EnviroInfo 2011: Innovations in Sharing Environmental Observations and Information
Copyright 2011 Shaker Verlag Aachen, ISBN: 978-3-8440-0451-9

Applying semantics in the environmental domain: The TaToo project approach

Giuseppe Avellino[1], Tomás Pariente Lobo[2], Gerald Schimak[3], Andrea Emilio Rizzoli[4], Pascal Dihé[5]

Abstract

The EU FP7 TaToo project major aim is to bridge the resource discovery gap. This is a well know problem regarding, firstly, the difficulty of discovering resources available on the Web, secondly, the access and possibly the 'visualisation' of these resources once discovered. TaToo is trying to solve the problem through a framework (based on semantic concepts) able to store in a knowledge base, meta-information (or metadata) associated with resources. Meta-information is then used by TaToo during the resource discovery process. Meta-information is provided by end users while discovering resources or by resource owners and, as stored in a triplestore, it consists of RDF triples made of Domain Ontologies concepts. Meta-information knowledge base makes possible, together with a reasoner, a resource model for annotation, and a set of Domain Ontologies, a semantically-enhanced resource discovery process.

The paper briefly shows the TaToo architecture focusing on the description of those components dealing with semantics. Then, it shows how semantics has been implemented, giving details on the Minimal Environmental Resource Model (MERM) and on the process of integrating Domain Ontologies to realise the ontology framework used in TaToo.

1. TaToo framework architecture

In order to allow an effective semantically-enhanced discovery process, TaToo needs meta-information, a massive amount of meta-information. The discovery process relies on this meta-information, so the more meta-information is available, the more the discovery process is effective and is able to discover resources the user is really searching for. TaToo planned to collect meta-information through two main approaches: end users associating meta-information with resources they have discovered: for instance telling what the resources are about; harvesting meta-information associated with resources by the resource owners (those making resource available i.e. publishing resources). In addition to these two approaches, meta-information can also be inferred through the support of a reasoning process able to derive new knowledge out of the one already available (acquired through the previous two approaches).

TaToo architecture has been designed to address all the requirements envisaged during the requirements elicitation process, in particular to provide all the tools (e.g. GUI for tagging and discovering resources) and the core framework required to allow the semantically-enhanced discovery process.

The complete TaToo architecture for the second project iteration is presented in Figure 1. The architecture is made of five tiers (each containing one or more building blocks): 'Presentation', 'Service', 'Business', 'Data', and 'Cross'. 'Presentation' contains components providing functionalities to interface the

[1] Telespazio S.p.A., Rome, Italy, email: giuseppe.avellino@telespazio.com
[2] Atos, Madrid, Spain, email: tomas.parientelobo@atosresearch.eu
[3] Austrian Institute of Technology, Seibersdorf, Austria, email: gerald.schimak@ait.ac.at
[4] IDSIA, Lugano, Switzerland, email: andrea@idsia.ch
[5] cismet, Saarbrucken, Germany, email: pascal.dihe@cismet.de

end user with the system: Tagging Tools, Search & Discovery Tools, and Evaluate / Validate Tools. In general, these components are supposed to be directly used by the end user and can be implemented as portlets (parts of a portal), external applications, browser plug-ins or APIs. The ‘Service’ tier provides web services offering interfaces to the underneath TaToo components. The ‘Business’ tier contains 'core' components implementing the TaToo business logic. In particular:

- The Clearinghouse: the central component for accessing the TaToo framework serving as an information exchange support between the core system components;
- The Semantic Processor: the main component dealing with semantics. It takes advantage of the TaToo ontology framework to provide functionality based on semantics. Basically, it is made of an application framework and a reasoner;
- The Resource Harvester: the component capable of retrieving external resources that could be either data or associated meta-information stored in catalogues, Web services or information contained in Web pages.

