EnviroInfo 2011: Innovations in Sharing Environmental Observations and Information

The TaToo Semantic Case – Requirements, impacts and application

Gerald Schimak[1], Bojan Božić[1], Alexander Kaufmann[1], Jan Peter-Anders[1], Pascal Dihé[2], Andrea Rizzoli[3] Tomás Pariente Lobo[4], Giuseppe Avellino[5]

Abstract

Up today numerous semantic developments, starting from ontologies and leading to semantic applications in various (mostly financial or business) domains have been developed and put on the market. Still real suitable systems or application fostering the enrichment of environmental resources in order to make them better understandable and/or discoverable for a broader community are missing or have not be thought of yet. TaToo (Tagging Tools based on a semantic discovery framework, FP7 project, www.tatoo-project.eu) intends to demonstrate how the enormous amount of environmental information can become better discoverable and understandable for a broad range of user, e.g. from environmental experts up to the man on the street.

Clearly upfront to the definition, specification and implementation of the TaToo framework and tools the requirements with respect to environmental resources handled in TaToo (e.g. environmental data, services, applications etc.) have to be defined. TaToo deals with three scenarios, playing a two folded role. On the one hand they are key providers for challenging requirements but they are also the evaluation facility to prove (as domain experts) the results of project. The proposed paper will describe:
a) the main categories of requirements (incl. the approach of requirements collection) written down e.g. for tagging, cross-domain annotation and discovery, meta-information, ontologies, etc.
b) brief overview on the tools and services for tagging and discovery provided by the TaToo project
c) description of Validation Scenario on Climate Twins Regions and its requirements as well as usage of TaToo developments
d) the expected impact on Europeans Single Information Space and its environmental society/community

1. Introduction

Looking for information seems so easy today. Tell your friend you are looking for something and his answer will be immediately "Google" for it. All of us tried it many times – and – were often frustrated because we got many hits that we clearly could read but not really interpret in the sense if they results are suitable for our needs – so we investigated spending hours and hours to get illuminated.

[1] AIT Austrian Institute of Technology GmbH, Vienna, Austria
email: {Gerald.Schimak, Bojan.Bozic, Alexander.Kaufmann}@ait.ac.at
2 cismet GmbH, Saarbrücken, Germany,
email: pascal.dihe@cismet.de
[3] IDISA, Lugano, Switzerland,
email: andrea@idsia.ch
[4] ATOS Origin, Madrid, Spain,
email: tomas.parientelobo@atosresearch.eu
[5] Telespazio spa, Rome, Italy,
email: giuseppe.avellino@telspazio.com

The TaToo Project idea comes exactly from that problem case mentioned, as one of our climate change experts discussed with us that it is impossible for him to get any additional information to data or models (he would luckily find at the internet) that allow him to interpret if these data or models would fit for his needs. What he means is that there is definitely no information about the model itself and also no information stemming from someone else who probably made already some experiments with that model or data. You see! – A big gap exists here between available data already published and the possibilities to take advantage or benefit of it. So, often a waste of time and in many cases a loss of money too.

Up to now appropriate means are missing to allow at first a meaning full search (i.e. semantically supported) as well as second the possibility to add additional (user or expert) information to a resource (data/services/models etc.) found in the Web. This is not only true for our expert mentioned above but also for the whole internet community. TaToo is taking up this challenge and will provide means – starting from requirements necessary to define the TaToo semantic framework architecture – up to services and tools that allow the enrichment of information as well as the discovery of these additional information items (Tags).

2. TaToo requirements

The TaToo requirements specifically the process of collecting requirements represents a consolidation point in the process of the TaToo development process. As a result these requirements describe what TaToo shall do for the Users and furthermore it is the reference for various downstream tasks, like:

- the design of the TaToo framework architecture;
- the specifications of the services and tools offered by TaToo;
- the verification and exploitation of the TaToo results.

