International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Semantic annotation of environmental resources: Does it really matter to resource users?

Saša Nešić[a], Andrea E. Rizzoli[a], Gerald Schimak[b] Giusepe Avelino[c], Jiří Hřebíček[d], Alexander Kaufmann[e], and Marcello Donatelli[f]

[a] *IDSIA - Dalle Molle Institute for Artificial Intelligence, Lugano, Switzerland (sasa@idsia.ch, andrea@idsia.ch)*
[b] *AIT Austrian Institute of Technology GmbH, Vienna, Austria (gerald.schimak@ait.ac.at)*
[c] *Telespazio S.p.A. - A Finmeccanica/Thales Company, Rome, Italy (giuseppe.avellino@telespazio.com)*
[d] *Masaryk University, Brno, Czech Republic (hrebicek@iba.muni.cz)*
[e] *AIT Austrian Institute of Technology GmbH, Vienna, Austria (alexander.kaufmann@ait.ac.at)*
[f] *IES-Agri4cast, Joint Research Centre, Ispra, Italy (marcello.donatelli@jrc.ec.europa.eu)*

Abstract:

Semantic annotation of environmental resources has become a reality attested by a significant amount of annotations created and published on the Web recently. A key question now is how to make practical and effective use of those annotations to aid users of environmental resources perform their activities more efficiently and effectively. In this paper we present and discuss results of an expert evaluation, which tried to address some aspects of this relatively complex question. The evaluation was conducted on the example of the TaToo semantic annotation and discovery approach. The overall results showed a positive attitude towards the usefulness of the semantic annotations and their potential to improve not only the discovery of environmental resources, but various aspects of their usage.

Keywords: Semantic annotation, resource discovery, expert evaluation.

1 INTRODUCTION

Semantic annotations of environmental resources, especially those complying with formal descriptions such as ontologies, represent conceptualized knowledge about the resources, their properties, and relationships among them (Uren et al., 2006). A number of approaches aiming at modelling and annotating environmental resources by adding formal semantics (e.g., metadata, ontologies, thesauri, and taxonomy) to them has been proposed over the last decade (Villa et al., 2009; Janowicz et al., 2010; Sicilia, 2006). Some of those approaches have been implemented, deployed, and exploited in real world scenarios producing significant amounts of semantic annotations. This process is currently ongoing and it is gaining momentum, therefore, the initial issue of how providing a sufficient amount semantic annotations for environmental resources is no more critical.

A key issue that needs to be properly addressed now is to discover how to get the most valuable benefits from the semantically annotated environmental resources. The knowledge provided by semantic annotations might be used not only for resource discovery but also to help environmental researchers (e.g., environmental policy and decision makers) performing their tasks more efficiently and effectively.

In this paper we state that to answer that question we must first identify the added value of semantically annotated resources to the resource users. We need to discover how and where the impact of using semantic annotations can be maximised. This can be achieved by asking proper questions to environmental researchers, asking them questions such as: Can semantic annotations shorten the time required to discover a resource? Do semantic annotations provide useful information about resources and their properties? Is information stored in semantic annotations easily interpretable and reusable? Can semantic annotations automate some environmental processes such as linking models, and preparing data presentations? Answering on this type of questions is not easy and it requires a comprehensive evaluation of not only the various semantic annotation approaches, but also of the software system implementing them.

In this paper, we present such an evaluation that we conducted on the TaToo semantic tagging and discovery approach (Pariente et al., 2011) and the corresponding software system implementing it (Dihe and et al., 2011). The evaluation engaged environmental experts and applied well-known evaluation methods such as a questionnaires and focus groups. We organize the rest of the paper as follows. In Section 2 we first outline main aspects of the evaluated TaToo approach. Section 3 is devoted to the description of the evaluation objectives, evaluation design and applied evaluation methods. Section 4 presents and analyzes collected evaluation results. We conclude the paper by discussing the evaluation outcomes and providing some summary remarks.

