International Congress on Environmental Modelling and Software

4th International Congress on Environmental Modelling and Software - Barcelona, Catalonia, Spain - July 2008

Jul 1st, 12:00 AM

Conceptual Model of Single Information Space for Environment in Europe

Jiří Hřebíček

K. Kisza

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Conceptual Model of Single Information Space for Environment in Europe

J. Hřebíček[a], **K. Kisza**[b]

[a] *Institute of Biostatistics and Analyses, Masaryk University, Kamenice 126/3, 625 00 Brno, Czech Republic (hrebicek@iba.muni.cz)*

[b] *Faculty of Informatics, Masaryk University, Botanícká 68a, 602 00 Brno, Czech Republic (kisza@iba.muni.cz)*

Abstract: The paper presents the further development of the conceptual model of the Single Information Space for the Environment in Environment (SISE) and compares this with an upper ontology concept of the SISE. The developed conceptual model of the SISE enables an implementation the vision of development an integrated, modern, common, shared and sustained Single European Information Space infrastructure for environmental information exchange and environmental management in Europe.

Keywords: Environmental data; Environmental information; Knowledge; Information space for environment; Conceptual model.

1. INTRODUCTION

The development of the Single European Information Space has been the first aim of the European i2010 strategy, [i2010, 2005] since 2005. The objective "ICT-2007.6.3: ICT for Environmental Management and Energy Efficiency" of the Seventh Research Framework Programme (FP7) specified this aim for the area of environmental protection and sustainable development. The aim of the development of the Single Information Space for Environment in Europe (SISE) was introduced by Schouppe [2008]: an Information and Communications Technology (ICT) research vision for real-time connectivity between multiple environmental resources which would allow seamless cross-system search, as well as cross-border, multi-scale, multi-disciplinary data acquisition, pooling and sharing. This aim of SISE is to provide some sort of integrated information space in which environmental data and information are combined with knowledge. This infrastructure will enable a "holistic view" and allow the processing of different types of environmental data and information to extract more knowledge for decision making (correlating information and data) that are not currently possible. Ongoing developments in the context of thematic environmental legislation of EU are increasingly recognising the need to adopt a more modern approach to the production, exchange and use of environmental data and information. Full attention will be on the optimisation of complex data flows across all decision levels, across borders and sectors in developed Shared Environmental Information System (SEIS) by the Go4 team (DG Environment, EEA, Eurostat and JRC) [COM46 final, 2008], [SEIS, 2008].

The development of the complete and complex SISE covering all interactions among environmental data, information and knowledge using current ICT tools is practically impossible [Pillmann et al., 2006]. There is very fast growth of amount of data, information and knowledge all over the world each day. Therefore, it is appropriate to develop a common methodology of building the basic conceptual model of the SISE, which enables the common overview on environmental data, information and knowledge in standardized way. This paper takes into account some of above mentioned challenges of the SISE and the SEIS and presents and compares two concepts: an upper ontology approach and author's conceptual model approach. The conceptual model of the SISE issues from ideas previously published by Hřebíček and Sluka [2003], Pillmann et al. [2006], Hřebíček et al. [2007], Hřebíček and Ráček [2007], and Schouppe [2008] and the paper summarizes the current results of author's team in the research of this topic.

2. MOTIVATION

Let us consider for a basic model of the SISE using of the upper ontology concept [Niels and Pease, 2001], [Pease, 2003], [Batres et al., 2007], [Villa, 2007]. The upper ontology concept offers fundamental structure and rule sets according to which is to build domain ontology models (for example ontology for medicine, financial engineering, etc) to achieve their compatibility. The domain ontology models for complete area of environment are not developed yet.

The general ontological model includes four basic elements: *individuals, classes, attributes* and *relations* and they are described in details e.g. in Wikipedia [2008].