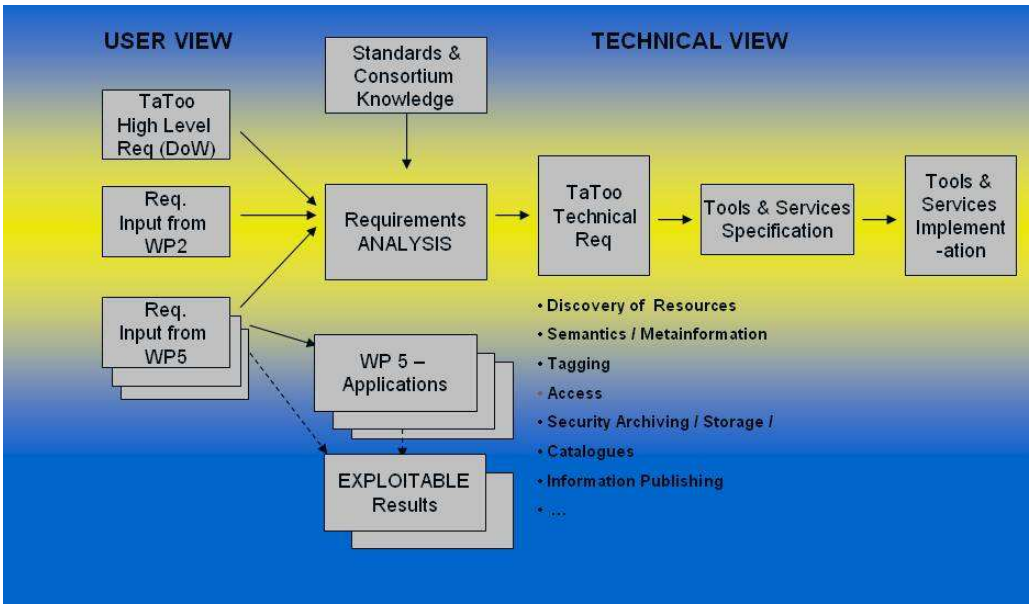

Figure 1: TaToo approach to technical requirements (Božić/Schimak 2010)

As mentioned before the TaToo technical requirements are the result of a consolidation process which was subdivided into four steps:

1. Discerning analysis of the applications oriented requirements documented during the process of deliverables preparation of for the three TaToo Validation Scenarios (Scenario 1: "Climate Change Twin Regions – Discovery Platform", Scenarios 2: "Agro-environmental management" and Scenario 3: "Anthropogenic impact and the influence of global climate change"), in order to identify the potential link with the technologies to be used in the project.
2. Analysis of the user needs and translation into relevant technical requirements.
3. Analysis of other useful sources of complementary requirements derived from: the knowledge and experience of the TaToo team (Schimak 2008) and other FP7 projects like SANY, ORCHESTRA or NEON, standardization bodies and user communities.
4. Addressing high level project requirements as described in the workplan of TaToo as measurable objectives.
5. Classification and formal documentation into the project's deliverable "D2.3.2 Requirements Document" (Božić/Schimak 2010) describing the scope (function), the realisation time (for example implementation in version 1 of TaToo) as well as the trace-back to the origin of the requirement. This Requirements Document serves as the reference for the preparation of the architectural design and specifications of tools and services.

The manifestation of the requirements splits into two main parts. Where the first part covers high level and generic requirements like:

a) Strategic positioning requirements obeying the Single Information Space for Europe (SISE)
b) General enterprise requirements which are mainly derived from the high-level objectives of TaToo as well as from requirements stemming from the Reference Model of ORCHESTRA (Usländer 2007) and SANY (Schimak 2008).

The second part specifically take care about requirements coming out of the analysis of the use cases defined by the different TaToo Validation Scenarios as well as a dedicated survey held during the course of requirements collection. This led to a set of requirements categories shown in the Figure 2. (Note: The ordering of the blocks does not have any specific meaning and additional blocks could be added if needed.).