2 TAGGING AND DISCOVERY OF ENVIRONMENTAL RESOURCES: THE TATOO APPROACH

The TaToo (Tagging Tool based on a Semantic Discovery Framework) is an approach for tagging/annotation and discovery of environmental resources that uses formal semantics defined in domain ontologies. The approach provides an environmental resource representation model, called MERM (Minimal Environmental Representation Model), that enables a number of ontological concepts from various domain ontologies to be added to a resource description (Pariente et al., 2011). Such added ontological concepts play a role of formalized semantics of the resource and can be used later on for the resource discovery. Resource descriptions, that is, instances of the MERM model are actually RDF (Resource Description Framework) graphs represented as sets of RDF triples. Moreover, the TaToo approach introduces an OWL- and SKOS-based ontology framework that allows different environmental domain ontologies to be aligned and used together when annotating environmental resources. In other words, the framework provides mechanisms for aligning the same or related concepts that are defined in different domain ontologies. Finally, the TaToo ontology framework comes together with a predefined set of three domain ontologies (an agricultural ontology, a climate ontology, and an ontology modeling anthropogenic impacts on the environment) that correspond to three validation scenarios designed to validate the TaToo approach. As we mentioned above, one of the main benefits from added ontological annotation would be to improve the resource discoverability. The TaToo approach also provides the resource discovery strategy that is based on the use of accumulated resource annotations and the reasoning performed at the level of underlying domain ontologies.

The TaToo approach is delivered by a software system, which according to the approach was named the TaToo system (Dihe and et al., 2011). The TaToo system is designed as a SOA (Service Oriented Architecture) system, which is based on the top of an RDF knowledge base (i.e., RDF repository) that stores the TaToo resource annotations. The system's services are organized into two layers: a core layer that provides implementation of the TaToo tagging and discovery functionalities, and a public service layer that provides user access to these functionalities. Considering the SOA nature of the system, a variety of users (i.e., client) applications and tools can be developed to let different user groups benefit from the TaToo tagging and discovery. Moreover, the TaToo system provides a Web portal consisting of a number of customizable portlets. In addition, it is

worth mentioning that the TaToo knowledge base is open to store annotations for any uniquely identified environmental resource regardless of the actual resource location.

3 THE TATOO EXPERT EVALUATION

We conducted the evaluation of the TaToo approach and the corresponding TaToo system in a form of an expert evaluation (Neilsen, 1994). Two main objectives of the evaluation were: 1) to evaluate the TaToo approach for semantic tagging and discovery of environmental resources, and 2) to evaluate the usability of the TaToo system and identify potential weaknesses. Besides these two main objectives, the TaToo expert evaluation was expected to help us in defining priorities for the following developments on the TaToo project. The evaluation was conducted as a part of the TaToo evaluation workshop that was organized during the ISESS (International Symposium on Environmental Software Systems) conference. In the rest of this section we first discuss the structure of the TaToo evaluation workshop and then describe the evaluation questionnaire. In the following section (Section 4) we present and analyze the evaluation results.

3.1 Structure of the TaToo evaluation workshop

The TaToo evaluation workshop was organized as a whole-day event consisting of the following sessions: 1) a conference parallel session devoted to the presentation of the TaToo approach; 2) a poster session describing various TaToo features and applications; 3) an interactive demonstration of the TaToo system; 4) filling the evaluation questionnaire; 5) a follow-up discussion session.

In the *presentation session*, which opened the evaluation workshop, TaToo project members provided theoretical background of the TaToo semantic tagging and discovery approach and they also discussed the architectural design of the TaToo system. The session was divided into 5 thematic talks, all of which were included in the proceedings of the ISESS conference (Nesic et al., 2011; Pariente et al., 2011; Dihe and et al., 2011; Kubasek et al., 2011; Ungar et al., 2011). Those talks reported most of the results achieved during the first year of the TaToo project. They provided a clear picture of what had been planned and what was realized during the first year, including the development of the MERM model and the domain ontology framework (i.e., domain ontologies of the TaToo validation domains), the design of the TaToo system architecture, the design of the TaToo validation scenarios, and the implementation of the TaToo system prototype.

The *poster session* was organized as a part of the main poster session of the ISESS conference. We prepared aTaToo poster and set up a looping PowerPoint slideshow, which were presented not only to the TaToo workshop participants but to all participants of the conference. By providing a brief summary of the TaToo approach the poster and slideshow were supposed to raise potential interests of the conference participants and attract them to take part into the TaToo expert evaluation.

The objective of the *demonstration session* was to demonstrate the functionalities of the TaToo system. In total, there were four demonstrations performed. The first one, also considered as introductory demonstration, was dedicated to the demonstration of the TaToo public services and the TaToo Web portal. The following three were dedicated to the demonstration of the TaToo tools implemented to support specific use-cases of the three TaToo validation scenarios. They included: 1) Climate Twins Application developed by the Austrian Institute for Technology (AIT), 2) the Anthropogentic Impact Validation Scenario by Masaryk University (MU), and 3) the Agro-Environmental Resource Publisher application, developed by the Joint Research Centre (JRC). All the demonstrations were performed in form of live demos.