Let us have two domain ontology models built according to upper ontology principles, then it is secured their compatibility in the most general form, i.e. there are no different definitions of equivalent classes in these models. The upper ontology concept is not suitable for more specific tasks (for example for the public access to environmental factors or effects), where the detailed level of the solution is needed. In this case it is necessary to use more detailed model, which is of much specific than the upper ontology model. If we have the class that provides the extraction of specific information type, (particularly for environmental information of factors or effects used for the decision making support), it is necessary to find out the primary data used for the extraction of this information. There are two basic possibilities for solving above problems:

- To define the class containing both primary data, and procedures for the information extraction in its attributes (this way is very unpractical – it could be represented by huge data and information aggregations which are very difficult to process – apart from data duplicity).
- To define special relations and classes enabling to form the information structure in the model (during the processing of such model it is necessary to know which relations and which classes were used and their context).

We developed the conceptual model formalized this second approach. Generally, there are many ways of formalization. We started from the object-oriented model of environmental information described by Hřebíček and Sluka [2003]. Hřebíček and Ráček [2007] have developed the basic conceptual model of the SISE and presented it on the conference ISESS 2007 in Prague. This version was generalised by Hřebíček et al. [2007], and the last renewed one is presented in the next chapter.

2. CONCEPTUAL MODEL

2. 1 Single Information Space for Environment in Europe

Let us consider a *network of constituents* consisting of four principal sets $\underline{I}$, $\underline{M}$, $\underline{O}$ and $\underline{A}$ which represent *universum* of classes of *information, methods, objects* and *attributes*, [Hřebíček and Ráček, 2007], [Hřebíček et al, 2007]. Every class of attributes, objects, methods and information includes also *meta-data description*. All four sets have a *tree structure* (continuous acyclic directed graphs), where nodes (classes) in lower tree layers inherit their structure (including also meta-data) from nodes in upper layers (Figure 1),.

Definition: Let us define the SISE as a quintuple: $S = [I, M, O, A, R]$ where I, M, O and A are domains of classes – tree structured – and R is a set of constituent inheritance and constituent aggregation relations (relation of inheritance, parent, child, predecessor, successor).

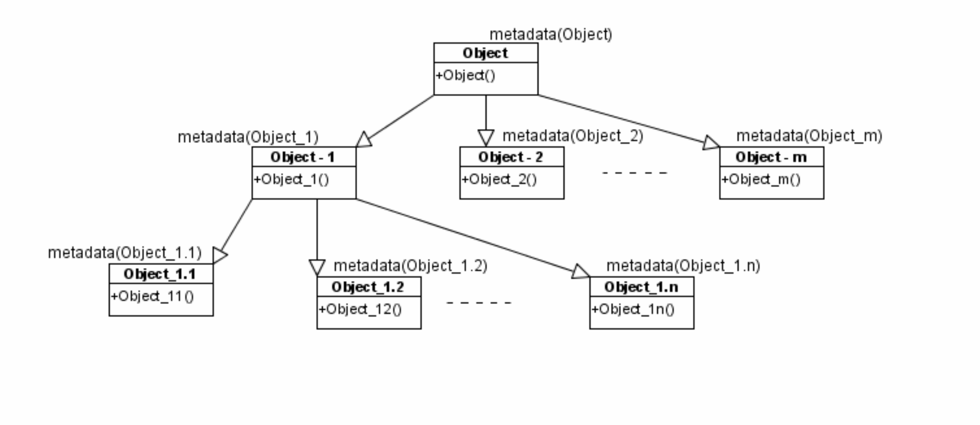

Figure 1: Common structure of object tree.

Following terms are used in the definition of the SISE:

- The tree is the continuous acyclic directed graph – the radical tree.
- The top element *v* is denoted as the *root* of the radical tree. The root *v* is the element, to which any edge is headed, there are only edges headed from it (at least one edge).
- The root is an element that represents the basic (the most general) type, from which are derived all other elements in the tree – *relation of inheritance.*
- For each element *w* is defined the *parent* element *p* as the element which is connected by *edge (p, w)* heading from *p* to *w* (every element different from the root element has only one parent). For the element *w* is defined the *child* element as any element *q*, which is connected by *edge (w, q)* heading from *w* to *q*. Each element except the leaf has at least one child (is able to have also more children). The leaf has no child.
- Any tree element in the tree with basic type of the root *v* is either empty structure or element of the type *v*, which is connected with the finite number of disjoint tree structures with the basic type *v* (mark them as *subtree*).

