19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011
http://mssanz.org.au/modsim2011

Environmental Information Enrichment – The TaToo Approach

G. Schimak [a], P. Dihé [b], T. Pariente Lobo [c], G. Avellino [d], L. Petronzio [d], A.E. Rizzoli [e], S. Nesic[e] , B. Božić [a]

[a] *AIT Austrian Institute of Information Technology, Austria; Email: {gerald.schimak, bojan.boziz}@ait.ac.at*
[b] *cismet GmbH; Email: pascal.dihe@cismet.de*
[c] *ATOS Origin, Spain; tomas.parientelobo@atosresearch.eu*
[d] *Telespazio spa, Italy; Email: {luca.petronzio, giuseppe.avellino@telespazio.com*
[e] *IDSIA, Switzerland; Email: {andrea, sasa}@idsia.ch*

Abstract: Difficult to find, difficult to understand! Can we trust this information? How reliable are these data? Would these data or model serve my needs? Uncountable questions are now being raised by numerous end users, after the result of their search query returned some (or tons) of information on the web.

TaToo tries to improve the success rate for reliable and understandable results on the web by a new and promising approach. TaToo is developing a semantic framework providing tools and services for information enrichment and discovery of environmental resources based on a service-oriented and semantically enhanced architecture.

Having available semantic descriptions of environmental resources clearly facilitates the search and discovery of information, as semantics provide "meaning" to items, and allows the search engine to go beyond simple statistical searches, which often return unrelated hits. But this needs a lot of meta-information and semantic information someone would have to enter. The gap between the need for semantics and its general availability is what TaToo is addressing. TaToo facilitates the annotation of environmental resources by providing tagging tools and services so that related to the user's domain and expertise they could enter (tag, comment, rate, etc.) a specific item from the world wide information pool. At the same time TaToo offers discovery tools and services based on the previously entered semantic annotation, thus allowing to return to the user search results improved with quality, uncertainty and ranking information.

TaToo relies on specially developed resource and semantic models, the Minimum Environmental Resource Model, called MERM, and the usage of bridge ontologies, in order to realize semantic interoperability between different environmental domains. Combined with several multilingual aspects TaToo will be able not only to cross-link information stemming from one language area but also from different ones.

This paper describes the architecture and the ontology framework of TaToo, then it details how we address interoperability issues using the MERM model and bridge ontologies; finally it describes the user interaction with the TaToo framework, describing the TaToo portal as the TaToo entry point.

Keywords: *Semantic Annotation, Discovery Framework, Ontology, MERM, Environment, TaToo, Knowledgebase, Information Enrichment*

1. INTRODUCTION

The TaToo project aims at exploiting a common practice among web users porting it to the environmental context: search, discovery and tagging of interesting resources. Tagging, that is attaching short descriptions to discovered resources, allow communities and user groups to label and classify resources, enriching information and enabling aggregators to display the most relevant ones according to the context.

TaToo aims for a community driven approach to information enrichment as a virtuous cycle (see Figure 1), addressing both experts, such as researchers and decision makers, and the wider public in order to setup, extend, use and promote their knowledge by using the TaToo framework as a knowledge sharing platform.

