau, 2007) and agriculture (AGROVOC, 2006). Not surprisingly, providing conceptual frameworks to express the vast realm of natural systems knowledge is far from trivial: natural systems sciences need to address the most complex domains and are still far from the general acceptance that much simpler, single-scaled conceptualizations, such as the Gene Ontology (Ashburner et al., 2000) can enjoy today.

In this section, we relate the vision of a knowledge-based approach to ecological and environmental modelling to the conventional notions of data-driven and declarative modelling. This discussion paves the way for that of explicit semantic modelling introduced in the next section.

3.1. Data-driven modelling

Datasets can be considered "static models" of systems: they always conform to conceptual models, and depend on (usually implicit) assumptions and world views just as much as dynamic models do (Villa and Costanza, 2000; Villa, 2001, 2007). Conventional data management practice separates "raw data" – actual numeric or categorical values – from "metadata", the information that allows a user to make sense of the numbers, providing needed spatial and temporal contexts, unit of measurement, method and accuracy information to complement the raw information.

In an ontology-informed framework, the starting point is always the accurate formalization of the concept that is quantified in the data, including the definition of its characteristics represented as properties. Whether the link to the ontology is made after the measurement (annotation) or beforehand, datasets can be seen as instances of an explicit, formal concept. Data are associated to concepts through one property: e.g. in a trivial case, a "*numeric-value*" property could link a concept to the raw data, represented as textual (literal) values. The remaining properties incorporate the "metadata", which in this case are defined by and connected to an explicit knowledge model. The benefits of this approach lie in the richness and logical consistency gained by using a formal ontology as the founding knowledge model for data and metadata. The latter are specified in forms that allow mediation, alignment and consistency checking across several sources; alignment between different knowledge models adopted can be performed by automatic reasoners if appropriate bridging information is supplied.

The ontology paradigm in its simplest incarnation offers a field-by-field substitution for the tabular and relational schemata conventionally used to represent data. Schema information contained in normative ontologies can for example be seen as an add-on to an existing database management system, implemented as a software "wrapper" for a relational database engine or other suitable technology (such as an XML database). In previous work, we have investigated how to automatically transcribe logical models expressed as ontologies into object-oriented models and relational schemata (Athanasiadis et al., 2007a, b).

In addition to that, ontologies offer a powerful synthetic way to specify both the data schema and the structure of the knowledge behind it. While a relational schema can be seen as expressing the structure (the "*what*") of the knowledge, ontologies also allow specifying the *how* and the *why* at the same time. The database paradigm that includes a rich knowledge model such as that specified by ontologies is often called *deductive* or *intelligent database* (Bertino et al., 2001), whose applications in the environmental domain are beginning to appear (Villa et al., 2007).

Many common metadata standards have been formalized as ontologies: e.g. the ISO 19115 standard for spatial information (ISO, 2004). Yet, these efforts typically stop at the level of controlled vocabularies: they capture syntax (names and information containment hierarchies) rather than meaning. Providing a set of ontologies to define the actual meaning of the metadata properties is a much more sophisticated activity, and lends itself to different interpretations that reflect the complexity of the domain, including the existence of different viewpoints that can sometimes be complementary and other times competing. The ontologies chosen ultimately depend on (and define) the applications; the choice becomes crucial in determining the scope of potential integration with other resources. *Semantic mediation* (Ludaescher et al., 2001) is the attempt to integrate data and models using sets of ontologies (see Section 3.2). Projects such as SEAMLESS (SEAMLESS, 2005), SEEK (Kepler, 2004; SEEK, 2004; Madin et al., 2007), ARIES (Villa et al., 2008) and IMA (Villa, 2001, 2007; Villa et al., 2007) are endeavoring to develop sets of ontologies that best fit large-scale application domains. Other projects have been relatively successful in more restricted domains, e.g. food web theory (Webs on the Web (URL)).

3.2. Systems theory and declarative modelling

Dynamic models, like data, always conform to a conceptualization, and no philosophical difference needs to exist between specifying data and models when this is done using ontologies (Villa, 2001, 2007). Any model that can be "declared" can also be specified as a set of instances of appropriate ontologies. The main difference between data and models is the increased conceptual richness necessary to describe how things are created, destroyed, and modified in time and space. This requires at least notions of linkage between concepts with causative or dependency relationships that are normally not necessary to specify data. It also requires developing ways to interpret this causality. The set of abstractions (concepts) that allows conceptualizing and expressing those cause–effect relationships and their results is the adopted modelling paradigm, of which there are many examples (e.g. ordinary differential equations, stock-and-flow, or individual-based). A modelling paradigm, like any consistent conceptualization, is amenable to being at least partially captured into the logical constructs of an ontology. In fact, most existing modelling software systems (e.g. STELLA: (Richmond, 2001); SIMILE: (Muetzelfeldt and Massheder, 2003)) conform to one implicit, usually simple, ontology, which defines their notion of entities familiar to the user, such as state variables, flux variables, etc. Advanced integrative systems (Kepler, 2004; Villa, 2007) can manipulate different ontologies, which, supplemented by the necessary software, enable them to orchestrate models that adopt heterogeneous modelling paradigms. Such systems are in the best position to enable integration of independently developed models adopting different paradigms into a higher-level, multiple-paradigm model.

In *declarative modelling*, model specification is based on the attributes and semantics of the natural systems, rather than on the algorithm (numerical integration procedure) that calculates their results. For example, the graphical equivalent of a simple predator–prey model (Fig. 2) shows how the algorithms that integrate the relative difference equations have been substituted by the statement of relevant variables and their mutual dependencies. Such declarative specifications use an implicit ontology of modelling entities that embodies a chosen paradigm (e.g. the stocks and flows that reinterpret state variables and rates of change), but the knowledge (meaning) about the environment must still be "inferred" from the names of the variables and the reconstruction of the process.