The *questionnaire session* was organized straight after the demonstrations, thus enabling the participants to provide an immediate feedback. The questionnaire was anonymous and time unlimited. The questionnaire was followed by the *discussion session*, organized

through several focus groups. The objective of this session was to let the TaToo project members discuss face-to-face with environmental experts about the TaToo semantic tagging and discovery approach and the system. The intended outcome was to identify potential limitations and weaknesses of the work done as well as to collect suggestions and recommendations for the upcoming development.

3.2 The TaToo evaluation questionnaire

The questionnaire focused on the two main objectives of the evaluation, that is, the evaluation of the TaToo tagging and discovery approach and the evaluation of the TaToo system (i.e., the TaToo services and tools). In addition to those two groups of questions, the questionnaire also included a set of "background questions" whose purpose was to obtain information about the evaluators' experience, familiarity, and involvement in domains of interest for the evaluation. In particular, the background questions were related to the participants' involvement in the environmental domain, which is the main domain of application of the TaToo approach, and the Semantic Web domain, which is a domain providing technologies and formalisms used in the realization of the TaToo approach. However, in order to keep focus on the main evaluation objectives we tried to limit the number of background questions. We now describe the evaluation aspects considered in both evaluation objectives.

Questions on the TaToo semantic tagging and discovery: This section of the questionnaire was designed to collect the participants' feedback about the TaToo approach as well as provide suggestions/recommendations of how to further enhance it. We asked 9 questions in total, 8 of which were multiple-choice questions and 1 was an open-ended question to let the participants express some additional thoughts and suggestions. All the multiple-choice questions were designed as positive statements to be rated according to a five-point Likert scale (*Strongly Disagree*, *Disagree*, *Uncertain*, *Agree*, *Strongly Agree*). The following are four aspects of the TaToo semantic tagging and discovery approach that we were focused on in the questionnaire:

- novelty of the approach,
- scalability of the approach across the environmental domain,
- reusability of the TaToo added semantic annotations (tags), and
- potential to improve the environmental resource discovery.

In addition, we also wanted to get the participants' opinion on the way the three TaToo validation scenarios utilize the TaToo approach in their use-cases, and can the validation scenarios provide a satisfactory validation of the TaToo approach. In particular, we asked the participants to evaluate the TaToo tagging and discovery use-cases within respect to each of the three validation scenarios.

Questions on the TaToo system: The purpose of this set of questions was to provide a preliminary evaluation of the usability of the TaToo system and to help us prioritize features to be realized in the next version of the system. Considering that the participants did not test the system on their own but watched the demonstrations, the usability related questions were rather general than specific. All together, we asked 6 questions, four of which concerned the usefulness of the TaToo system and two were about the user satisfaction with the system's functionalities. Moreover, we also wanted to know the participants' opinion about features that we might consider in the future development of the system. Therefore, we asked the participants to assign their priorities to the following list of features:

- multilinguality,
- more advanced user interface of the TaToo portal,
- TaToo add-ins for conventional Web browsers, and
- additional, client-side tools and applications.

We concluded this section of the questionnaire by an open-ended question that enabled the participants to write additional suggestions of how to further enhance the existing functionalities of the TaToo system as well as to propose new functionalities.