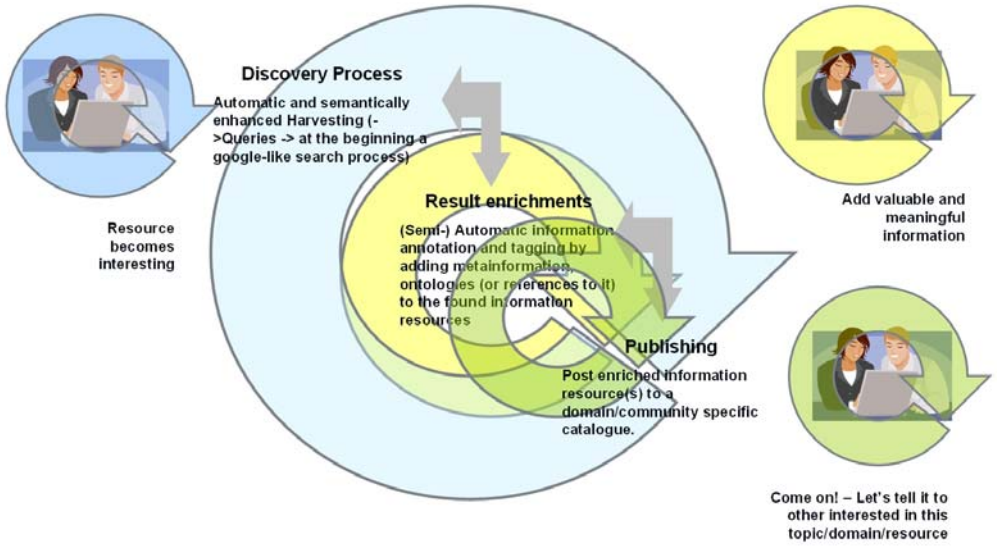

Figure 1. Cycles of information enrichment

2. THE TATOO ARCHITECTURE

The TaToo project intends to provide functionalities to tag resources with metadata in order to improve future discovery through the information enrichment process described above. In order to enable this, the TaToo framework, developed within the TaToo project, implements a client-server architecture, including a set of system components (on the server side) providing the implementation of the functionality, and a set of user components (on the client side) helping the end users interact with the system components through graphical user interfaces. The TaToo framework therefore offers facilities for data and information discovery, visualisation, and evaluation/validation, as well as tagging/annotation by humans and the harvesting of meta-information about resources by (semi-) automated processes.

The architecture of the TaToo framework consists of an implementation independent Functional Purview and an Implementation Purview (Dihé et al., 2011). Implementation independence at the architecture level leads to a more sustainable architecture that is not vulnerable to technology changes in the future, and it is thus able to accommodate changes in technology without changing the architecture itself. This is especially important because TaToo intends to make use of existing solutions, such as widely adopted semantic frameworks like Sesame and Jena.

The Functional Purview is mapped to a concrete implementation platform, resulting in a reference implementation of the TaToo framework. The Implementation Purview complements the Functional Purview in the areas of implementation and technology. The focus of the Implementation Purview lies on how a certain component is realised in detail, the formal specification of the component's interfaces and information exchanged, and the specification of a Minimum Environmental Resource Model (MERM), which will be described later on.

The elements of architectural design that define the overall structure of the TaToo framework are Tiers and Building Blocks. Tiers logically separate general concerns, like presentation and business logic, while

Building Blocks and the components defined therein share a common set of properties and organise and group more concrete functionalities.

Figure 2, shows the Building Blocks of the TaToo framework architecture. The diagram illustrates some essential components and the most important information flows. The TaToo framework architecture is designed as an n-tier architecture that currently consists of the following 4 tiers (Dihé et.al., 2011): The Presentation Tier, which is concerned with user interaction and the presentation and aggregat-ion of information. The Service Tier decoupling the Business Tier and Presentation Tier, as well as serving as a layer to enforce interoperability through the provision of well defined and self-describing service inter faces. The Business Tier, which is responsible for the core functionality (business logic) of the TaToo System. The Data Tier, which is concerned with the storage of semantically enriched information, and other data (registered resources to be harvested, user's information, etc.).