Lemma 1: If *p* is the child of *q* and at the same time *q* is the child of *r*, then *p* is child of *r*.

Assumption 1: Let sets *I*, *M*, *O* and *A* are defined as domains of *information, methods, objects* and *attributes* and we assume, that the constituent aggregation is hidden in next mandatory rules:

- For every $i \in I$ exist set *M'* and relation *r* where $M' \subseteq M$ and *r(i, M')* is valid;
- For every $m \in M$ exist set *O'* and relation *r* where $O' \subseteq O$ and *r(m, O')* is valid;
- For every $o \in O$ exist set *A'* and relation *r* where $A' \subseteq A$ and *r(o, A')* is valid.

The simple example of the above structure is presented on Figure 2.

2.2 Representation of information (knowledge)

Let us consider:

- Information is represented as a continuous acyclic directed graph – *four-leveled radical graph.*
- The top element *v* is denoted as the root of this radical graph. The *root v* is the element, to which any edge is headed, there are only edges headed from it. The top element belongs to *I* set.

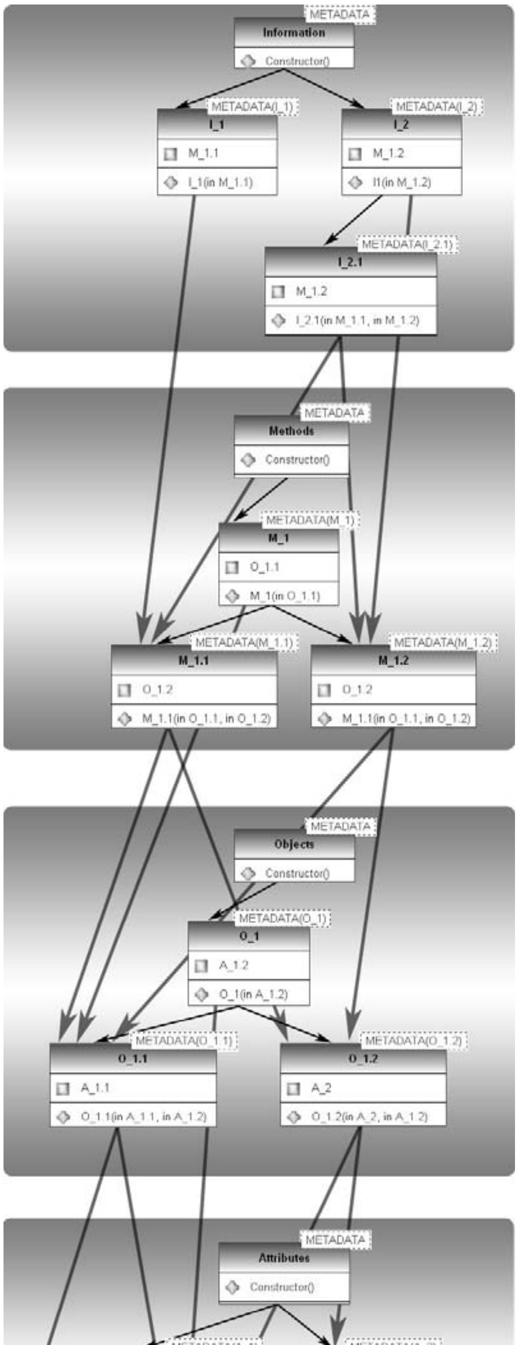

Figure 2: Basic structure of Environmental Information Space.