Ontologies can support declarative modelling by providing, at the same time, schemata for model declaration and meaning for these schemata. Instances of ontologies represent declaratively expressed models that refer to concepts laid out in the ontologies. Such declarations contain enough information to enable a software infrastructure to simulate the behavior of the systems represented over a user-defined temporal and spatial extent. Thanks to the rich meaning made possible by ontologies, a workflow environment can properly connect models to data, and feed quantities calculated by simulation to other models in the same environment.

4. Semantic environmental modelling

In semantic modelling, all concepts used to model a natural system are explicitly defined by ontologies. The semantics of a model is the intersection of two languages: one that describes the natural knowledge, with concepts such as population, individual or growth rate, and one that represents the modelling process, with concepts such as variable, stock or flow. The declaration of a model in a semantically enriched way, using ontologies, can be achieved by specifying:

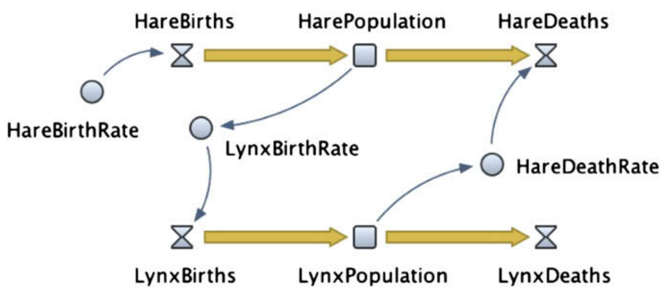

Fig. 2. Simple predator–prey system modeled according to common conventions of graphical stock-and-flow languages. Squares represent state variables, circles represent parameters and functions, and the remaining symbols represent rates. Influence arrows represent dependence of a rate or variable value on the current value of others. The ecological knowledge is represented implicitly in the variable names and model structure.

a. the modeled entities, by identifying the relevant concepts and properties and creating instances of them for each model component;
b. the underlying relationships among these entities, capturing the structure of causality in the system as understood by the modeller.

Software implementations can automatically check that a semantically enriched model specification is self-consistent, and with appropriate support information, can go as far as ensuring matching contexts of both space and time for all entities before the model can be accepted or calculated. This preliminary semantic validation prevents users from defining inconsistent models and helps to retrieve compatible data sources from databases when the model is applied. This kind of inference can be used to delegate to an ontology-aware software system much of the difficulties of properly designing and deploying a complex simulation model, while facilitating its application by non-scientists such as decision makers, and, at the same time, ensuring proper, correct use of both model and data and the soundness of the results. Ontologies are also valuable for adding "meta-information" related to both model equations and variables, maximizing the potential for reuse. This can be achieved by decoupling model interface elements and equations, which allows specifying model interfaces elements (variables or parameters) as measurements (Athanasiadis et al., 2006). This way a model exposes as inputs not only a set of numeric values, but also a clear specification of dimensions, units and spatio-temporal context. Because semantic links among models can be logically verified, using ontologies in model design defines a clear path to specifying reusable model components (Athanasiadis, 2006).

Two distinct approaches to incorporating ontologies in model design can be identified. In the *mediation* approach, concepts from ontologies supplement conventional data and models to facilitate integration and reuse of the information. In the *knowledge-driven* approach, datasets and models are directly represented as instances of ontologies and embody a statement of the system conceptualization, enabling machine reasoning about the system structure that can lead to more sophisticated applications. We describe both approaches in the rest of this section.

4.1. The mediation approach: data integration in analytical workflows

The mediation approach for environmental knowledge representation consists of the enrichment of existing data and environmental models with formal semantics to connect measurements and variables to the identity of the observable entities they quantify. This approach is gaining ground as a mean to overcome significant obstacles to the reuse of available computerized legacy models or archived data sources. As discussed earlier, the main reason for this situation is the typically poor semantics for representing environmental information and the absence of a standardized framework to properly captures the represented knowledge in the forms of annotations. In such a situation, managing and processing environmental information require ever-increasing efforts. *Scientific workflows* define information flows among existing simulation models and data sources, and typically require intensive pre-processing activities for data manipulation, scaling, formatting, and so on. Without appropriate semantic information, the knowledge that makes such tasks possible must be supplied externally, at the cost of considerable human effort, only little of which can be reused in a different context.

A semantically aware framework using the mediation approach is typically data-centric, as it utilizes ontologies for describing the representational contexts of environmental data, either residing in databases or produced by models. From a representational point of view both models and databases can be considered as "*data nodes*" that provide environmental data, operating either as *data sources* (databases, files, datasets, etc.) or as *data converters* that include both environmental models and pre- and post-processing algorithms (such as statistical data reduction or calculation of indicators from model results). The only difference between the two kinds is that in the first case data are permanently archived and are available via querying, while in the second are dynamically produced, and are made available via computation.

Defining the representational context of knowledge captured is an important activity for annotating properly data sources and data converters, and is critical for reusing them by constructing *scientific workflows* (Ludaescher et al., 2005; Athanasiadis, 2007). A scientific workflow is a pathway between two or more processing steps, along which data are transformed until a desired result is reached. Users assemble workflows, by connecting data nodes together and linking the results to storage or visualization facilities. A workflow environment is thus a "blackboard" for users to assemble the information flow needed to address a particular problem. In functional terms, a dynamic model can often be seen as a workflow, because in the end all models process input information and produce a set of output results. The similarity, however, does not hold when models and workflows are considered in semantic terms: the meaning of a model is not to produce outputs, but to describe a natural process. In fact, a more correct generalization sees workflows as special cases of models, whose "paradigm" entails data transfer and transformation along the connections of an artificial system.