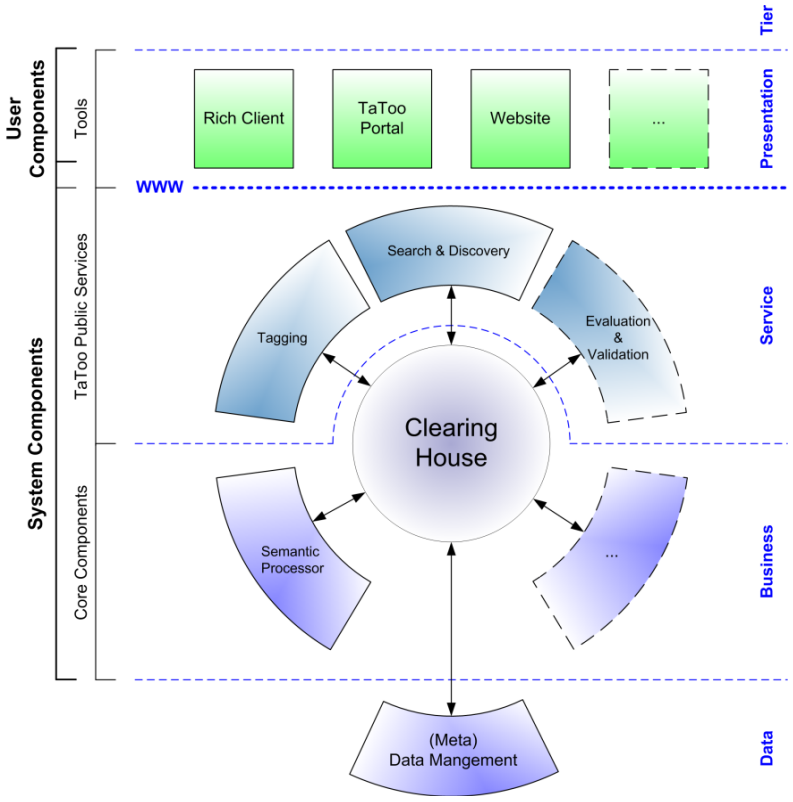

Figure 2. The TaToo Framework Architecture and Building Blocks (Dihé et.al., 2011)

Each client component (e.g. TaToo Tool), has a server-side counterpart that is a (web) service. Besides Tools for validation and evaluation of resources and related semantic meta-information (TaToo Tags), the primary tools provided by the TaToo Framework Architecture are those to tag and discover environmental resources:

- **Tagging Tools** offer a GUI allowing the user to tag arbitrary environmental resources with concepts selected from different domain ontologies.
- **Discovery Tools** provide a GUI suitable for various discovery strategies and thus functionality to search and discover environmental resources through the key concepts and relations chosen from the corresponding ontology domain. They allow users to view, in a user-friendly manner, the results of a query, and to interact with them.

3. THE TATOO SEMANTIC APPROACH

The TaToo project grounds its approach on semantic technologies (Berners-Lee et al., 2001). Semantic technologies allow for a machine-based shared representation and interpretation of knowledge, which rests on the adoption of shared ontologies. Therefore, the establishment of an ontology framework allowing for the production of formal environmental resource descriptions is in the inner core of TaToo.

Although there are ontologies dealing with environmental issues, presently there is no encompassing ontology detailing such a broad domain that fits the needs of TaToo. Therefore, a realistic objective of TaToo in this regard is to allow sub-domain experts to model their own sub-domain and at the same time provide the maximum degree of integration between such sub-domain models. Because of these reasons the ontology framework should ensure *semantic interoperability* between different domain ontologies while ensuring their modular nature.

3.1. Ontology Framework – TaToo's basis for semantic interoperability

There are several approaches to achieve semantic interoperability when dealing with different ontologies. Wache (Wache et al., 2001) defined three ways to integrate ontologies by using single ontology, multiple ontology or hybrid ontology approaches. TaToo follows the hybrid approach as shown in Figure 3.

The mapping process is done manually by defining equivalency and subsumption (is-a) relationships between concepts, object properties, data properties and individuals defined in a so-called bridge ontology and the corresponding entities in the domain ontologies. These mappings are the ontology alignments. As a result, the bridge ontology shows the entire ontology framework as one big knowledge base.

In TaToo, there is also the need to uniformly describe environmental information resources. We think of an environmental resource as a web resource (being a web page, a document, a model, a service, etc.), which is identified by an URI. In this sense, the Minimal Environmental Resource Model (MERM) is defined as a part of the shared ontology that will contain the minimal set of cross-domain concepts and properties.