- There will not be used the relation of inheritance for purposes of this graph, but new relation *R* is defined: *R:* M^n -> *I*, or O^n -> *M*, or A^n -> *O*.

- The root element of each relation is just one element from set *I* (i.e. the first graph level). Its successors (i.e. the second graph level) are elements from set *M*. The successors of these elements from set *M* (i. e. the third graph level) are elements from set *O*. The successors of these elements from set *O* (i.e. the fourth graph level) are elements from set *A*.
- For each top element *w* is defined the *predecessor* element *p* as the element, which is connected by the *edge (p, w)* heading from *p* to *w*. For each top element *p* is defined the *successor* element *w* as the element, which is connected by *edge (p, w)* heading from *p* to *w* (is able to have also more successors).

Lemma 2: It is valid:

- If *w* belongs to the *A*, then all its successors belong to the *O.*
- If *w* belongs to the *O*, then all his successors belong to the *M.*
- If *w* belongs to the *M*, then all his successors belong to the *I.*
- If *w* is successor of *u*, then for no child is *w* his successor (i.e. w is not successor of no element which is element of the subtree, where the root is *u*).

3. IMPLEMENTATION IDEA

3.1 Semantic web services

The base sets of elements of the SISE are elements on the different implementation level in the graph representing concrete information. Implemented elements are the same (constant) for all users (primary data, a mathematical definition of methods, etc). Unimplemented elements are represented like interfaces which are necessary to implement according to concrete user possibilities, without necessity to know whole "domain knowledge". Both implemented elements and interfaces represent "domain knowledge". They are interconnected with the ontology representing "operational knowledge".

Information is represented by the graph that contains implemented elements and unimplemented elements. Unimplemented elements are defined as common interfaces with input and output limitations. Every such element can be implemented like service. Type of those services refers to the type of the element (belonging to base set):

- *M – method* – a service enables obtaining information from aggregate data (elements of set *O*).
- *O – objects* – a service enables data access.
- *I – information* – a service enables the aggregation of basic information – outputs of methods – for obtaining information in the way requested by user.

Any information can be present like a process *T* that has defined the plan, which has to be executed to achieve the requested result (information). The process *T* is the complex process that is able to contain sub-processes that must be scheduled and executed (elements of sets *M, O, A*). Each of these sub-processes has its input and output limitation (characteristics). Each of those subtasks has preconditions and post conditions that must hold, as does *T* itself. Those preconditions and post conditions are properties that must hold either prior to or posterior to the planning, scheduling, or execution of that task/subtask. There are constraints on those phases of the complex service enables to get complex information (the element of *I*).

We can apply this model to describe any information, data or service and their relations in SISE. It enables contextual reasoning and the automated transition (coercion) of semantics from one context to another. This approach enables integration of the data and metadata associated with a service together with specifications of its properties and capabilities, the interface for its execution, and the prerequisites and consequences of its use.

For example in SISE we can describe the configuration for any information from its components according to a required specification of information as a quintuple *[R‘, I‘, M‘, O‘, A‘]*, where objects are connected with relations and:

- *A‘⊆A (i.e. I is a subset of I);*
- *O‘⊆O;*
- *M‘⊆M;*
- *I‘⊆I;*
- *R‘⊆R.*

These configurations cover the domain knowledge. They describe attributes, objects, methods and relations among them, necessary to get any information defined in SISE.

If we consider some area of interest we can take it as a quintuple *[R‘, I‘, M‘, O‘, A‘]* where elements are on various level of implementation. Implemented elements are the same for any user who use the SISE (mathematical definitions, common data …). Unimplemented elements are represented as interfaces which are necessary to implement according to possibilities of concrete user without need to know whole domain knowledge. We suggest implementing these interfaces as semantic web services. Both implemented parts and unimplemented parts represent the domain knowledge. They are connected by ontology relations represent the operational knowledge.