In this sense, a workflow is amenable to the same ontology-based description as any other model: concepts of "input", "output", "processing step" can be specialized as needed using ontologies that can describe all steps of any workflow environment. The most important use of ontologies in workflow environments, however, is another: to allow the system to enforce meaningful, correct connections between inputs and outputs, and – if necessary and possible – insert transformation steps in the workflow that guarantee a proper match. The operation of enforcing and supporting semantic consistency along data paths in workflows is usually called *semantic mediation* (Kohler et al., 2000; Ludaescher et al., 2001) and it is essential to guaranteeing correct results (Athanasiadis, 2006; Villa, 2007), particularly when processing steps are heterogeneous and users are not domain experts. To allow semantic mediation, all inputs and outputs must come tagged with concepts from ontologies that are known to the workflow environment; the latter needs to use a reasoner program to ensure the consistency of concepts along each connection made by the user. The operation of associating real-world entities (e.g. input and output "ports" of workflow components) to concepts from ontologies is usually called *semantic annotation* (Kiryakov et al., 2003), and it is done by the same actors that have developed the models or processing steps. Conceptual compatibility of data nodes interfaces (inputs and outputs) is tested with a reasoning operation that checks if the data being fed from an output to an input represent the same concept. In a simple example, a model X is "packaged" as a workflow component and all its inputs are semantically annotated by its developer according to a set of commonly understood ontologies. The semantic annotation operation requires that all the conceptual details of each "port" (or interface element) are understood and appropriately defined. As an example, an input X_i representing temperature at the earth's surface may require that the temperature is expressed as monthly data over the simulated timespan, and the model has only been calibrated for temperatures in the 19–30 °C range, so it should not accept data outside of these boundaries. Semantic annotation is a way to express such conditions, which commonly are only expressed verbally in the model's documentation, in a formal and machine-readable way. In order to do so, an ontology is created to define model *X*, with concept defined for each exposed "port". Using restrictions and concepts from appropriately linked ontologies, the concept definition associated with input X_i may look similar to that shown in Text box 1.

When a semantically annotated model is used in a workflow, inputs and outputs are connected by the user. For example, a time series of temperature data retrieved from a database may be connected to input X_i. Upon connection, a semantically aware workflow environment can ensure the appropriate match between the input and the output by feeding the respective semantic annotations to a reasoner and ask if they describe the same concept (a classification operation). A reasoner can make the necessary inferences to deduce the equivalence of types that have different names, based on their properties. In some cases, a verdict of conceptual compatibility may be sent back to the processing environment to further check if a transformation needs to be inserted in the dataflow to make the input and output numerically compatible. A typical case is conversion of compatible numeric values that are expressed in different units of measurement. In a more complex example, requiring a higher sophistication of the processing environment, consider a data source of weekly data rather than monthly. A sophisticated workflow environment can understand that the data need to be aggregated into a monthly time scale before being passed to the model that uses them, and direct the workflow environment to create a transformation step to perform the aggregation and insert it between the data source and the model. Although decisions of this complexity are beyond the capabilities of first-order reasoning alone, software systems can be assisted by reasoners and ontologies to determine the proper operations to apply in order to mediate different, compatible contexts for the information exchanged in dataflows.

The mediation approach can be successfully applied to conventional declarative modelling. A dynamic system can be seen as described by a set of *state variables* whose numeric state changes under the influence of phenomena defined in instantaneous terms (*rates*) through differential equations. A commonly accepted declarative modelling interpretation of such systems, incorporated in several visual modelling environments such as Stella (Richmond, 2001) and Simile (Muetzelfeldt and Massheder, 2003), uses the abstractions of *stocks* and *flows* to refer to state variables and rates (Fig. 2). The use of these abstractions has become commonplace in ecological and environmental modelling, to the point that they constitute a good starting point for conceptualizing a modelling ontology that can be widely understood.

A declarative modelling language defining phenomena in terms of stocks and flows is highly enriched if these dynamic aspects can be related to the definition of the actual physical entities where they take place. So for example, instead of simply providing the stock and the two flows that describe change in population size, we can adopt a language that allows us to define the population entity, and include in its definition the fact that the "population numerosity" variable represents its state, and the phenomena of birth and death are the processes that influence it. By specifying the population in terms of its actual semantics (including, e.g. the property of being a population *of* hares) we greatly facilitate the process of connecting this model to existing data for initialization, and of ensuring that any connection of its output to other components in a workflow is meaningful. Model design is also greatly facilitated by the fact that the semantics of a population captured in the ontology can serve as a blueprint for the design of the model.

Text box 1. Semantic annotation of a required model input.

```
I :: =
   is-a: Temperature,
   vertically-distributed-in: PlanetarySurface,
   has_unit: Fahrenheit,
   max-value: (is-a: Temperature, has-value: 30.0, has-unit: Celsius)
   min-value: (is-a: Temperature, has-value: 19.0, has-unit: Celsius)
   distributed-in: (is-a: TimeSpan, step: 1, has-unit: Month).
```

4.2. Knowledge-driven modelling

If the mediation approach enriches an existing representation with formal knowledge, in knowledge-driven modelling the formal knowledge structure *is* the model. This is accomplished by defining environmental models (either static or dynamic) directly as instances of adequately expressive ontologies. Because different ontologies can be freely combined in a conceptualization, knowledge about the natural domain and the modelling paradigm can be pulled together to completely define a natural system, pairing knowledge about model dynamics with the definition of the relevant ecological entities. When the model is stated as instances in a knowledge-based framework, a reasoner-supported system can be employed to render models declaratively and pass them to an execution environment for simulation. A model is at this point no longer just an annotated *data node*, semantically equivalent to a *data source*; at the same time, it does not operate as a black box whose interface only (inputs and outputs) is documented, as in the mediation approach. On the contrary, a model expressed entirely using ontologies becomes a statement of the logics that constitutes the understanding of the modeled system, and captures the processes involved and its interpretation on behalf of the modeller in full semantic detail.

To explain how this approach can be realized, we use the common "hare and lynx" predator–prey system. Typically, a predator–prey system evokes a nonlinear differential equation system or a discrete "stock and flow" model in the mind of most ecologists. Figs. 3 and 4 explore this model in a knowledge-based framework: in Fig. 3, the "identity" of the system is captured in a particular instant of time, while in Fig. 4 enough knowledge is added to enable a suitable software system to infer a dynamical model by using machine reasoning alone.

In both cases, the "hare-lynx" system uses concepts from biodiversity ontologies that describe populations and communities. In the same way, it can be coupled with a taxonomy ontology that allows defining species unambiguously by referring to species identifiers from known repository services such as that provided by GBIF (GBIF, 2004). Using taxonomic resolution services can be key to integrating models with species data from remote datasets that have been annotated with the same conventions.