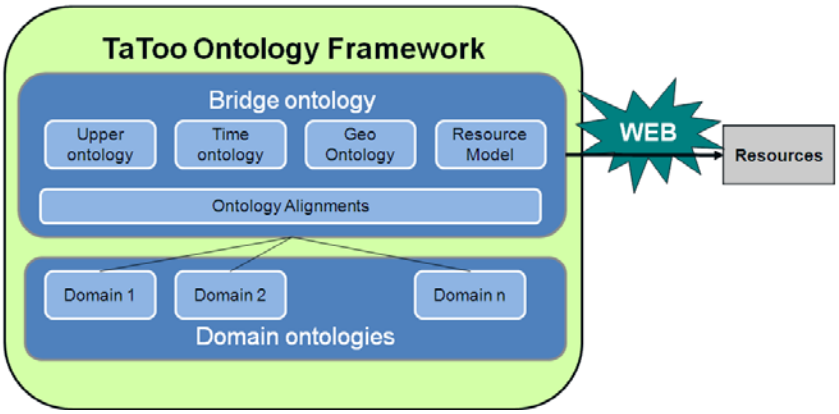

Figure 3. TaToo high-level ontology framework (Pariente-Lobo et al., 2011)

The Minimal Environmental Resource Model (MERM). MERM describes the structure of resources and annotations in TaToo, being a reference data model for both services and user interfaces. As any other ontology developed in the scope of TaToo, MERM has been developed following the NeOn methodology (Gómez-Pérez, 2010). The NeOn methodology encourages reusing knowledge as much as possible. To that extent, MERM contains, besides the POSM ontology (Pedrinaci et.al., 2010), concepts from several other ontologies like SIOC (http://www.sioc-project.org), FOAF (http://www.foaf-project.org), Dublin Core (http://dublincore.org) and O&M (http://seres.uni-muenster.de/o&m/O&M_discussion_paper.pdf). MERM is still work in progress so the addition of new models for several resources is expected during the course of the project. Furthermore, in TaToo we deal with a special kind of annotations, the so-called evaluations, which are annotations relating to another annotation instead of to a resource. Evaluations are created by users in order to evaluate how accurate or useful a resource annotation is. This information is used during the discovery process to improve discovery results and accuracy (see also Pariente-Lobo et al., 2011).

3.2. Integrating Domains

The TaToo ontology framework, as described above, comprises the bridge ontology, a number of domain ontologies, and a number of alignment ontologies. The framework is not limited by a number of different domain ontologies, that is, by a number of different environmental sub-domains whose resources are managed by the TaToo system. However, in order to be able to plug-in a domain ontology in the TaToo ontology framework, the domain ontology needs to be accompanied by an appropriate mapping interface. This mapping interface is identified in the framework as an alignment ontology.

In TaToo we use our alignment/mapping strategies (Pariente-Lobo et al., 2011) applied to three domains ontologies, which are specific to the TaToo validation scenarios. These scenarios are used as proof-of-concepts and case-studies throughout the project life. The first ontology describes the agro-environmental domain, the second ontology is related to the anthropogenic impacts of global climate change, and the third ontology contains concepts related to the comparison of climatic conditions in different regions. These three domain ontologies are considered to be an integral part of the TaToo ontology framework thus justifying the proposed ontology integration approach. Examples can be found in (Pariente-Lobo et al., 2011).

In case of domain ontologies we plan to extend them in future by adding new domain specific concepts and relationships that are currently not provided. In case of the bridge ontology we want to identify which of the

concepts from the domain ontologies can be generalized and potentially become members of the MERM ontology. Thus it allows the integration of new domains and fosters the interoperability between different domains. It can be stated that it is not needed to replace the bridge, respectively the MERM, when new domain ontologies are added. We only need to establish new alignments thus heavily reducing the required work ..