DISCOVERY	**ARCHIVING / STORAGE**
SEMANTICS	**DATA QUALITY**
TAGGING	**USER COMPONENTS**
ACCESS	**ACCESS**
SECURITY	**SYSTEM ADMINISTRATION**

Figure 2: TaToo technical requirements categories (Božić/Schimak 2010)

2.1 High Level and Generic Requirements

These requirements are setting the generic constraints to the architectural design of the semantic framework as well as service network with the aim of maximising its flexibility and ability to adapt to changing technologies and functional requirements.

- The TaToo architecture shall make use of proven concepts and standards in order to decrease dependence on vendor-specific solutions and help ensure the openness of the TaToo network and support the evolutionary development process of the architecture.
- The components involved in a TaToo framework shall be loosely coupled, where loose coupling implies the use of mediation to permit existing components to be interconnected without changes. Extensibility/Flexibility is needed in different aspects:
  - Extensibility of domains: the scope of TaToo shall not only be limited to a specific thematic domain dealt with in the TaToo validation scenarios. Thus TaToo must allow integrating new domains.
  - Extensibility/Flexibility of functionalities: TaToo shall not be a "closed" system with a fixed set of functionalities: it must be possible plug-in new services with additional functionality into a TaToo semantic framework.
- The TaToo architectural process shall be such that the TaToo architecture and the architecture of components are decoupled. This means that a component shall be seen as a black box, i.e., no assumptions about its inner structure are made when designing the architecture. This requirement facilitates the flexibility to exchange or replace components.
- The TaToo architecture shall be designed to allow security mechanisms to be incorporated. These mechanisms include user management (authentication, authorisation), as well as control of access to data, services and tools.

Furthermore, the TaToo Architecture shall be able to take scalability issues into account, like the following aspects:

- Type and amount of concerned resources (for discovery and tagging)
- Discovery and tagging services and tools
- Domains (domain ontologies)
- Size and level of detail of tagging information (meta-information)

2.2 Technical Requirements and Categories

The categorisation was done as a result of the needs and gaps analysis (which comprised user questionnaires as well as a technology survey) performed during the initial phase of the project. Here we briefly present the main categories as shown in the figure above. The detailed list of requirements can be found in the public available project deliverable, called "Requirements Document" available at the TaToo web site: www.tatoo-project.eu;

2.2.1 Discovery and Harvesting

Discovery: In general, we define the process of discovery and / or search as the retrieval by a system of a set of resources that satisfy an information need expressed by a user. In the scope of TaToo search is the process of expressing the need for information by the user, and discovery is the process performed by the system to retrieve the results.

In TaToo, we intend to use semantic search, a kind of search that makes extensive use of domain knowledge encoded in the form of ontologies. Ontologies are related to resources through annotations stored as metadata. It is for this reason that the search in TaToo will always be metadata-based.

A classic information retrieval system usually provides a crawling functionality in order to gather information from external resources in a (semi) automatic way. This process creates an inverted index on which the discovery is made later. In the same way TaToo harvests additional meta-information from known catalogues and repositories and consolidates the annotations in the system knowledge base, where the discovery is made afterwards.

Harvesting: One of the main objectives of a metadata-based system is the automatic or semiautomatic metadata acquisition starting from resources. However, this goal is far from being achieved due to the fact that (semi)automatic annotation acquisition is a very complex process and depends heavily on the resource itself. TaToo approaches this problem through the harvesting process.

TaToo harvesting process aims to generate metadata usable by TaToo starting from existing metadata stored in catalogues. Taking (semi)structured metadata as starting point will facilitate the automatic acquisition process, making it affordable in the scope of TaToo. Since the acquisition process depends on the structure of the catalogue, it is necessary to configure and programme specific harvesters for each catalogue.