The knowledge model in Fig. 3 is essentially a semantically rich dataset: it states that at a certain time two populations coexist in a community, and have specified numerical abundances. Yet, it's much richer than a typical dataset with conventional metadata, or a conventional model component with the associated documentation, because of the extra knowledge layer expressed by linking to ontologies. The system as defined this way is essentially the non-dynamic model of a two state variables system.

The switch between a static and a dynamic model is operated by providing information on the causal connections among the different entities, and on how these connections influence the evolution of state variables through time. Causality in conventional models is usually expressed through equations associated with the value of variables. Equations, by referring to the values of other

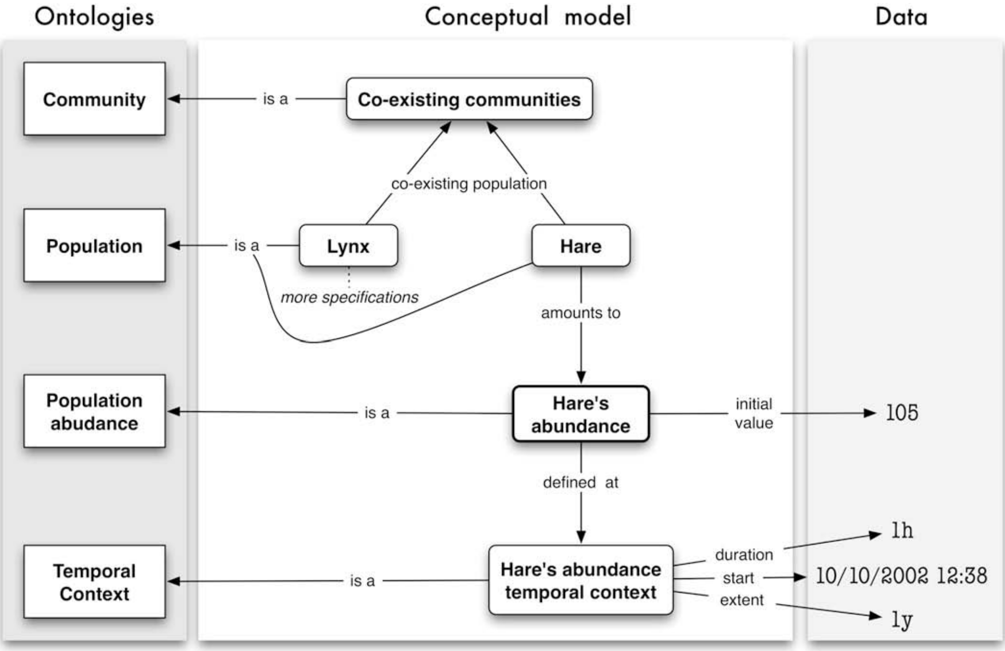

Fig. 3. Portions of a possible conceptualization of a predator–prey community. The system is conceptualized in a static way, meaning that the temporal variability is captured as multiple data points and no attempt is made to describe the causal factors that make the abundances vary with time.

variables, implicitly define causal relationships that correspond to computational dependencies. An ontology-based framework can make these dependency relationships explicit, and add semantics to them by further specializing these dependencies. For example, a generic *depends-on* relationship can be specialized into a *flows-into* relationship between a state variable and a flux variable (rate). The presence of this relationship in the knowledge-rich definition of a model informs the underlying software architecture that the flux must be integrated over time in order to assess its contribution to the value of the state variable. In other words, equations in a model definition can be seen as statements of specialized dependency relationships. The notion of variable, so central to conventional approaches, can similarly be enriched and made dependent on the modeled entity. For example, in an individual-based paradigm, variables describe quantitative traits of modeled individuals, but maintain the link to the individual which is the main entity considered. No conflicts need exist between paradigms, whose conceptual boundaries often become blurred when a explicit knowledge-based approach is used, particularly if notions of scale are formally defined (Villa, 2007).

The system conceptualization sketched in Fig. 4 adds enough information to the system of Fig. 3 to define how changes in time are enacted. To this purpose, the specification of the population abundances is extended to inform the system of how they change in time. The abundance of the lynx population is now not only a numeric value exposed in the model interface, but it is also made

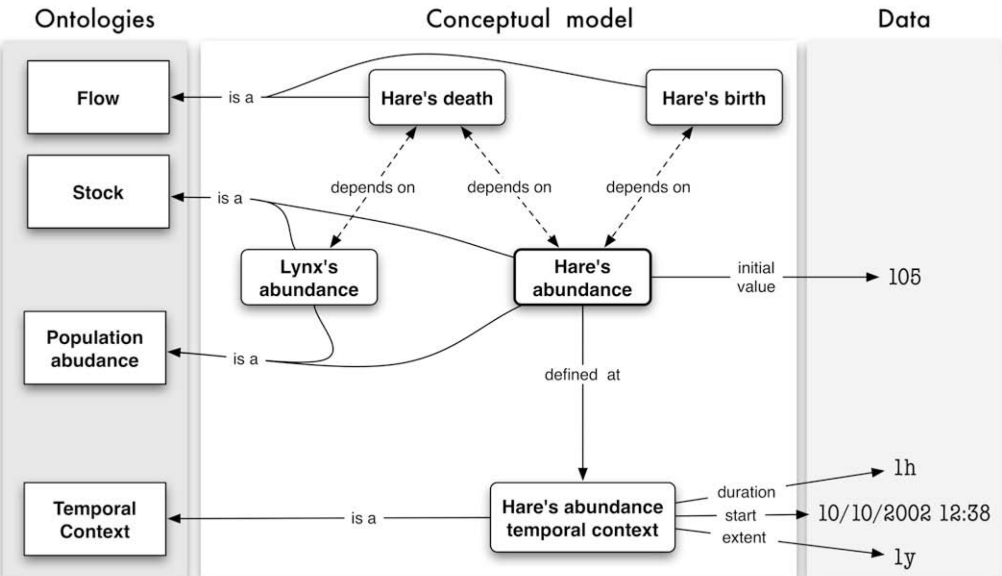

Fig. 4. The same system as in Fig. 3, with the addition of causal relationships from a modelling ontology. Such details allow an appropriate software system to infer a simulation workflow for the system.

a *stock*, by defining it as an instance of a "stock" concept from a modelling ontology that has been added to the conceptualization. Similarly, its time specification declares that while its current value is an hourly measurement equal to the initial value of the system in Fig. 3, there is an *extent* now defined, that expresses the fact that such hourly values will exist over the span of one year, determining a multiplicity of states for the system. Corresponding flows are added and linked to the existing model to express the factors that influence the change in the stocks. A knowledge representation framework for environmental modelling can analyze such declaration and decide that in order to know the numeric state of the system over the given extent, the stock and flow identities require difference equations to be defined, and the initial state must be extended over the time extent by integrating the flows.