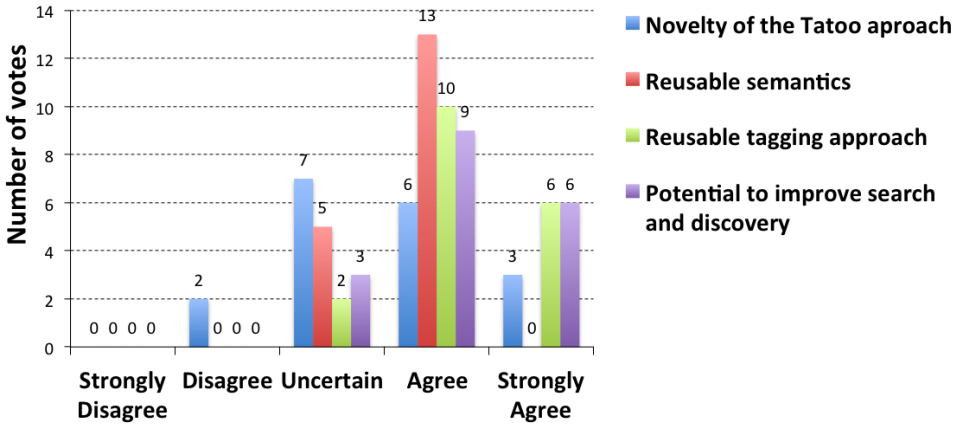
4 Evaluation Results

A total of 18 participants took part in the TaToo expert evaluation and completed the questionnaire. Considering the participants level of domain expertise, 6 participants declared themselves to be environmental experts, there were 10 participants who had done some research in the domain, and 2 said they were aware of the main research topics in the domain but had never contributed to any of them. Moreover, since the TaToo approach is based on the application of the Semantic Web technologies (e.g., ontologies, RDF and SPARQL) to the environmental domain, we wanted to know about the participants familiarity with these technologies as well. The questionnaire showed that all of the participants had heard about them and 8 participants had also used some of them. There were 2 participants who declared themselves as Semantic Web experts and 1 who was involved in the development of some semantic Web technologies. Taking all this into account, we could state that we had a set, albeit small, of very qualified evaluators with respect to both expertise in the environmental domain and familiarity with the semantic Web technologies.

4.1 Evaluation results obtained through the questionnaire

As described in Section 3.2, the questionnaire considered the following four aspects of the TaToo semantic tagging and discovery approach: the novelty, the scalability, the reusability of the semantics, and the potential to improve the discovery of environmental resources. Figure 1 illustrates the questionnaire results for each of them.

Figure 1: Semantic tagging and discovery

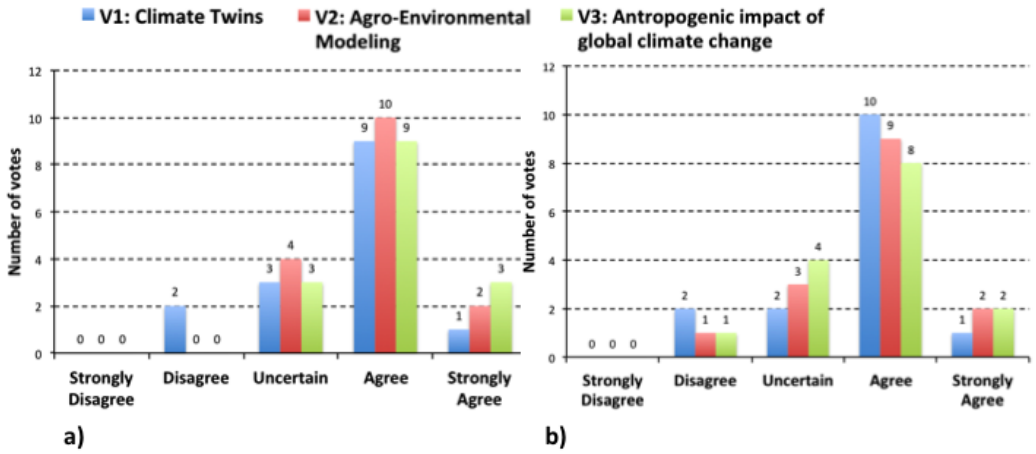


Regarding the novelty of the TaToo approach, the questionnaire showed that the participants opinions were divided but still most of them (6 Agree and 3 Strongly Agree) were positive. We interpret a high uncertainty among the participants answers (9 Uncertain) as a fact that most of them had in mind a number of similar approaches, which utilize/apply the semantic Web technologies to solve different issues in the environmental domain. However, according to our knowledge, none of the existing approaches provides an integrated framework for semantic annotation, search and discovery of environmental resources as it is proposed by the TaToo approach. Regarding the approach scalability across the environmental domain, the reusability of the TaToo semantics, and the potential of the approach to improve the search and discovery of environmental resources, the results were quite positive having Agree and Strongly Agree as dominant answers.

Furthermore, we evaluated the TaToo approach by considering its application in concrete

environmental sub-domains such as the domains of the three TaToo validation scenarios (i.e., V1: Climate Twins, V2: Ago-Environmental Modeling and V3: Anthropogenic Impact of Global Climate Change). We actually wanted to see if the TaToo tagging and discovery workflows were satisfactory for these different domains, thus being able to generalize the conclusions afterwords. As we can see from Figure 2, the results for all the three validation scenarios regarding both tagging (Figure 2a) and discovery (Figure 2b) were highly positive.

