A Model for Semantic Annotation of Environmental Resources: The TaToo Semantic Framework

Tomás Pariente[1], José María Fuentes[1], María Angeles Sanguino[1], Sinan Yurtsever[1],
Giuseppe Avellino[2], Andrea E. Rizzoli[3], Saša Nešić[3]

[1]ATOS Origin, Madrid, Spain
{tomas.parientelobo, jose.fuentesl, maria.sanguino, sinan.yurtsever}@atosresearch.eu
[2]Telespazio, Rome, Italy
[2]giuseppe.avellino@telespazio.com
[3]IDSIA, Lugano, Switzerland
[3]{andrea, sasa}@idsia.ch

Abstract. During the past years huge amounts of resources in the environmental domain have been published on the internet. To facilitate search and discovery of relevant data among an ever increasing mass, the use of tags has been suggested. Yet, the use of non-formal tags for annotating resources allows simple categorization and search capabilities, but it does not provide the means to create cross-domain annotations. On the other hand, ontologies are a shared and formal conceptualization of a given domain and they can be used to formalise tags. The use of formal semantics for tagging allows taking advantage of the reasoning and inference power of the ontologies to create richer resource annotations enhancing the discovery process. In the environmental domain there is a clear need of frameworks and tools allowing formal tagging and discovery. In this paper we discuss about the definition of a Semantic Framework helping the tagging and discovery process of environmental resources. Moreover, we also report on the definition of a model to describe environmental resources allowing cross-domain annotation and search.

Keywords: environment, discovery, annotation, ontology, cross-domain, search, tagging, semantics.

1   Introduction

Since its inception, the Web has brought a revolution on how to publish, transmit and consume information. While at first the number of information publishers on the Web was limited, now almost everyone publishes personal and / or professional information on the Web. Much of this growth is due to the fact that the expertise needed to publish information on the Web is quite small. Basically, publishers only need to own information and organize it in a way that makes it possible to be accessed by consumers. However, the ease of publication, which originally facilitated the growth of the Web, is not necessarily matching the ease of discovery and access to data. In professional environments, where information needs are highly complex, it is necessary that machines, and not just humans, understand the published information

to improve its exploitation capabilities. Meta-information and semantics associated with meta-information are the concepts we need to introduce in order to allow for automated machine-processing of information. And such concepts are at the foundations of the TaToo project.

The TaToo project, which started in 2010, aims at improving the discovery and exploitation capabilities of the published environmental information by providing a way to semantically enrich environmental resources. TaToo is built on top of three fundamental pillars:

- A framework to create complex and formal annotations. Informal annotations are relatively good for human interpretation, but fail to help machines to understand the meaning and perform more advanced tasks based on those annotations. Annotations should be then formal and shared across the domain.
- Usage of ontologies to enhance discovery capabilities and to prevent interoperability issues. TaToo provides an ontology framework that allows cross-domain and cross-language tagging and search. Producing annotations is a time consuming task. TaToo offers both the automatic harvesting of meta-information from existing resources, and a manual interface for humans; both approaches are based on the above mentioned ontology framework.
- A community driven approach. TaToo pays special care in providing easy-to-use and effective tagging interfaces to enable and facilitate the tagging process. TaToo is strongly inspired by existing social bookmarking initiatives, such as reddit, StumbleUpon, Digg, etc. In this sense, TaToo allows the user producing their own annotations, and share them with the community, starting an information enrichment cycle. On the other hand, TaToo differs from the aforementioned social approaches in the use of formal semantic annotations.

This paper focuses on the TaToo ontology framework. The first objective of TaToo is the establishment of an ontology framework that allows the production of formal resource descriptions. Subsequently, the TaToo system is taking advantage of domain ontologies and formal resource descriptions to improve the exploitation of environmental resources by improving the discovery processes, enhancing the resource presentation and providing interoperability between different fields of the environmental domain with the aim of allowing cross-domain search.

Much of the functionality provided by TaToo relies on ontologies. Ontology development is a complex, time-consuming task involving different roles (ontology developers and domain experts) and activities (requirements acquisition, design, etc.). Therefore, from the ontological point of view, one of the biggest challenges of TaToo is to use or produce a formal description of the environmental domain that is at the same time good enough to cover the project objectives, and on the other hand simple enough to avoid unnecessary complexity and gruesome maintenance. The environmental domain is actually a very broad domain that includes several sub-domains dealing with such diverse areas as climatology, toxic compounds fate assessment, crop management, and so on.