4. TATOO WEB PORTAL

The TaToo framework architecture describes a set of components (User Components), which allow the user to access all the functionality TaToo is providing. These components can be of different type, used for different purposes and in the context of different application domains. In fact, some of these tools are customised to address particular requirements from different users from specific application domains, the previously mentioned validation scenarios: in particular TaToo is considering those domains from the TaToo Validation Scenarios, e.g. Masaryk University (MU), Austrian Institute of Technology (AIT) and the Joint Research Center (JRC).

Assuring the visibility of the User Components is of vital importance for a successful and widespread user adoption of the TaToo Semantic Framework. In order to provide a general purpose tool, available on the Web with no particular restrictions, TaToo is providing, among others, a web portal: a central and user-friendly user-configurable entry point.

The TaToo web portal is able to provide all the TaToo functionality, in particular:

- searching for resources;
- adding tags and annotations to resources (discovered or directly provided by the user through an identifier – an URI);
- evaluating resources (possibly visualising them) and validating previously added tags and annotations.

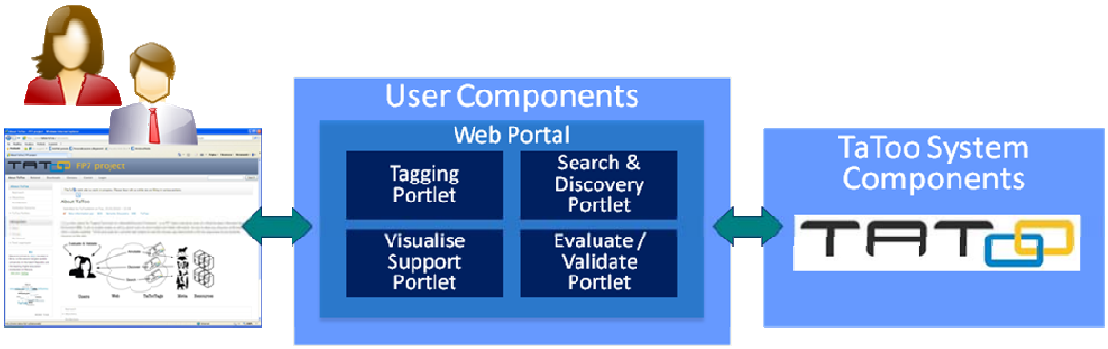

Figure 4. TaToo web portal

The TaToo web portal represents the main entry point for the TaToo semantic framework and its functionality. Due to the nature of a web portal the user has access independently of his location to all the tagging, discovery, and evaluation / validation functionality (portals benefit of Web ubiquity increasing the number of possible users or the number of contacts from a single user). It is composed of a set of portlets that match with the set of services provided by the TaToo semantic framework: tagging portlets, discovery portlets, and evaluation / validation portlets. Portlets are developed independently of the portal itself, and are loosely coupled with the portal. The idea is that a portlet implements a single functionality. Of course, more than one portlet can be developed to provide the same functionality. TaToo's architectural design identifies four basic portlets providing tagging, search and discovery, results visualisation and evaluate / validate functionality. Currently the processes for validation and ranking are under discussion. Specification and development of related functionality will be provided in 2012. All these portlets are pluggable user interfaces making up the TaToo portal interface (see Figure 4).

At present time the TaToo web portal is made of three pages referring to different thematic areas: 'Tagging', 'Discovery', and 'Masaryk University'. Of course the user can configure the portal as they like putting all the desired portlets in a single tab, adding new tabs, etc. The 'Tagging' page has been created to contain all tagging related portlets. At the present time it contains the TaToo tagging portlet (see Figure 5). The aim of this portlet is to provide generic tagging functionality. It is made of three panels: 'Resources', 'Tagging',

'Tag Details', plus a status bar ('Status'). The 'Resources' panel allows the user to provide a set of resources identified by URIs and to search for already provided tags shown in the 'Tag Details' panel. The 'Tagging' panel instead can be used to add tags to one or more resources (selected from the ones listed in the 'Resources' panel) providing properties and values basing on the domain ontology selected by the user through the specific combo box (tags are then stored as RDF triples in the TaToo semantic framework).