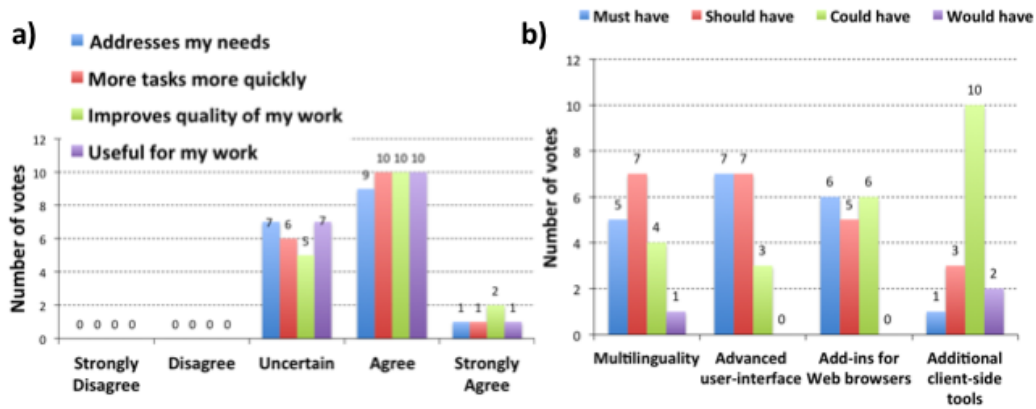
Figure 2: Semantic tagging a) and discovery b) in the TaToo validation scenarios



The questionnaires section dedicated to the evaluation of the TaToo system was composed of three parts: the questions related to the system usefulness, the questions related to the user satisfaction, and the questions related to priorities for the future development of the system. By the time of the evaluation the TaToo system prototype was operational but not feature-complete, so that we decided to demonstrate the system functionalities by performing a set of live demos instead of conducting real usability tests with end users. This enabled us to get some preliminary feedback on the TaToo system.

Regarding the system usefulness we asked the participants to rate the following four statements: 1) TaToo system has potential to address my work-related needs, 2) TaToo system could enable me to accomplish tasks more quickly, 3) TaToo system could improve the quality of my work, and 4) Overall, I find the TaToo system useful for my work. The overall results (Figure 3a) were quite positive in a sense that none of the four statement got any Strongly Disagree and Disagree as well as all of them had Agree the most.

Figure 3: a) Usefulness of the TaToo system; b) Priorities for the TaToo future development



The user satisfaction, was evaluated by two statement: 1) I will recommend the TaToo Services and Tools to my collaborators, and 2) Overall, I am satisfied with the functional-

ities provided by the TaToo system. The results were mostly positive although with some degree of uncertainty. We explain that uncertainty by the fact that the participants did not directly interact with the system, but watched the demonstrations.

A concluding set of questions in the questionnaire was dedicated to prioritizing features for the future development. We considered the following four features: 1) multilinguality, 2) advance user-interface of the TaToo Portal, 3) TaToo add-in extensions for conventional web browsers, and 4) additional client-side tools and applications. The participants were asked to assign priorities to the listed features by applying a four-level, Mast-Should-Could-Would rating scale. By contrasting the collected results for all four features (Figure 3b) we can say that the participants set slightly higher priority to the first two (i.e., mutilinguality and advance user-interface of the TaToo Portal) comparing to the third one, (i.e., TaToo add-in extensions for web browsers), whereas the priority of the fourth one (i.e., additional client-side tools and applications) was far lower.

4.2 Evaluation results obtained through discussions in focus groups

After the completion of the questionnaire all interested participants were invited to an open discussion about different aspects of the TaToo approach and the system. The idea was to divide the participants in several focus groups of domain experts and talk with them in order to get their opinions, criticisms, suggestions and recommendations. Here after we sum up key findings from the discussion.

- Ontologies should be applied in a way so that everybody (including those who even do not know what is an ontology) can use them.
- The TaToo system should provide a software support for a domain-ontology mapping to the TaToo ontology framework that would enable end-users to add and use their own domain ontologies within the system.
- The TaToo System should provide support for collaborative activities trough mechanisms such as social networks. Community members should be able to rate/evaluate resources as well as contact resources providers.
- Collecting/adding meta-data to resources should be at least semi-automated. Initial set of meta-information should be automatically devised. It should also trigger an information enrichment cycle including user-added metadata.
- The TaToo approach should be inline with the linked data principles, which among others requires the TaToo managed data to be linked to other data available on the Web in order to provide right context.
- Besides providing resource annotations (meta-information), the TaToo System should provide mechanisms for accessing and browsing/viewing actual resources from their original locations in cases in which it is possible.