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 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 this, the ontology framework should ensure semantic interoperability between different domain ontologies while ensuring their modular nature.

2 Ontology Framework

There are several approaches to achieve semantic interoperability when dealing with different ontologies. Watche [2] defined three ways to integrate ontologies by using single ontology, multiple ontology or hybrid ontology approaches.

The single ontology approach uses one global ontology that provides a shared vocabulary for the specification of the semantics. Thus, all information sources are related to one ontology, which usually is a combination of several specialized ontologies. This approach is suitable for solving integration problems where all information sources to be integrated provide nearly the same view on a domain.

In the multiple ontology approach each information source is described by its own ontology. The main advantage of this approach is that no common and minimal ontology commitment about a global ontology is needed. To achieve a common understanding of the information sources, defined ontologies must be related. To relate several "source describing" ontologies, several inter-ontology mappings must be introduced. In practice, these inter-ontology mappings are very difficult to define because of the many semantic heterogeneity problems, which may occur, and the situation is aggravated when there are many domain ontologies to map.

The hybrid approach is based on using a common shared ontology and a set of local or application ontologies that are mapped uniquely to the shared vocabulary. In this case the local ontologies would extend the vocabulary to the needs of a given domain and the interoperability is achieved based on the shared ontology. TaToo follows the hybrid approach as shown in Fig. 1.
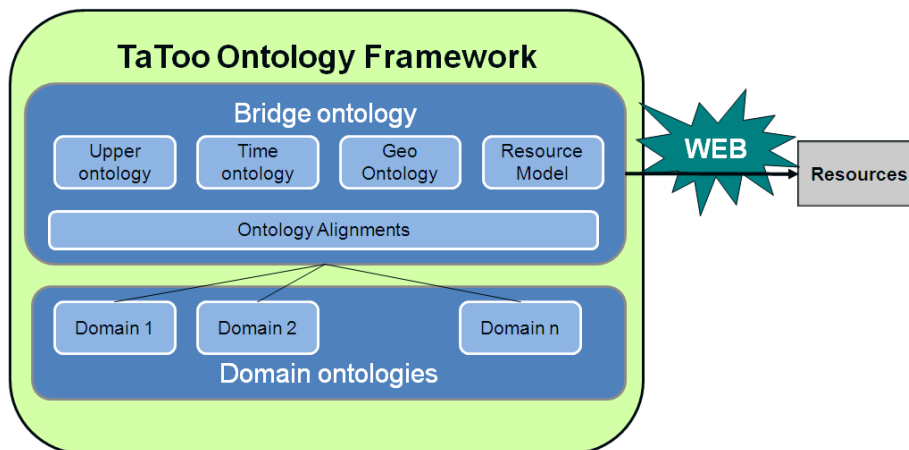

Fig. 1. TaToo high-level ontology framework

In order to apply the hybrid approach, a bridge ontology that will become the upper layer of the integrated ontology framework has to be developed. Our bridge ontology includes concept and property definitions from widely adopted ontologies and necessary concept definitions in order to ensure the mappings between domain ontologies. Any domain ontology should be mapped properly to the bridge ontology in order to be usable by the integrated ontology framework. So the system interacts with the domain ontologies via the bridge ontology.