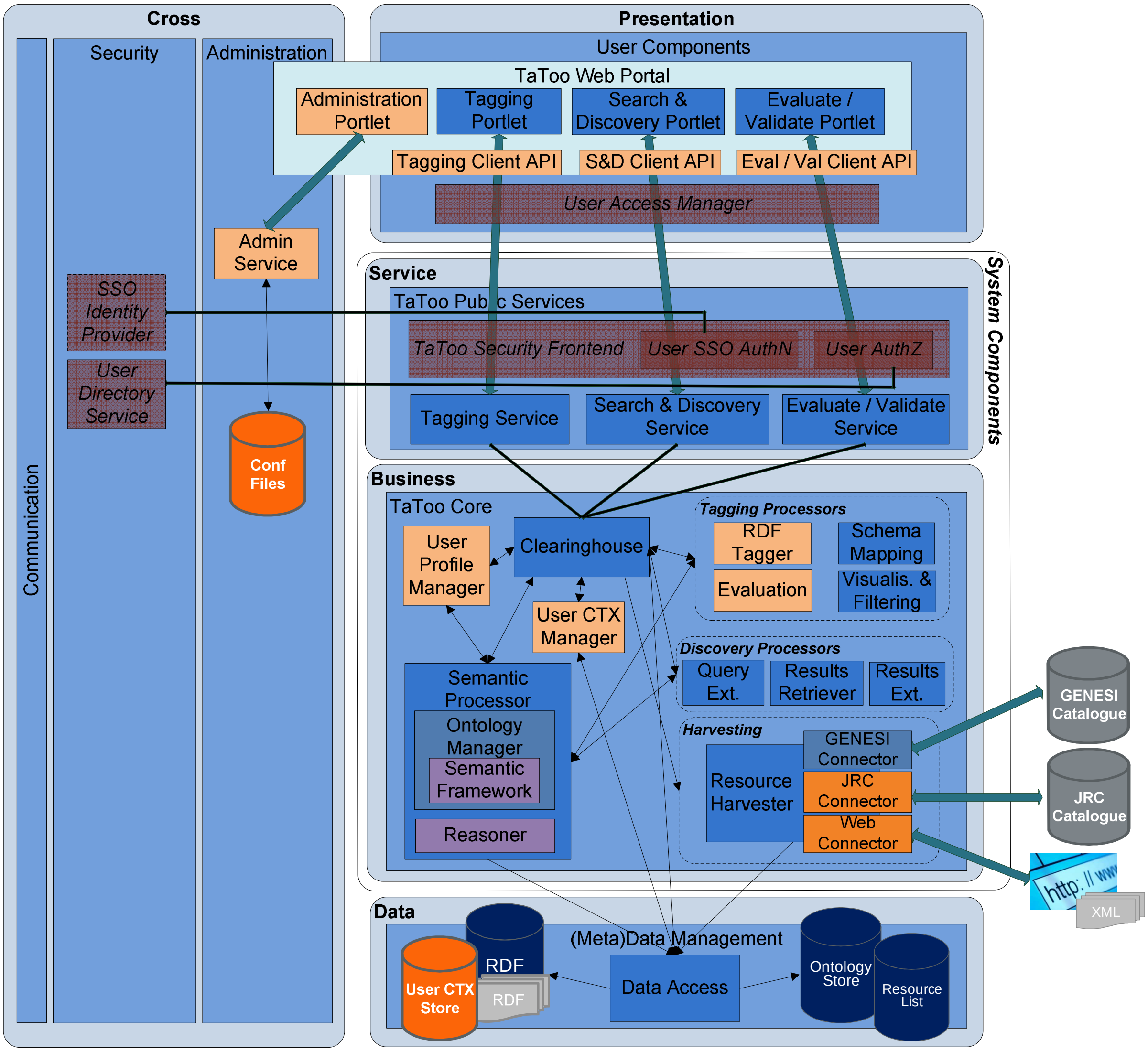

Figure 1 TaToo architecture overview

Copyright 2011 Shaker Verlag Aachen, ISBN: 978-3-8440-0451-9

The ‘Data’ tier deals with the storage functionality in general (meta-information, list of sources to be harvested for meta-information, user context, and so on).

Finally, the ‘Cross’ tier provides cross-cutting functionality. This tier foresees a security framework for securing access to TaToo functionalities and stored data, and a set of administration functionalities.