Generally, ontology defines a common vocabulary for researchers whose need to share information, data and services in domain. It includes machine-interpretable definitions of basic concepts in the domain and elations among them. The main advantages of using ontology in our model for SISE are that ontology enables mainly:

- to analyze domain knowledge and make domain assumptions explicit;
- to share common understanding of the structure of information among people or (and) specialized software technologies (computer agents). For example, we can suppose several different web sites (each with its own database) contain environmental information. If these web sites implement the same underlying ontology (defined in SISE), then computer agents can easily extract and aggregate related information from these different sites. This aggregated information can be used as an answer to user queries or as an input data to other applications.
- to reuse of this domain knowledge (after making analysis);
- to separate domain knowledge from the operational knowledge;
- to provide a large extent of flexibility and expressiveness, the ability to express semi-structured data and constraints, and support types and inheritance.

As an instrument for implementation of services we suggest use Web Ontology Language (OWL-S) [OWL-S, 2004]. OWL-S (formerly DAML-S) is an ontology of services that makes these functionalities possible. In this submission we describe the overall structure of the ontology and its three main parts: *the service profile for advertising and discovering services; the process model, which gives a detailed description of a service's operation*; and *the grounding, which provides details on how to interoperate with a service, via messages*, [OWL-S, 2004].

OWL-S takes a mostly reactive planning view of the semantics of web services. The reactive planning view means that services and their complex compositions are generally viewed as a three-phase operation: *planning, scheduling*, and *execution*. There is some set of objectives or goals that a developer or user wants to achieve. This set might be viewed as the rationale for the desired web service. One might have multiple plans (various compositions of web services) that could achieve those desired goals. A given plan is selected or composed from a library or registry of services/plans. That plan can be represented as a more-or-less complex task or process model, [SWS, 2005].

This model is suitable for our idea of tree representation of any information defined in S which was mentioned above.

To realize our vision of the SISE, it is necessary to implement theoretical model S in some pilot project and fully describe formalization processes and form recommendations for used technologies. If we concern SISE as an abstract upper model *S*, it will be necessary to define common Semantic constraints (preconditions and postconditions) for defined interfaces of services.

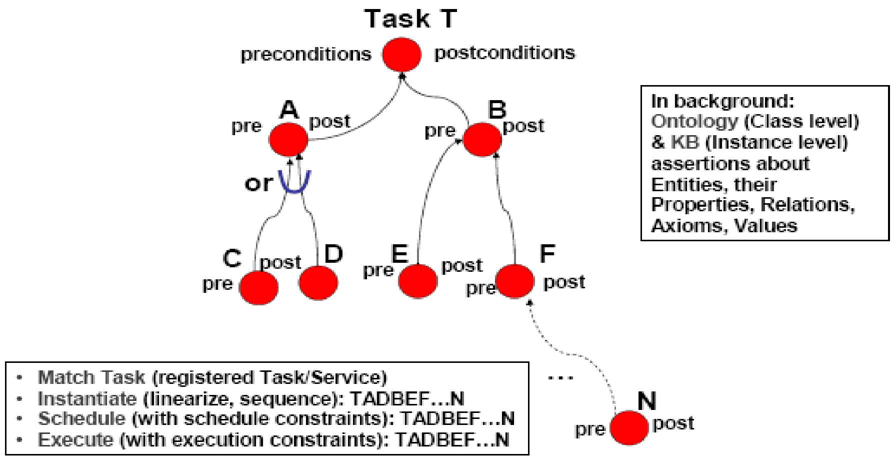

Figure 3. Planning, Task, Process Representation. Source [SWS, 2005].

4. CONCLUSION

Interaction among various digital data, information and knowledge sources are necessary for building SISE, e.g. ensure so - called semantic interoperability. Semantic interoperability is conceptual formulation of metadata structure that allows semantically combine data elements from different schematics, dictionaries and other sources and makes possible to search information across heterogeneous distributed data source. By the help of semantic interoperability are solved e.g. problems, when individual sources use various terms for description of the same term or on the contrary use same terms for various notions.