The mapping process is done manually by defining equivalency and subsumption relationships between concepts, object properties, data properties and individuals of bridge ontology and the domain ontology. From the point of view of the bridge ontology, the entire ontology framework is seen as one big knowledge base. In Fig. 2, where the current version of the TaToo ontology framework is shown, some example mappings between ontology elements from the bridge ontology (crossed-out circles in green) and the domain ontologies (crossed-out circles in gray) are depicted to represent the cross-domain alignment process.
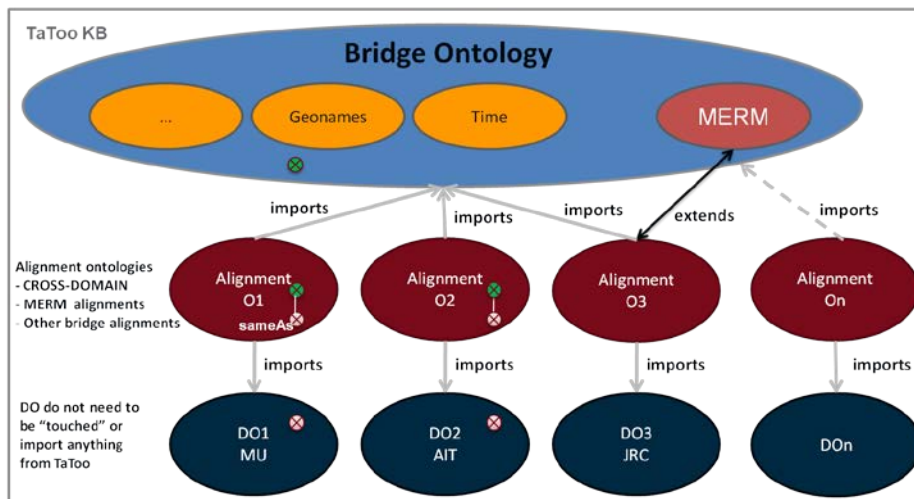

Fig. 2. Current version of the TaToo ontology framework

Some of the selected ontologies included in the TaToo bridge ontology are GeoNames, an ontology for representing geospatial expressions, and OWLTime, an ontology for representing temporal expressions. There are already implemented reasoners which especially support both ontologies for geospatial and temporal reasoning. One of them is OWLIM, an OWL semantic repository, which supports geospatial reasoning exclusively for the GeoNames ontology. As TaToo is expected to offer users the possibility to create time and location based annotations and temporal and spatial queries, these ontologies become an essential part of the bridge ontology. In particular, OWLIM has been selected to be used as the semantic repository and the availability geospatial reasoning has been one of the reasons for this choice. OWLIM supports OWL2-RL, a web ontology language profile inspired by rule-based languages. OWL2-RL is much more efficient and scalable compared to

OWL2-DL but less expressive. This is the reason why OWL2-RL has been chosen. Because of the constant incremental annotated data that will be manipulated, a need for more efficient reasoning arose and OWL2-RL is aimed to cover these kinds of needs.

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 a URI. For future releases, the usage of an upper-level ontology to perform more elaborate cross-domain mappings will be investigated. The Semantic Web for Earth and Environmental Terminology (SWEET) [1], a widely-used set of ontologies providing semantic descriptions of Earth system science, is a good candidate for the TaToo ontology framework.

3   The Minimal Environmental Resource Model

A minimum resource model is defined as the largest common denominator between a set of heterogeneous description formalisms related to a common resource. Minimum resource models have been applied in other research areas with satisfactory results. In [4] a description of an open system for web services publication and discovery based on a description of a minimal service model resulting in an ontology for service modelling called POSM is presented. Functionally, the minimum resource model aims at a similar objective as the bridge ontology: providing a common framework between different domain ontologies. However, conceptually, the minimum resource model is an effort to identify a minimal model that, without limiting the expression of specific domains, acts as a reference for past and future applications in a specific area. This minimisation effort and referencing aim is what differentiates the bridge ontology from a minimum resource model and it is what makes it a valuable resource.

One of the most relevant parts of the TaToo bridge ontology is 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 [3]. The NeOn methodology encourages reusing as much as possible. To that extent, MERM contains, besides the already mentioned POSM ontology, concepts from several other ontologies like SIOC[1], FOAF[2], Dublin Core[3] and O&M[4]. **Fig. 3** presents the main classes of MERM: *Resource, Annotation*, and *ResourceAccessInfo*.