2.2.2 Semantics

In linguistics, semantics is the subfield that is devoted to the study of meaning, as inherent at the levels of words, phrases, sentences, and texts. In computer sciences, semantics is framed in the Semantic Web field that enforces a vision of the web in which data is semantically annotated so automatic agents can understand this data, enhancing its exploitation and automation capabilities. In the scope of TaToo, environmental resources will be semantically enriched to improve its exploitation capabilities (discovery, publishing, accessing, etc).

Ontologies: Adding semantics is a very abstract process which is normally done by linking data with concepts described in ontologies. The most widely used definition for ontologies is that used by Gruber (Gruber, 1993): an ontology is a formal, explicit specification of a shared conceptualisation. More specifically, an ontology provides the tools to describe a domain in terms of classes, attributes and relationships. In TaToo, ontologies have a dual role, on one hand, they describe several domains related with each TaToo validation scenario, on the other, they provide a common conceptualization for TaToo tools and services on how to describe and manage resources, annotations, evaluations, etc

Meta-information: Typically many information items or resources are available but have no accompanying meta-information. Therefore it is important for us to enable users to add additional descriptive information (tagging of resources – see relevant chapter in this doc thereafter). On the other hand, it is important that the user could easily extract existing metadata that already comes with the resources in order to use them for further interpretation of other resources.

A common structure for metadata / meta-information for TaToo resources for the evaluation process has to be defined and made available. Also a physical place where to store this information is needed.

2.2.3 Tagging

The users typically want to enrich resources (e.g. data, models, services) by adding additional data on the topic, e.g. hints for future users of the data or model; quality, usability, suitability or even links to other similar resources. The practice of adding meta-information to resources allows the information enrichment process that is at the base of the TaToo project aim.

Tagging: Tagging is understood as the process of adding information (as part of the metadata / meta-information of an item) to a certain artefact or piece of interest. Tagging is not only important for users wishing to add useful metadata to resources they can find in a second time, but it is considered one of the most important functionality the TaToo project is providing.

Annotation: In TaToo we perform tagging using ontologies. According to the TaToo understanding and definition, the process of tagging with semantic information is called annotation.

From the requirements perspective the term tagging is more generic. Therefore, in this section we use the term tagging to refer generically to the process of creating semantic metadata.

2.2.4 Access

Discovery and tagging often goes hand in hand with facilitating access to the discovered resources on the one side and accessing the meta-information ("tags") maintained by TaToo about the resource of interest on the other side.

Access to TaToo meta-information which involves besides reading also creating and updating meta-information will take place in a regulated environment, the TaToo Semantic Service Environment and Framework, featuring well defined service interfaces and supporting standards-based access control mechanisms. In contrast, accessing the resource itself, especially if it is an application, a mathematical model or a service, may be a way more complex. Therefore, TaToo has to provide appropriate meta-information on how to access the resource, including information about specific security restrictions that may apply when requesting access.

Furthermore, access to TaToo meta-information is characterised by the fact that it is either performed through public interfaces (user interfaces and service interfaces) exposed by the TaToo Service Environment and Framework and/or within the framework or the framework components (e.g. the TaToo portal accessing the TaToo Clearinghouse) itself. Standards'-based access control mechanisms at interface level for access performed through public interfaces have to be established at least for those operations that add new or modify existing meta-information. As an essential requirement as well as an essential feature of TaToo is, to not only provide access to meta-information about a resource but also the appropriate means to allow and ease the access to the resource itself. Uniform and transparent access to such disparate types of resources like databases, documents, services, models, etc. has to be supported and adequate meta-information about the various access mechanisms has to be provided.

2.2.5 Security

Since TaToo will facilitate discovery and access of a wide range of environmental resources and web-based services we have to think about how security can be provided in such an environment. Some of those services or resources may require a registration or usage fee, may be subject to copyright or lawful restrictions or may not be available to the public for some reason (Dihé, 2010).

A service provider who is willing to make non-public services discoverable, accessible and tag-able by TaToo should be offered the possibility to control who can do what with his services. But not only the external services and resources accessed by TaToo may require protection. Also the access to TaToo's public services (e.g. tagging and search) must be regulated to prevent potential misuse.