For the discovery two portlets are currently deployed: the Hierarchical Search and the Result Presentation portlets. The Hierarchical Search portlet is the first implemented portlet to perform discovery of resources taking advantage of metadata (tags) stored in the TaToo semantic framework knowledge base. The search is hierarchical as the portlet presents a tree where the user can navigate to search the desired concept from the relevant domain ontology. The Result Presentation portlet is used to show the found resources together with so far associated tags as the result of the user navigation.

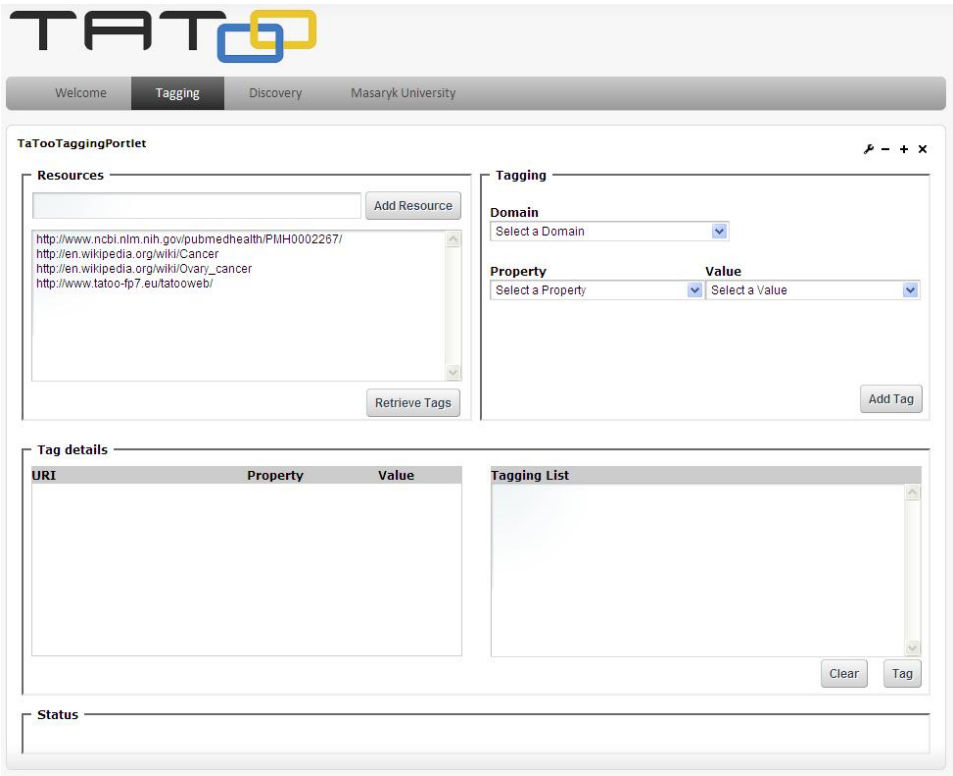

Figure 5: TaToo Portal Tagging Tab

For example the Masaryk University page has been set up to provide one of the TaToo Validation Scenario a public environment where they are the owners and that can be configured following specific scenario needs. The page contains a customisation of the Tagging portlet for the Masaryk University and represents an example of how custom User Components (in this case portlets) can be provided to address specific requirements from different user communities. In particular, with respect to the generic Tagging portlet, the customisation is located in the "Tagging" panel. Properties and values are concepts taken from the Masaryk University domain ontology and the user is guided through fixed combo boxes to properly annotate specific domain resources. E.g. the "Diagnose" label is mapped to the ontology defined property "refersToDisease" and the values are exclusively instances of the ontology class "Disease".

5. DISCUSSION AND CONCLUSIONS

The TaToo approach aims at improving the search and discovery of environmental resources by means of a collaborative and social approach based on the semantic annotation of environmental resources.