There are limits to the declaration of models in current ontology frameworks, particularly if reasoning capabilities must be preserved. These limitations stem from the fact that any dependency other than linear requires higher-order logics statements to be fully formalized, and the handling of higher-order logics is beyond the capabilities of existing reasoning systems. The dominant paradigm, Description Logics (Baader et al., 2003) can only operate on a subset of first-order statements that guarantee decidability and computability in finite times. For example, in description logics it is possible to express the fact that the birth rate of the hare depends at any given time on both the abundance of the hare and that of the lynx; it is, however, impossible to express the nonlinear dependency that translates the notion of predation. Therefore, machine reasoning cannot be used to assess facts relative to the species interaction unless additional information is given. The mechanism of interaction between the species can only be captured with externally processed information such as an equation, whose syntax can be expressed through a literal property and is easily parsed by software, but whose logical underpinnings remain obscure to a reasoner.

Even within the constraints of first-order logics, reasoning can be profitably used to enforce sound designs and consistent definitions of models. As an example, a biodiversity ontology used to model an ecological community can state that coexistence of populations in a community (e.g. as captured in the coexisting-population relationship) also implies coexistence in space and time, a constraint that can be checked by an appropriately configured semantic modelling system which will refuse, for example, to simulate a system where lions prey on dolphins simply by checking population distribution data.

A knowledge-driven specification enables an appropriately configured system to automatically generate an algorithm (the actual simulation model) capable of generating a simulated dataset from given values of parameters and initial conditions. In this specific case, the system in Fig. 4 is easily compiled into a system similar to that of Fig. 2, which can later be calculated by numeric integration. A semantically enabled environment can target different compilation or execution environments, for example compiling the simulation in a high-level programming language or into a declarative workflow system.

In the knowledge-driven approach, what has conventionally been called the modelling paradigm can be incorporated to a certain degree in the reasoning strategy that "resolves" the specification into a running simulation. As a result, the choice of modelling paradigm may become less relevant to the modeller. The real paradigm in such a system is captured in the ontology used to lay out the conceptual definitions of the system. Finer-grained paradigm choices such as how to interpret the differential equations composing the system (e.g. continuous time vs. discrete time) can be absorbed in the knowledge base of the framework; users only need to concentrate on accurately describing the system using known, well-documented concepts and their relationships. To this end, knowledge modelling tools such as GrOWL (Krivov et al., 2007) can be used effectively. For example, an individual-based solution for the system of Fig. 4 is also possible; decisions can be taken automatically based on the ontologies themselves and on the comparison of efficiency metrics. For example, Villa (2007) describes a hybrid model where reasoning is used to decide the best strategy to model a coupled system only after independent conceptual models have been merged into a higher-level one.

No matter what strategy has been chosen to simulate the system, the result of simulating a system will produce a semantically enriched result that preserves the original identities. For example, simulating the system of Fig. 4 may produce a result very similar to the static system of Fig. 3, only with multiple abundance values per each time step. The system will automatically define a process in terms of variables, inputs, outputs, concepts that users won't need to manipulate unless they want to, and use it to "resolve" the dynamic formulation into numeric states. The dynamic specification, shown partially in Fig. 4, is considerably more verbose, but still recognizable as an extended version of the previous "static model". It is noteworthy how metadata, units, and other information are the same: storage, query or inference can use the same infrastructure for both the static and dynamic model. The boundaries of "dataset" vs. "model" can be effectively blurred by a full semantic specification, reducing to a choice of level of detail and representational framework for a similarly conceptualized system.

5. Discussion and perspectives

Advanced knowledge-based systems, e.g. IMA (Villa, 2001, 2007) and Kepler (Kepler, 2004), are being prototyped and have been applied to selected case studies (Pennington et al., 2007; Villa et al., 2007, 2008). Such systems are typically not committed to a particular set of concepts, with the possible exception of a small core ontology, carefully designed for generality, paradigm neutrality, and extensibility. The uncoordinated extensibility of such environments allows domain experts to produce knowledge describing specific disciplinary contexts without having to employ specific tools. Users can adopt the necessary concepts to produce representations of natural systems; the software architecture, informed by the ontologies, can resolve the system model into numeric states. Such prototypes are laying the groundwork to make a purely semantic approach possible, where all technological details related to the calculation of the model are inferred by machine reasoning based on the logical assertions that define the model, while remaining hidden from the user. The rationale for this approach is the notion that an accurate description of nature's entities needs also to be complete: if enough knowledge has been given to allow a full description of the system, the description also embodies enough information to allow a system to calculate the corresponding model.

The advances discussed so far introduce potentials that were obviously not available to ecologists and environmental scientists before ontologies were available. Even the "lighter" semantic approach, which uses formal concepts to enrich conventionally specified model and data structures, adds a powerful dimension to modelling because of the novel integration potential. Adding to that, initiatives in environmental planning (Villa et al., 2008) and ecology (SEEK, 2004) are being paralleled by others in agriculture (SEAMLESS, 2005), geosciences (GEON, 2005) as well as genomics and other branches of natural sciences. Most of these initiatives are coordinating in order to make their core ontologies eventually converge towards commonly accepted standards. At the same time, obstacles to widespread adoption of semantic approaches remain that may reduce or postpone the benefits resulting from these new opportunities. In this section, we briefly review the most relevant potential benefits and the main challenges to adoption.