---

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

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

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

[4]Observations   and   Measurements   ontology,   http://seres.uni-muenster.de/o&m/O&M_discussion_paper.pdf
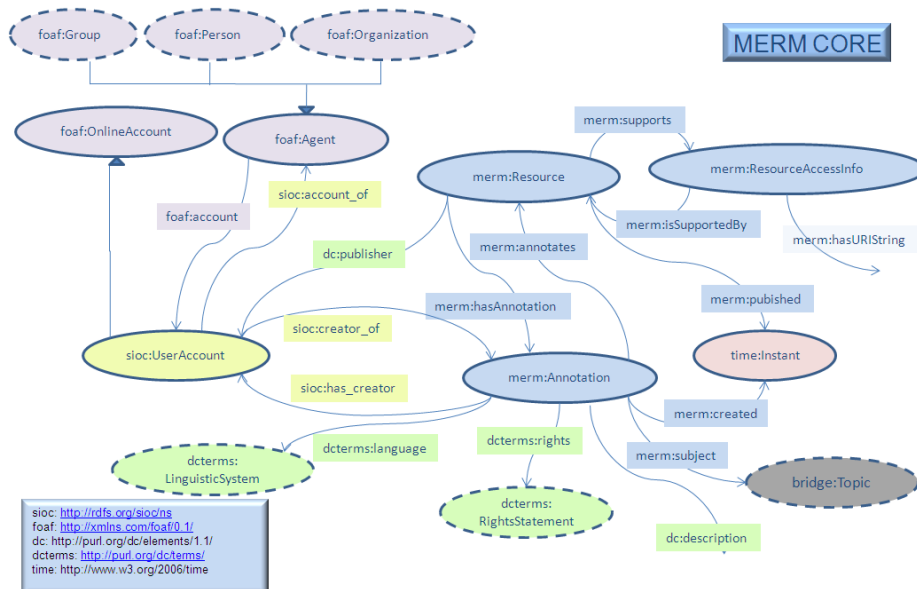
Fig. 3. MERM core classes and relations

The *Resource* class represents a resource in the TaToo system. It includes resource management information as author, owner, provider, date of creation, etc. A *Resource* usually has access information presented in the *ResourceAccessInfo* class. The *ResourceAccessInfo* class contains all information needed to access a resource. This information is very heterogeneous ranging from a simple URL to a complex WSDL depending on the nature of the resource. Finally, a resource presents a set of annotations, represented by the *Annotation* class. An annotation describes what a resource is about. Except for some management information, the content of an annotation depends on the type of the annotated resource. An annotation for a web document would be different from an annotation for a web service, being the "some resource is related to some domain concepts" the simplest kind of annotation. Resources in TaToo are retrieved taking their annotations as basis. In the current version, MERM distinguishes between three kinds of annotations:

- *TimeSeriesAnnotation* that describes time series following the model proposed by the Observations and Measurements Ontology.
- *WebServiceAnnotation* that describes web services following the model proposed by POSM.
- *WebAnnotation* that describes web resources following the model proposed by SIOC.

MERM is still work in progress so the addition of new models for several resources is expected. **Fig. 4** shows in details the model adopted for *TimeSeriesAnnotation*.
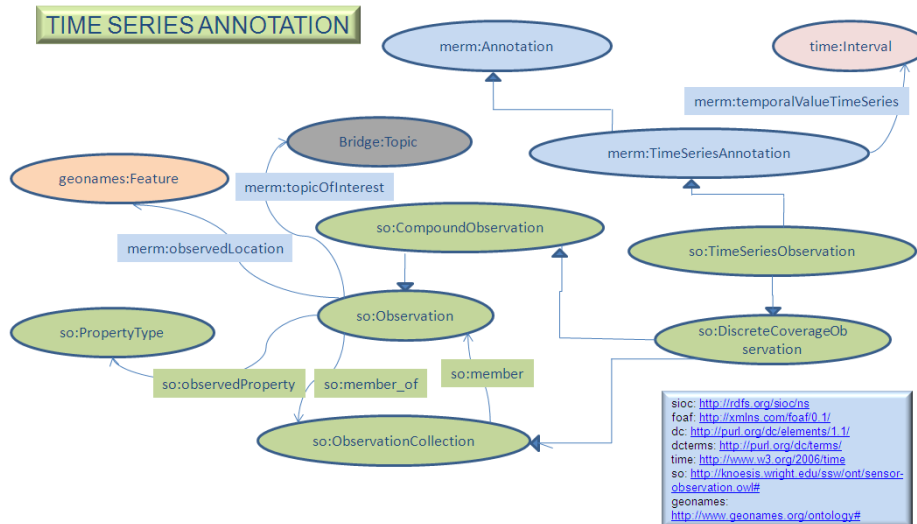
Fig. 4. *TimeSeriesAnnotation* model

Finally, there is also a special kind of annotations, the evaluations, which are annotations that relates to another annotation instead 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.