In this paper we have presented the framework architecture of TaToo and its ontological framework. The software architecture has been inspired by the Service-Oriented-Architecture paradigm, and it draws from previous experiences matured in the ORCHESTRA, a project, where a reference framework for the delivery of SOA for environmental services was developed (Uslaender, 2007). The advantage of such an architecture

is the full decoupling of the service framework from the user applications. TaToo now is developing a client application implemented as a web portal, but third parties will also be able to access the open architecture of TaToo.

The ontological framework is centered about the concept of a Minimal Environmental Resource Model. The idea is to use an ontology to define this minimal model that should be as generic as possible, in order to be shared among the various specific domains which make up the overall environmental sciences domain (e.g. air, water, groundwater, agriculture, atmosphere, hydrology, etc.). The domain specific ontologies will be connected to the MERM via bridge ontologies in order to establish links across domains.

As of writing this paper, we haven't yet explored in detail a number of highly important issues, such as the performance of various semantic search algorithms, and the scalability of the TaToo approach. Other open issues are the methodologies for collaborative ontology development, which might be needed when adding a new application domain to the TaToo ontological framework.

The project is about to enter its final stage in year 2012 and so far it is under intense development by a team of partners including large enterprises such as ATOS (Spain) and Telespazio (Italy), a small-medium enterprise such as cismet (Germany) specialized in the provision of environmental software solutions to regional and local authorities, and large and medium research institutes and universities such as AIT Austrian Institute of Technlogy (Austria), Masaryk University (Czech Republic), the Joint Research Center Ispra (Italy), IDSIA USI/SUPSI (Switzerland).

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nr. 247893.
Thanks to the TaToo Consortium for the support of this paper.

REFERENCES

Dihé P., Frysinger S., Güttler R., Schlobinski S., Petronzio L., Denzer R., Nešić S., ParienteLobo T., Schimak G., Hřebíček J., Donatelli M. (2011). An Architecture for the Semantic Enhancement of Environmental Resources in Environmental Software Systems - Frameworks of eEnvironment, 9th IFIP WG 5.11 International Symposium, ISESS 2011, Brno, Czech Republic, June 27-29, 2011, Proceedings by Jiří Hřebíček, Gerald Schimak, Ralf Denzer (Eds.), IFIP Advances in Information and Communication Technology Volume 359, pp. 372-364; DOI: 10.1007/978-3-642-22285-6; www.springerlink.com.

Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American, May 2001.

Gómez-Pérez A., Motta E., Suárez-Figueroa M.C., (2010). NeOn Methodology in a Nutshell. http://www.neon-project.org/nw/NeOn_Book, 2010.

Pariente-Lobo T., Fuentes J.M., Sanguino M.A., Yurtsever S., Avellino G., Rizzoli A.E., Nešić S. (2011). A Model for Semantic Annotation of Environmental Resources: The TaToo Semantic Framework; in Environmental Software Systems - Frameworks of eEnvironment, 9th IFIP WG 5.11 International Symposium, ISESS 2011, Brno, Czech Republic, June 27-29, 2011, Proceedings by Jiří Hřebíček, Gerald Schimak, Ralf Denzer (Eds.), IFIP Advances in Information and Communication Technology Volume 359, pp.419-427; DOI: 10.1007/978-3-642-22285-6; www.springerlink.com

Pedrinaci C., Liu D., Maleshkova M., Lambert D., Kopecky J., and Domingue J. (2010). iServe: a Linked Services Publishing Platform. Workshop: Ontology Repositories and Editors for the Semantic Web at the 7th Extended Semantic Web Conference (ESWC) 2010. Heraklion, 2010.

Usländer, T. (2007), Reference Model for the ORCHESTRA Architecture v.2.1, OGC "Best practices" paper No. 07-097, http://www.opengeospatial.org/standards/bp, (2007)

Wache H., Vogele T., Visser U., Stuckenschmidt H., Schuster G., Neumann H., Hubner S., (2001). Ontology based integration of information: a survey of existing approaches. In: Proceedings of the IJCAI„01:17th *International Joint Conferences on Artificial Intelligence*, Seattle, WA, USA, pp.108–117.