The technologies supporting semantic interoperability are very popular and exploited nowadays (semantic technology), especially then ontology, which compared to for example UML (Unified Modeling Language) offer some other benefits.

The amount of existing ontological models is rising very quickly. Mostly they are focused on some specified domain (domain ontology) or systems in companies (application ontology). To ensure interoperability of these systems, it is necessary to solve compatibility of systems and it covers as a first step – to standardize model for environmental data, information and knowledge.

The usage of the developed model will enable more exactly find out (identify) which elements (constituents) and in what way they have been used in the extraction of information any time, without necessity to know more details about the meaning of single relations and classes in the whole model context.

Nowadays it is just theoretical model which flows from earlier research and we try to verify it on implementation pilot in biodiversity and waste management area.

ACKNOWLEDGEMENTS

The paper is supported by the Ministry of Environment of the Czech Republic (project No. SP/4i2/26/07), further by the Ministry of Education and Youth (project INCHEMBIOL, No. MSM0021622412).

REFERENCES

Batres, R., Matthew West, M., Leal, D., Price, D., Masaki, K., Shimada, Y., Fuchino, T. and Naka, Y., An upper ontology based on ISO 15926, *Computers & Chemical Engineering*, 31(5-6), 519-534, 2007.

COM46 final, Towards a Shared Environmental Information System (SEIS), Commission of the European Communities, 2008.

Hřebíček, J., and Sluka, J., Model of the Object-Oriented Environmental Information - A New Approach in the Environmental Data and Information Management. In: *17. International Conference Informatics for Environment Protection. The Information Society and Enlargement of the European Union*. Marburg: Metropolis Verlag, Germany, 381-388, 2003.

Hřebíček, J., and Ráček, J., Environmental Information Space I. Theory and Practice. In: *6th International Symposium on Environmental Software Systems (ISESS 2007).* Guelph: IFIP WG 5.11, 123-133, 2007.

Hřebíček, J., Kisza, K., Štefaník, M., and Ráček, J., Environmental Information Space II. *3rd Internationl Summer School on Computational Biology*. Masaryk University, 158-170, 2007.

i2010, COM/2005/0229 final: “i2010 – A European Information Society for growth and employment” <http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm>, 2008/04/30

Maedche, A., Ontology Learning for the Semantic Web. Kluwer Academic Publishers, N.Y., 2002

Niles, I., and Pease, A., Towards a Standard Upper Ontology. In: Proceedings of the 2nd International Conference on Formal Ontology in Information Systems, Chris Welty and Barry Smith, eds, Ogunquit, Maine, 2001

OWL-S, http://www.w3.org/Submission/OWL-S/, 2004

OWL, <http://www.w3.org/TR/owl-features/>, 2008/04/30

Pease, A., The Sigma Ontology Development Environment. In Working Notes of the IJCAI-2003 Workshop on Ontology and Distributed Systems, August 9, Acapulco, Mexico, 2003.

Pillmann, W., Werner Geiger, W. and Voigt K., Survey of environmental informatics in Europe, *Environmental Modelling & Software*, 21(11), 1519-1527. 2006

SEIS, http://ec.europa.eu/environment/seis/index.htm, 2008

Schouppe, M., Towards a Single Information Space for the Environment in Europe. Concept and purpose. Expert Consultation Workshop, European Commission Information Society & Media Directorate-General, Unit ICT for Sustainable Growth. Brussels, 15 February 2008. <ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/sustainable-growth/workshops/workshop-20080215-mschouppe_en.pdf>, 2008/04/30

SWS, Frameworks for Semantics in Web Services Workshop, MITRE Position Paper, 2005

Villa, F., A semantic framework and software design to enable the transparent integration, reorganization and discovery of natural systems knowledge. *Journal of Intelligent Information Systems*, 29(1), 79-96, 2007

Wikipedia, Ontology (information science), <http://en.wikipedia.org/wiki/Ontology_(computer_science)>, 2008/04/30