5.1. New opportunities

5.1.1. Multi-paradigm modelling

Modelling at the conceptual level allows users to employ a language that's tailored to the knowledge domain of reference, adding the necessary dynamic information to the definition so obtained, and letting appropriate infrastructures define the corresponding computing workflows. Because each modelling paradigm can be described by a set of ontologies handled by matching software, a system can be extended so that, for example, the flow/stock model can coexist with others where the behaviors of organisms are modeled individually, instead of relying on community-level state variables. The details of the scheduling and the interactions between different calculation workflows can be sorted out automatically. For example, a knowledge-explicit approach can greatly ease the specification of hybrid models, such as those that are most necessary in decision making, e.g. landscape models (best modeled as a spatially explicit process based model) coupled with human component models that react to changes in the landscape and influence it in turn (best modeled as individuals moving on the landscape and reacting to its change). Ontologies can identify the common ancestor concepts that allow both the modeller and the infrastructure to represent the coupled model consistently or to seamlessly merge the independently developed components.

5.1.2. Automated contextualization in space and time

Models are necessary because the phenomena they describe vary in time and space. The property of being distributed in time and space causes a multiplicity of states for the variables of a system. When we model the temporal or spatial heterogeneity in a system, what is being modeled is not the system itself, but more accurately the context of its observation. Temporal and spatial scales of observation can be changed, so that a dynamic model appears constant, or so that what appears to be a static variable reveals fine-grained internal dynamics. By virtue of their relative conceptual independence, time and space can be modeled independently from the abstract conceptualization of natural systems; the definition of the contexts of time and space can be connected to that of the entities by coupling the definition of the temporal and spatial contexts of interest with that of the modeled entities and their behavior.

An important consequence of adopting a knowledge-explicit approach to modelling is that when space and time become part of the allowed semantics, there is no need for specially tailored knowledge or tools for basic spatially explicit modelling, because such functionalities can be invoked as necessary by the knowledge-based system, and the paradigms necessary to enact it are automatically integrated into the specification. It is for example conceivable to make a non-spatial model spatially explicit by simply describing one or more of its components as distributed in space (Villa, 2007). A properly configured system can propagate the notion of space in one concept to the whole conceptual network, or mediate competing representations by operating transformations, e.g. to propagate coarse polygon data over a fine-resolution grid.

As a generalization of the contextualization mechanism in time and space, it is possible to imagine other reasons for a multiplicity of states in the simulation of a model. As an example, it is possible for the behavior of the system to depend on conditions or parameters whose values are not known exactly, but can be assumed to have different values according to different likely scenarios or interpretations. It is useful to be able to model this situation as a "classification" context, which can be included in a modeled system to cause the automatic replication of the simulation as many times as the number of competing explanations of its behavior. This becomes very useful in what-if prediction, sensitivity analysis and optimization.

5.1.3. Model discovery in databases

In a distributed database context, semantically aware modelling opens novel and exciting perspectives such as "model-driven query" (Villa, 2007). An appropriately general version of a model can become a powerful discovery tool that can be used as a constraint over a distributed knowledge base or semantic web in order to discover new knowledge in a fully automated way. As an example, the concept of a species–area relationship can be defined as a property of any set of coexisting populations of the same taxon. A semantically aware infrastructure can automatically translate this logical statement definition into a query, and launch an iterative process that identifies all possible instances of species–area relationships represented by the population data stored in a database. A food web can be described in a similar way. It becomes possible to match an abstract model structure to a distributed database that is semantically annotated. By describing patterns of interest in terms of ontologies, repositories of environmental knowledge with sufficient semantic information can be searched to automatically discover patterns and relationships that have traditionally taken lengthy investigations to find, even when only the necessary data are present in a database.

5.2. Challenges and barriers to adoption

Despite the clear potential offered by semantically aware technologies applied to environmental modelling, only limited case studies are available, and as a consequence the practicality of large-scale adoption remains to be fully understood. While it is clear that the advantages of semantic specification of natural systems will play a role in the future of environmental modelling, the extent to which this will happen will depend on both technical feasibility and acceptance by researchers.

5.2.1. Technical feasibility

While by no means trivial, the engineering aspects of higher-level knowledge-based systems are well within reach, and prototypes of knowledge-based modelling systems are in use today (Villa et al., 2007). Ontology frameworks and knowledge editors allow convenient specification of knowledge and provide standardized platforms for sharing it. High quality reasoner programs are available as open source and can be easily run on ordinary computers. Indeed, at the time of this writing, natural systems science seems well on its way to the adoption of the mediation approach. Projects such as ARIES, SEAMLESS, SEEK and GEON strongly emphasize the role of ontologies to describe data products and analytical steps and enable their composition.

Yet, Description Logics limits the complexity of what can be obtained through reasoning quite severely, preventing for example the analysis of the consequences of nonlinear interactions in models. These limitations, inherent to the theoretical aspects of knowledge representation and unlikely to be solved in the short term, prevent a true "logical modelling" from existing: any nonlinear model will need to be generated from the first-order description of the systems, but no automated inference can inform us about their likely behaviors without simulating the systems first.

A larger, if less clearly identified, problem has to do with the lifecycle of shared knowledge and the alignment of annotated content with meanings that shift as the knowledge evolves. First of all, there is no "right" ontology for any domain; the choice of conceptualization depends on the purposes as well as on often rather personal viewpoints and assumptions. In fact, Goguen (2005) refers to ontologies as *theories*, and highlights the need and the possibility of developing bridges between different ontologies that respect the diversity of viewpoints without preventing cross-reasoning. Unfortunately, knowledge alignment algorithms remain to this day experimental and complex, requiring significant human supervision, and one-to-one concept alignment is often impossible. So the usefulness of semantic approaches to modelling for integration and harmonization can be limited by the degree of sharing of the viewpoints expressed in the ontologies used.