1.2 TaToo components dealing with semantics

In its architecture TaToo identifies a set of components dealing with semantics. These components are essential to compose annotations based on RDF triples made of Domain Ontologies concepts, and to make the semantically-enhanced discovery process possible. Components dealing with semantics are in principle all part of the ‘Business’ tier and in particular part of the TaToo Core building block.

The Semantic Processor is the main semantic component. It deals with ontologies (in particular with the entire TaToo ontology framework): takes care of the functionalities related to retrieving / storing / loading TaToo ontologies, stores RDF triples in the triplestore as part of the knowledge base (through the Data Store), makes possible the reasoning on the knowledge base. To be able to operate, it is made of an Ontology Manager dealing with all the ontology related stuff (taking advantage of a library providing semantic framework functionality such as APIs for manipulating RDF, supporting SPARQL query, and others), and a reasoner (an external and already available library for reasoning on the knowledge base).

To identify already available products, components, or technologies in general to be adopted in TaToo, a SotA analysis has been performed. It resulted in the identification of some technologies used in the context of TaToo. Several semantic frameworks, reasoners, and triplestores have been considered such as Jena[4], Redland[5], RDFSuite, RDF2GO, Pellet, KAON2[6], OWLIM[7], Sesame[8] (OpenRDF). The analysis results brought to the adoption of Sesame as semantic framework / triplestore, and OWLIM TRREE as reasoner.

Figure 2 shows how the Semantic Processor (Semantic Framework plus Reasoner) and the RDF store (triplestore) have been implemented through the adoption of Sesame plus OWLIM. The choice has been motivated by the fact that Sesame plus OWLIM seems to be the fastest and the more scalable available solution allowing storage plus reasoning. Other projects have adopted the same solution (e.g. SOA4ALL[9]). Sesame is a well-known semantic framework for most of the people related to the semantic stuff and the installation and configuration procedures are quite straightforward. Furthermore, Sesame provides a web console and APIs for configuration allowing, among the others, the configuration of the type of persistency desired for the RDF (in memory, file system, database). The OWLIM reasoner (TRREE) has a good level of support of RDFS, OWL Horst, and OWL 2 RL, yet not complete support for OWL-DL. Even if OWLIM is limited to a certain level of expressiveness in the ontology (ontology language support), it is rather good to achieve TaToo objectives. Very often, a little of semantics can do the job without the needs of really expressive languages, which also entail really high elaboration time for reasoning and inference (low performance). The combination of Sesame plus OWLIM allows TaToo to have a good coverage of the functionalities to be offered by the Semantic Processor, the Ontology Manager, the reasoner and the triplestore.

---

[4]http://jena.sourceforge.net/
[5]http://librdf.org/
[6]http://kaon2.semanticweb.org/
[7]http://www.ontotext.com/owlim/
[8]http://www.openrdf.org/
[9]http://www.soa4all.eu/