5 Discussion and Conclusions

First of all, taking into account all aspects of the conducted evaluation, in our opinion the environmental domain (i.e., environmental communities) has become highly familiar, if not deeply aware, of the Semantic Web technologies and formalisms. Terms like 'semantic annotation', 'metadata', 'ontologies', and even more technical such as 'RDF', 'OWL', and 'SPARQL' have become quite common in environmental communities. Moreover, the use of semantic annotations for the resource description and the resource discovery later on, was quite clear and did not need a particular justification.

Considering the evaluation of the TaToo tagging and discovery approach, we could say that the possibility of having reusable semantic annotations across the environmental information space as well as the scalability of the approach across the different environmental domains attracted the most attention and received a highly positive feedback. We remind here that these two characteristics of the approach were enabled by the use of standardized Semantic Web formalisms such as RDF, OWL, and SKOS and the introduction of the extendable ontology framework.

Considering the usability of the TaToo system, the results showed that the system has the potential to support environmental researchers to execute their tasks more effectively and efficiently. However, a lack of multilinguality at the level of ontology representations as well as the user interface level could considerably diminish expected benefits. Accordingly, we set multilinguality and an advanced user interface that supports it as our priorities for the following project developments.

In addition, through the discussion with the evaluators several new issues were triggered, such as hiding a complexity of the ontology framework from the users, providing a semi-automated annotation support to avoid a cold start (i.e., absence of resource annotation at the beginning), providing a user interface support for aligning domain ontologies, and introducing semantic links between resources in addition to the semantic annotations.

To sum up, the evaluation results clearly showed that semantically annotated environmental resources have potential to aid environmental resource users when performing different environmental activities. Accordingly, the environmental domain is ready for the next step, that is, the exploitation of semantically enriched environmental resources.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the EC's 7th Framework Programme (FP7/2007-2013) under Grant Agreement Number 247893.

REFERENCES

Dihe, P. and et al. An architecture for the semantic enhancement of environmental resources. In *Proceedings of the 9th IFIP WG 5.11 ISESS 2011*, volume 359, pages 372–384. Springer, 2011.

Janowicz, K., S. Shade, A. Broring, and C. Stasch. Semantic enablement for spatial data infrastructures. *Transactions in GIS*, 14(2):111–129, 2010.

Kubasek, M., J. Hrebicek, J. Kalina, L. Dusek, and I. Holoubek. Semantics annotations of ontology for scenario: Antropogenic impact and climate change issues. In *Proceedings of the 9th IFIP WG 5.11 ISESS 2011*, volume 359, pages 407– 418. Springer, 2011.

Neilsen, J. Usability engineering. *San Diego: Academic Press*, pages 115–148, 1994.

Nesic, S., A. E. Rizzoli, and N. I. Athanasiadis. Towards a semantically unified environmental information space. In *Proceedings of the 9th IFIP WG 5.11 ISESS 2011*, volume 359, pages 407–418. Springer, 2011.

Pariente, T., M. J. Fuentes, M. A. Sanguino, S. Yurtsever, G. Avelino, A. E. Rizzoli, and S. Nesic. A model for Semantic Annotation of Environmental Resources: The TaToo Semantic Framework. In *Proceedings of the 9th IFIP WG 5.11 ISESS 2011*, volume 359, pages 419–427. Springer, 2011.

Sicilia, M. A. Metadata, semantics and ontology: providing meaning to information resources. *Int. Journal of Metadata, Semanatics and Ontologies*, 1(1):83–86, 2006.

Ungar, J., J. Peters-Andres, and W. Loibl. Climate twins – an attempt to quantify climatological similarities. In *Proceedings of the 9th IFIP WG 5.11 ISESS 2011*, volume 359, pages 312– 320. Springer, 2011.

Uren, V., P. Cimiano, J. Iria, S. Handschuh, M. Vergas-Vera, E. Motta, and F. Ciravegna. Semantic annotation for knowledge management: Requirements and survey of the state of the art. *Journal of Web Semantics*, 4(1):14–28, 2006.

Villa, F., I. N. Athanasiadis, and A. E. Rizzoli. Modelling with knowledge: a review of emerging semantic approaches to environmental modelling. *Environmental Modelling and Software*, 24(5):577–587, 2009.