Even when ontologies are agreed upon, they are likely to evolve in time as the understanding of the natural systems changes. The problem of keeping existing annotations and conceptualization aligned with an evolving knowledge base is similarly complex. The management of the lifecycle of knowledge is a highly researched area whose future results will decide much of the practicality of adopting semantic approaches in modelling. The current state of the art in knowledge-driven modelling is mostly using conceptualizations developed ad hoc to reflect the needs of the project, and is by consequence relatively immune to these problems. Alignment issues will become more significant as the adoption of semantic annotation increases, or when knowledge-driven modelling approaches become mainstream.

5.2.2. Adoption

At this stage, it is hard to ascertain whether the adoption of ontology-driven approaches by the larger community will be smooth. This factor becomes even more important in the semantic approach, where ontologies incorporate paradigms for specification and reasoning as opposed to the simpler role of "tagging" information for validation and mediation of workflows, and are therefore bound to create more controversy.

As a consequence, acceptance problems may hamper the large-scale usage of semantic approaches. Even when the difficulties of developing, storing and maintaining large ontologies are successfully addressed, their recognition and acceptance will likely remain difficult for some time. For now, knowledge-based approaches are a little-understood black box in the minds of many scientists. Knowledge-based computing will require a paradigm shift in the way environmental scientists think about modelling, whose feasibility will depend on scientists feeling that the knowledge incorporated in models is *their* knowledge. Bridging technologies are available that allow us to merge, to a certain extent, concepts from different ontologies into larger-scale ones. Yet, integration remains difficult and confidence in new technology is typically low at first. The progress of environmental modelling in the next decade will probably clarify many aspects of the feasibility of these new approaches.

Acknowledgements

This work is supported by the US National Science Foundation, award DBI 0640837 (ARIES) for Villa, and by the European Union, award 010036-2 (SEAMLESS), for Villa, Athanasiadis and Rizzoli. Three anonymous reviewers contributed greatly to clarifying and improving the manuscript.

References

AGROVOC, 2006. The AGROVOC multilingual dictionary of the United Nations food and agriculture organization. www.fao.org/agrovoc/.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. Nature Genetics 25, 25–29.

Athanasiadis, I.N., 2006. An intelligent service layer upgrades environmental information management. IT Professional 8, 34–39.

Athanasiadis, I.N., 2007. Towards a virtual enterprise architecture for the environmental sector. In: Protogeros, N. (Ed.), Agent and Web Service Technologies in Virtual Enterprises. Idea Group Inc., pp. 256–266.

Athanasiadis, I.N., Villa, F., Rizzoli, A.E., 2007a. Ontologies, JavaBeans and Relational Databases for enabling semantic programming. In: Proceedings of the Thirty-First IEEE Annual International Computer Software and Applications Conference (COMPSAC), vol. 2. IEEE, pp. 341–346.

Athanasiadis, I.N., Villa, F., Rizzoli, A.E., 2007b. Enabling knowledge-based software engineering through semantic-object-relational mappings. In: Proceedings of Third International Workshop on Semantic Web-Enabled Software Engineering, Fourth European Semantic Web Conference, pp. 16–30.

Athanasiadis, I.N., Rizzoli, A.E., Donatelli, M., Carlini, L., 2006. Enriching software model interfaces using ontology-based tools. In: Third Biennial Meeting of the Int'l Environmental Modelling and Software Society. Burlington, VT, USA.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (Eds.), 2003. The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press.

Baker, K., Benson, B., Henshaw, D.L., Blodgett, D., Porter, J.H., Stafford, S.G., 2000. Evolution of a multisite network information system: the LTER information management paradigm. Bioscience 50, 963–978.

Batzias, F.A., Siontorou, C.G., 2006. A knowledge-based approach to environmental biomonitoring. Environmental Monitoring and Assessment 123, 167–197.

Beckett, D., 2004. RDF/XML syntax specification (revised). http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210.

Bertino, E., Catania, B., Zarri, G., 2001. Intelligent Database Systems. Addison-Wesley, New York.

Brilhante, V., 2005. Ecolingua: a formal ontology for data in ecology. Journal of the Brazilian Computer Society 11 (2), 61–78.

Ceccaroni, L., Cortés, U., Sànchez-Marrè, M., 2004. OntoWEDSS: augmenting environmental decision-support systems with ontologies. Environmental Modelling & Software 19 (9), 785–797.

Chau, K.W., 2007. An ontology-based knowledge management system for flow and water quality modeling. Advances in Engineering Software 38, 172–181.

Chen, Z., Gangopadhyay, A., Karabatis, G., McGuire, M., Welty, C., 2007. Semantic integration and knowledge discovery for environmental research. Journal of Database Management 18, 43–68.

Duepmeier, C., Geiger, W., 2006. Theme park environment as an example of environmental information systems for the public. Environmental Modelling & Software 21, 1528–1535.

GBIF, 2004. Global biodiversity information facility. http://www.gbif.org.

GEON, 2005. GEON cyberinfrastructure for the geosciences. http://www.geongrid.org.

Goguen, J., 2005. Information integration, databases and ontologies. http://www.cs.ucsd.edu/~goguen/projs/data.html Available from:

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

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human–Computer Studies 43, 907–928.

McGuinness, D., Fox, P., Cinquini, L., Benedict, P.W.J., Garcia, J., 2007. Current and future uses of OWL for scientific data frameworks: successes and limitations. In: Golbreich, C., Kalyanpur, A., Parsia, B. (Eds.), Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions, p. 258.

Guinness, D.L.M., van Harmelen, F. (Eds.), 2004. OWL Web Ontology Language Overview. http://www.w3.org/TR/owl-features/.

ISO, 2004. ISO/TC211 Geographic Information Metadata Standard.

Jones, M.B., Berkley, C., Bojilova, J., Schildhauer, M., 2001. Managing scientific metadata. IEEE Internet Computing 5, 59–68.

Keller, R.M., Dungan, J.L., 1999. Meta-modeling: a knowledge-based approach to facilitating process model construction and reuse. Ecological Modelling 119, 89–116.

Kepler, 2004. The Kepler project. http://www.kepler-project.org.