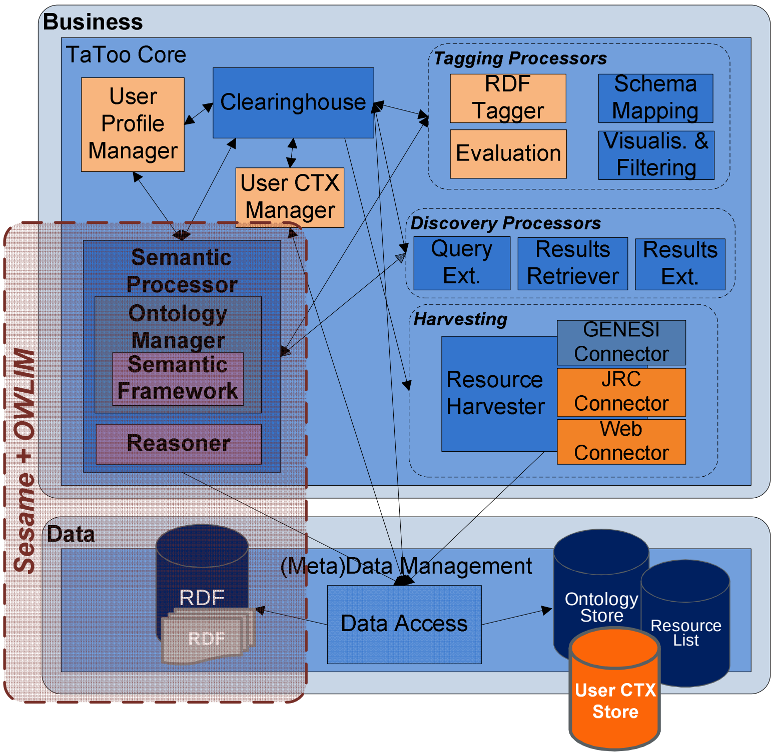

Figure 2 TaToo architecture core components

In addition to the Semantic Processor, which plays the major role, other components provide functionalities supporting the semantics adoption in TaToo. Of course, the 'Presentation' tier tools, interacting with the end user, have to manage domain ontologies to present the end user with concepts they can use for providing annotations (tags). Presentation of domain ontologies concepts can be processed through filtering functionalities provided by the Visualisation & Filtering component. The Schema Mapping component is provided with the business logic required for converting received triples in the proper RDF adhering to the structure adopted for the annotations defined in TaToo by the MERM. The RDF Tagger takes care of getting RDF triples from the Presentation tier tools to store them in the triplestore (possibly converting them through the Schema Mapping Component). Analogous functionality as the Schema Mapping component has the Resource Harvester component, which gets RDF from the various connectors converting it before storing in the triplestore.

3. Semantics in TaToo

TaToo semantic components have been briefly presented from an architectural point of view. The aim of this section is to go in depth semantics describing how ontologies together with MERM make the TaToo ontology framework and how this framework is used in TaToo.

3.1 Use of ontologies

Search (discovery) can be broadly defined as the retrieval by a system of a set of resources that satisfy a user need. TaToo relies heavily on the concept of semantic search. Applying semantic technologies to search for resources usually means first to describe the resources using meta-information (or metadata) and later to apply semantically-enhanced algorithms to the underlying meta-information dataset. This meta-information is composed by formal tags associated with the resources coming from existing ontologies. Ontologies can be defined as explicit and shared conceptualizations of a domain and provide additional reasoning power to the search.

There are several approaches to semantically describe resources or data. The approach followed by the Linked Data[10] initiative consist on publishing the resources by actually adding a separate layer of meta-information (RDF) that describe the resource and points to it following the so-called Linked Data principles. In Linked Data the resources are harvested and published in datasets that can be linked to other existing datasets thus allowing the search by navigation paradigm. Another approach consists of tagging the resources using terms from existing and shared ontologies. While in Linked Data the objective is publishing and linking entire datasets, tagging using ontologies aims at gathering a minimal set of annotations that help on the search process. TaToo is following this last approach. In this sense, it is worth noticing that TaToo does not gather environmental-related resources from external repositories, but rather meta-information about those resources. The TaToo approach is therefore entirely based on applying semantic search techniques over a controlled set of meta-information, using the ontology corpora defined within the scope of the TaToo framework, which can be extended for future domains.

Ontology development is a very complex and time-consuming activity. It involves several actors, ranging from domain experts to ontology developers, which have to apply their respective skills to come up with a suitable ontology that describes one domain for specific needs and requirements. In the last years several ontologies have emerged in multiple domains that can be of interest of TaToo. However, there is no a single ontology that fits it all in the environmental domain. The approach followed in TaToo has been to generate a set of cross-linked domain ontologies to some simple upper-level concepts in order to facilitate the cross-domain tagging and search. Adding new domains to this ontology network consists of align the new ontology to the upper elements to connect to existing network. The main ontology in this upper layer is baptised within TaToo as the Minimal Environmental Resource Model (MERM).

3.2 The Minimal Environmental Resource Model

TaToo has defined MERM as an ontology that acts as the backbone for the annotation process. Essentially MERM consists of basic properties for describing a resource and common properties for the annotation of resource types. Conceptually, MERM is an effort to identify a minimal model that can be used to typify the annotations of a given resource, pointing to other domain elements. In this sense, it acts as a connector between the resource meta-information structure and the domain topics or tags identified during the annotation process.