TaToo services, especially those services having public interfaces and allowing manipulation meta-information have to be access controlled. Access control encompasses registration / management of identities ("users"), their authentication ("login") and the enforcement of the access restrictions (authorisation). Access control, Authentication and Authorisation is understood as defined in OASIS Reference Architecture for Service Oriented Architecture Version 1.0, http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.pdf

2.2.6 Archiving/Storage

The TaToo Clearinghouse as on of the Core Components plays the role of organising the semantic information provided on environmental resources. It is a central component for accessing the semantic annotations storage and serves also as an information exchange support between the TaToo system components.

All descriptive information provided for a specific resource (data, service, model, etc.) either via annotations, tagging of resources, ontologies etc. have to be stored at a central place and made available and accessible (if not secured) by everybody.

2.2.7 Data Quality

TaToo provides metadata to allow a better interpretation of reliability and suitability of data through the possibility to evaluate discovered information. Quality indicators have to give information (to end users) about data quality, reliability, uncertainty and so on. This quality information is intended to be used for better interpretation of suitable search/discovery results by the end user.

At first, the quality criteria are going to be defined by domain experts. At later time User Communities are intended to enhance these criteria and furthermore TaToo takes their knowledge into account for evaluation and ranking (classification) of information related to data quality.

2.2.8 User Components

TaToo is providing a set of Tools as User Components (e.g. portal, tools, clients) in order to establish the TaToo functionalities, like the tagging functionality thus enabling the enrichment process, which in turn enables the semantically enhanced search and discovery process.

In the context of TaToo, the term tool is intended as a front-end component, generally with a graphical user interface, which allows the user (residing in the Presentation tier) to interact with the system taking advantage of the provided functionality. A TaToo tool acts as a (Web) client of a Server side TaToo (Web) services.

It is possible to consider a TaToo tool as a portal providing a set of functionality through a set of configurable portlets, or the single portlet itself. A TaToo tool can also be a browser plug-in (for instance, implementing a tool bar) or a client side application. But for the moment, the TaToo framework shall be accessible and exploitable through a Web portal. This makes possible taking advantage of the provided functionality theoretically without any other need apart from a Web browser and an Internet connection. The implementation shall be done using portlet technology. Portlet technology allow us to develop applications by aggregation of pluggable user interface software components that are managed and displayed in a web portal and executed into a portlet container.

2.2.9 Visualisation

In order to evaluate resources and then provide appropriate tags and/or annotations for these resources, the user should be supported by tools, applications, or other types of user components. An appropriate visualisation (respectively information representation) has to be provided to enable the user to understand and possible further process these informations.

All resources that TaToo is considering are identified by URIs. Once discovered, the resource URI has to be referenced and the obtained resource representation has to be either visualised (e.g. a picture, diagrams, or a document), or somehow elaborated (e.g. a set of raw data).

We know that it is impossible to design and implement a web-based application able to visualise any resources and to let, at the same time, the user apply tags and semantic annotations to that resource. For this

reason, the major strategy will be to design the visualisation part of the TaToo system as an extensible application that could be enhanced in its functionalities, e.g. by specific plug-ins.

2.2.10 System Administration

TaToo will provide functionalities for administrating the TaToo system, together with a simple client side component (GUI) or a specific portlet accessible only to authorized users. Administration functionalities will include: system general maintenance, monitoring, configuration, management of users or user groups, etc.

3. Tools and Services

This section provides only the descriptions of two TaToo Public Services namely for Tagging and Discovery and related Tools. The full set of TaToo services as well as their interface descriptions and functionality and TaToo tools can be found in (Dihé/Schlobinski 2010).