4   Integrating Domains

The TaToo Ontology framework, as it was described in Section 2, comprises the bridge ontology, a number of domain ontologies, and a number of alignment ontologies (see Fig. 2). 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 this section we describe two possible alignment/mapping strategies on the example of three domains ontologies, specific to the validation scenarios of the TaToo project, namely the JRC, MU, and AIT ontologies. The JRC ontology describes the agro-environmental domain, the MU ontology describes the anthropogenic impacts of global climate change, and finally the AIT 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. Moreover, we also plan to use them in the evaluation of the semantic tagging and discovery capabilities of the TaToo system.

The first mapping strategy, adopted for the JRC ontology, uses the *rdfs:subClassOf* construct defined in the *RDFS* to map concepts of the domain ontology to the

concepts of the MERM and bridge ontology. For example, concepts describing some domain specific environmental resources should be specified as sub-concepts of the *merm:Annotation* concept or its sub-concepts (*merm:TimeSeriesAnnotation*, *merm:WebService-Annotation*, and *merm:WebAnnotation*). Moreover, the concepts introduced to conceptualize some domain specific features (topics), which are supposed to be a part of the resources' annotations, should be defined as sub-concepts of the *bridge:Topic* concept. To illustrate this mapping strategy on a practical example, in Figure 5 we list a snippet of the JRC ontology mapping specifying a resource annotation type that describes software development kits:

```
<owl:Class rdf:ID="SoftwareDevelopmentKit">
        <rdfs:subClassOf rdf:resource="merm:Annotation"/>
</owl:Class>
```

Fig. 5. A snippet of the JRC ontology mapping

The second mapping strategy, used for the AIT and MU ontologies, assumes the usage of the *owl:equivalentClass* and *owl:sameAs* constructs defined in OWL, plus the *skos:broadMatch* and *skos:narrowMatch* constructs defined in the SKOS[5] (Simple Knowledge Organization System) vocabulary to map concepts and properties of a domain ontology to the concepts and properties of the MERM and bridge ontology. The difference between the OWL and the SKOS constructs is that the latter ones provide a looser mapping. There is also a slight difference between the two OWL constructs regarding the strength of the mapping they provide. The meaning of the *owl:equivalentClass* is that the two concepts have the same set of individuals. However, the *owl:equivalentClass* does not imply the concept equality. Concept equality means that the concepts have the same intentional meaning. Real concept equality can be expresses by the *owl:sameAs* construct. In Figure 6 we show a snippet of the MU ontology mapping that specifies the mapping of the "*Diagnose*" concept from the MU domain to the *merm:Annotation* concept of the MERM ontology.

```
<owl:Class rdf:about="Diagnose">
    <owl:equivalentClass rdf:resource="merm:Annotation"/>
</owl:Class>
```

Fig. 6. A snippet of the MU ontology mapping

The development of both bridge ontology (including MERM) and domain ontologies (i.e., JRC, AIT, and MU) in the TaToo Ontology framework is an evolutionary process. 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.

---

[5]http://www.w3.org/TR/skos-reference/

5 Conclusions

In this paper we have presented the approach of the TaToo project to the development of an ontology framework for the semantic annotation of environmental resources. As the environmental domain is composed by a wide number of sub-domains, we chose a hybrid approach to be able to integrate ontologies for the various sub-domains. Such an approach required the development of a bridge ontology and a series of alignment ontologies to map concepts from domain specific ontologies to the bridge ontology. A minimum environmental resource model has also been developed to be included in the bridge ontology in order to provide a minimal set of common concept to be shared across the different domains and facilitate alignments.

The TaToo ontology framework is functional to the declared objectives of the TaToo project that is to allow for the semantic annotation of environmental resources in order to enhance and facilitate search and discovery.

References

1. Raskin, R.G., Pan, M.J., 2005. Knowledge representation in the semantic web for earth and environmental terminology (SWEET). Computers & Geosciences 31, 1119-1125.
2. 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.
3. Gómez-Pérez A., Motta E., Suárez-Figueroa M.C., NeOn Methodology in a Nutshell. http://www.neon-project.org/nw/NeOn_Book, 2010.
4. Pedrinaci C., Liu D., Maleshkova M., Lambert D., Kopecky J., and Domingue J.: 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.