[10] http://linkeddata.org/

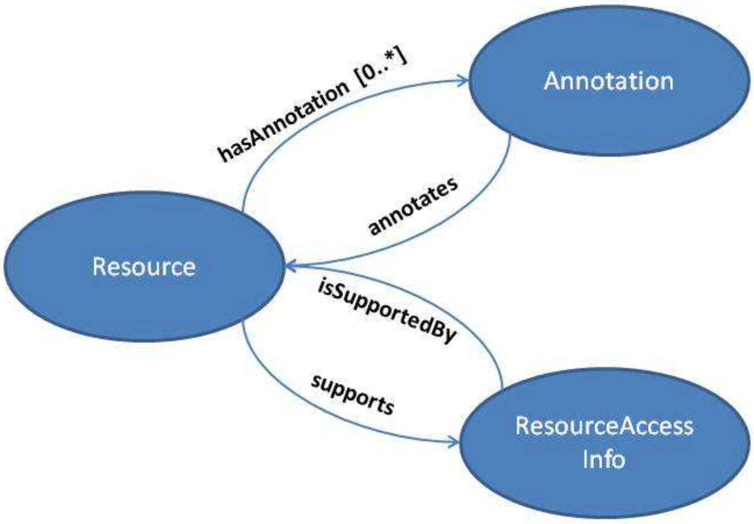

Figure 3 MERM core classes and relations

MERM has been developed following the NeOn methodology (Gómez-Pérez 2010). MERM reuses parts of other well-known ontologies and vocabularies, such as for instance SIOC[11], FOAF[12], Dublin Core[13] and O&M[14].In Figure 3 a reduced view of MERM is shown.

The Resource class represents the meta-information about a resource in TaToo, including properties to define author, owner, provider, date of creation, etc. A Resource could have (but it is not mandatory in TaToo) access information presented in the ResourceAccessInfo class. This information could be very heterogeneous ranging from a simple URL to a complex WSDL depending on the nature of the resource. Last but not least, the annotations or tags related to a specific resource are individual of the Annotation class. An annotation describes what a resource is about and the link to the domain topics. The Annotation class has been sub classified in several annotation types widely used in the environmental domain (web document, web service, time series, etc.). The model is extensible and can be further tailored for other annotation types and reused in other domains.

3.3 The TaToo ontology framework and knowledge base

As explained before, TaToo bridges different domains by using upper level common elements and the MERM ontology. This approach is inspired in the hybrid ontology approach as described by Wache (Wache 2001), as shown in Figure 4.

---

[11]The SIOC initiative (Semantically-Interlinked Online Communities), http://www.sioc-project.org/

[12]The Friend of a Friend (FOAF) project, http://www.foaf-project.org/

[13]Dublin Core metadata iniciative, http://dublincore.org/

[14]Observations and Measurements ontology, http://seres.uni-muenster.de/o&m/O&M_discussion_paper.pdf