Khatri, V., Ram, S., Snodgrass, R.T., 2006. On augmenting database design-support environments to capture the geo-spatio-temporal data semantics. Information Systems 31, 98–133.

Kiryakov, A., Popov, B., Ognyanoff, D., Manov, D., Kirilov, A., Goranov, M., 2003. Semantic annotation, indexing, and retrieval. Lecture Notes in Computer Science 2870, 484–499.

Kohler, J., Lange, M., Hofestadt, R., Schulze-Kremer, S., 2000. Logical and semantic database integration. IEEE International Symposium on Bio-Informatics and Biomedical Engineering, 77–80.

Krivov, S., Williams, R., Villa, F., 2007. GrOWL, visual browser and editor for OWL ontologies. Journal of Web Semantics: Services and Agents on the World Wide Web 5 (2), 54–57.

Lee, T.B., Hendler, J., Lassila, O., 2001. The semantic web. Scientific American 284, 28–37.

Lee, S., Wang, T.D., Hashmi, N., Cummings, M.P., 2007. Bio-STEER: a semantic web workflow tool for grid computing in the life sciences. Future Generation Computer Systems 23, 497–509.

Ludaescher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, M., Lee, E.A., Tao, J., Zhao, Y., 2005. Scientific workflow management and the Kepler system. Concurrency and Computation: Practice and Experience 18, 1039–1065.

Ludaescher, B., Gupta, A., Martone, M.E., 2001. Model-based mediation with domain maps. Seventeenth International Conference on Data Engineering (ICDE) Heidelberg, Germany.

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., Villa, F., 2007. An ontology for describing and synthesizing ecological observation data. Ecological Informatics 2, 279–296.

Muetzelfeldt, R., 2004. Declarative modelling in ecological and environmental research. In: European Commission Directorate-General for Research.

Muetzelfeldt, R., Massheder, J., 2003. The Simile visual modelling environment. European Journal of Agronomy 18, 345–358.

NASA, 2006. Semantic web for earth and environmental terminology (SWEET). http://sweet.jpl.nasa.gov.

Parr, C., Parafiynyk, A., Sachs, J., Ding, L., Dornbush, S., Finin, T., Wang, T., Hollender, A., 2006. Integrating ecoinformatics resources on the semantic web. In: Proceedings of the Fifteenth International World Wide Web Conference. ACM Press, pp. 1073–1074.

Paterson, T., Kennedy, J.B., Pullan, M.R., Cannon, A., Armstrong, K., Watson, M.F., Raguenaud, C., McDonald, S.M., Russell, G., 2004. A universal character model and ontology of defined terms for taxonomic description. Lecture Notes in Computer Science 2994, 63–78.

Pennington, D., Higgins, D., Peterson, A.T., Jones, M.B., Ludaescher, B., Bowers, S., 2007. Ecological niche modelling using the kepler workflow system. In: Workflows for Science: Scientific Workflows for Grids. Springer Verlag, New York, pp. 91–108.

Raskin, R.G., Pan, M.J., 2005. Knowledge representation in the semantic web for earth and environmental terminology (SWEET). Computers and Geosciences 31, 1119–1125.

Richmond, B., 2001. An Introduction to Systems Thinking: STELLA Software. High Performance Systems Inc.

Rizzoli, A.E., Donatelli, M., Athanasiadis, I.N., Villa, F., Huber, D., 2008. Semantic links in integrated modelling frameworks. Mathematics and Computers in Simulation 78, 412–423.

Robertson, D., Bundy, A., Muetzelfeldt, R., Haggith, M., Uschold, M., 1991. Eco-logic: Logic-Based Approaches to Ecological Modelling. MIT Press, Boston, MA.

Salles, P., Bredeweg, B., Araujo, S., 2006. Qualitative models about stream ecosystem recovery: exploratory studies. Ecological Modelling 194, 80–89.

Scholten, H., Kassahun, A., Refsgaard, J.C., Kargas, T., Gavardinas, C., Beulens, A.J.M., 2007. A methodology to support multidisciplinary model-based water management. Environmental Modelling & Software 22 (5), 743–759.

SEAMLESS, 2005. SEAMLESS IP home page. http://www.seamless-ip.org.

SEEK, 2004. Enabling the Science Environment for Ecological Knowledge. http://seek.ecoinformatics.org.

Tiller, M., 2001. Introduction to Physical Modelling with Modelica. Springer, New York, NY.

Villa, F., 2001. Integrating modelling architecture: a declarative framework for multi-paradigm, multi-scale ecological modelling. Ecological Modelling 137, 23–42.

Villa, F., 2007. A semantic framework and software design to enable the transparent integration, reorganization and discovery of natural systems knowledge. Journal of Intelligent Information Systems 29, 79–96.

Villa, F., Ceroni, M., Krivov, S., 2007. Intelligent databases assist transparent and sound economic valuation of ecosystem services. Environmental Management 39, 887–899.

Villa, F., Ceroni, M., Krivov, S., Johnson Jr., G.W., Bagstad, K., Batker, D., Portela, R., Honzak, M., 2008. The ARIES project: artificial Intelligence for Ecosystem Services. Ecoinformatics Collaboratory white paper, UVM. Available from: http://esd.uvm.edu/uploads/media/ARIES.pdf.

Villa, F., Costanza, R., 2000. Design of multi-paradigm integrating modelling tools for ecological research. Environmental Modelling & Software 15, 169–177.

Wand, Y., Storey, V.C., Weber, R., 1999. An ontological analysis of the relationship construct in conceptual modelling. Acm Transactions on Database Systems 24, 494–528.

Webs on the Web (URL). Webs on the Web project: home page. http://www.foodwebs.org/index_page/wow2.html.

Wenzel, V., 1992. Semantics and syntax elements of a unique calculus for modelling of complex ecological-systems. Ecological Modelling 63, 113–131.

Williams, R.J., Martinez, N.D., Goldbeck, J., 2007. Ontologies for ecoinformatics. Web Semantics, Science, Services and Agents on the World Wide Web 4, 237–242.