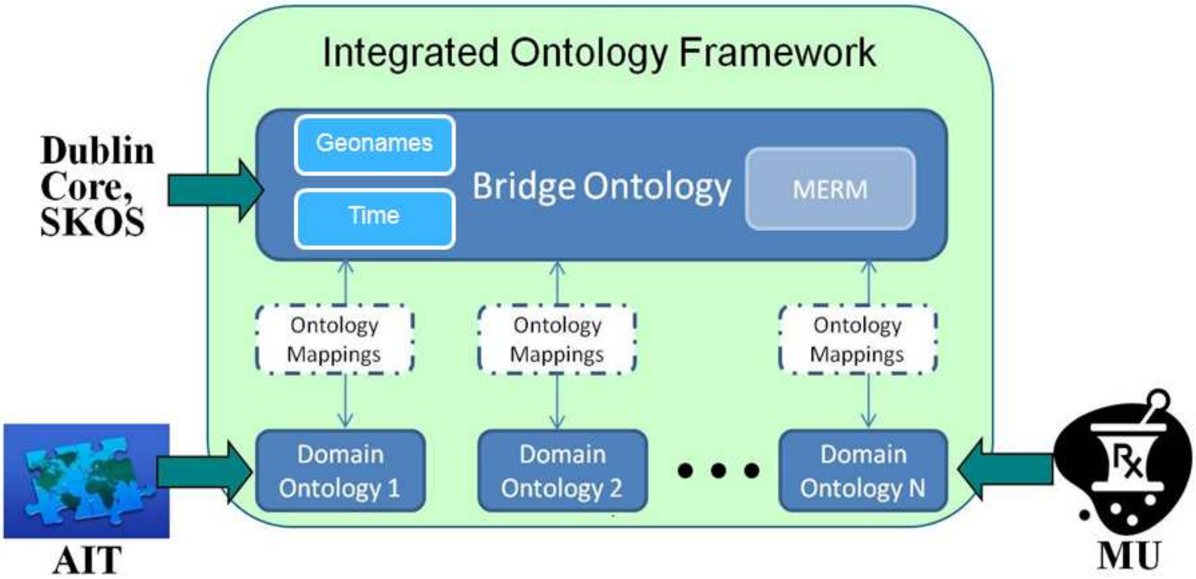

Figure 4 TaToo high-level ontology framework

The TaToo Bridge ontology is an upper layer of the TaToo ontology framework including elements from widely adopted ontologies. For now, the Bridge ontology imports by default the MERM ontology defined within TaToo, the GeoNames[15] ontology for describing geographical locations, and some parts of the OWL Time[16] ontology from the W3C for time handling. It also uses parts of Dublin Core and SKOS[17]. As depicted in the previous figure, every domain ontology has to be aligned to the bridge ontology in order to be part of the TaToo ontology framework by creating ontology mappings to the bridge elements. The mapping process can be done manually by defining equivalency and subsumption relationships between concepts, object properties, data properties and individuals of bridge ontology and the domain ontology. Automatic or semi-automatic ontology alignment tools can also be used to ease the mapping process.

The TaToo resource tagging process consists of the creation of annotations based on the existing ontologies as set of RDF triples. These triples are stored in a knowledge base (the TaToo semantic processor) that is implemented as a Sesame triplestore repository plus the OWLIM, which is a semantic framework that relies on Sesame and provides fast reasoning and SPARQL extensions for geo-spatial queries.

4. Conclusions

Bridging the resource discovery gap is challenging. TaToo effort is focused on this, designing and implementing a framework able to achieve it through the adoption of semantics. An approach based on semantics where resources are annotated through terms from domain ontologies has been chosen. The effectiveness of the proposed approach will be verified promoting the tagging of resources (by end user and resource owners), in order to count on a large amount of meta-information. Domain ontologies have been

[15]http://www.geonames.org/
[16]http://www.w3.org/TR/owl-time/
[17]http://www.w3.org/2004/02/skos/

developed to conceptualise domains from three validation scenarios: Climate change twin regions from Austrian Institute of Technology (AIT); Agro-environmental management from Joint Research Centre (JRC); Anthropogenic impact and global climate change from Masaryc University (MU). These ontologies will be extended and modified to better address validation scenarios needs for resource discovery. As TaToo allows plugging in new ontologies to the ontology framework, effective discovery of resources based on validation scenarios ontologies will be a visible result for stating the validity of the proposed solution.

Acknowledgment: The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nr. 247893T

Bibliography

Dihé P., Avellino G. (et al.): D3.1.2 TaToo Semantic Service Environment and Framework Architecture – V2. TaToo second year project deliverable

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

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

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

Rizzoli A.E, Schimak G. (et al.) (2010): TaToo: Tagging environmental resources on the web by semantic annotations; Proceedings of International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.) http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Schimak, G., Rizzoli, A.E., Avellino, G., Pariente Lobo, T., Fuentes, J.M., Athanasiadis, I. (2010): Information enrichment using TaToo's semantic framework; Proceedings of Metadata and Semantics Research (MTSR) Conference 2010, Alcala de Henares, Spain, S. Sanchez-Alonso, I. N. Athanasiadis, E. García-Barriocanal, N. Palavitsinis, http://www.ieru.org/org/mtsr2010/