3.1 Tagging

In the Web context, the term 'tagging' concerns the attachment of user-defined labels (tags or descriptive keywords) to resources in order to describe, organize and share these resources. Tagging allows social information exchange and democratizes the process of classification as tags are not exclusively provided by a central authority (normally owning the resources). In general, tagging is a way for collecting meta-information. The assignment of meta-information to resources makes possible: Finding resources; Sharing resources; Managing owned resources; and Linking resources each other.

The main aim of the TaToo project is bridging the resource discovery gap. In order to improve the discovery process, being able to discover resources basing on Semantic Web (Web 3.0) concepts, resources are semantically annotated (tagged) through tags (meta-information) based on terms expressed as concepts from relevant domain ontologies .

3.2 Tagging Service

The Tagging Service (Petronzio/Avellino 2010) is the service responsible for all the operations related to the tagging of resources. It interacts with the tagging User Component that demands updating of meta-information for a resource and the TaToo Clearinghouse, a TaToo Core Component, to access to the meta-information repository, but also to retrieve ontology information if needed.

Information exchanged with the tagging User Component and the Clearinghouse includes:

- Identification of a resource or a list of resources
- Meta information structured according to some schema (general, resource type specific and/or domain specific)
- "Semantic information" given back to the User Component to support the creation of additional meta-information to be associated in a further tagging request by the user component
- Domain identifier of a domain selected by the user
- Ontology representation in xml-based format

The Tagging Service is a TaToo Public Service (Dihé 2011; Dihé/Schlobinski 2010) that allows User Components, in particular the Tagging Portlet, to access the tagging functionality offered by the TaToo System and exposed through the public interface of this service. A Tagging Service receives tagging requests from the different User Components to:

- associate meta-information with resources;

- access already available meta-information;
- update meta-information associated with the resource;
- delete meta-information;
- access ontologies.

The Tagging Service, in its first version, is limited to the creation of new tags and the read-only access to existing tags. Currently, tagging is limited to simple semantic tagging which means choosing from a number of terms of the selected ontology and the supported formats for ontologies and tags are limited to RDF and OWL. More sophisticated tagging possibilities, update and delete operations and other client-specific formats will be added successively. The access to ontologies is realised by a dedicated Ontology Retrieval Interface.

Currently we have two tagging processes in mind:

- Simple tagging process: the tagging User Component provides meta-information (tags) to be associated with a resource. This meta-information has to be stored via the Clearinghouse in the TaToo meta-information store.
- Complex tagging process: The tagging User Component asks for existing meta-information about a resource, updates this meta-information with additional information and requests that the new meta-information be associated with the resource.

The Tagging Service is supposed to be an interoperable and standalone Web server, in particular a Web Service W3C compliant taking advantage of widely adopted standards such as XML, WSDL, and SOAP.

The service specification of Tagging is comprised of the following interfaces: The Tagging Interface, that includes all operations related to the tagging functionality and the Ontology Retrieval Interface, that includes all operations related to the retrieval of ontologies to support tagging functionality.

Where the Tagging Interface contains the following operations: addTag, associates a single tag with exactly one resource; addTags, associates different tags with at least one resource; getTag retrieves a specific tag of exactly one resource; getTags retrieves all tags of at least one resources. And the Ontology Retrieval Interface contains: getOntology, retrieves an ontology related to a certain domain; getMERMOntology, retrieves an ontology containing the MERM.

Tags must be encoded in RDF and ontologies in OWL.

3.3 Tagging Tool – Tagging Portlet

The Tagging Portlet (Petronzio/Avellino 2010) allows a user to tag information about a discovered or already known resource, where a resource could be a data source, a service, a Web page, etc. The user is prompted by the portlet with a selection panel to choose terms from an ontology to create tags for a resource or a set of resources. When the user adds tags, the Tagging Portlet contacts the TaToo Tagging Service to update the information related to the resource in the TaToo meta-information repository.

The Tagging Portlet displays to the Web portal user the result of the tagging operation whether it is successful or unsuccessful. As already mentioned before, the Tagging Portlet is a TaToo User Component. It is part of the TaToo Web Portal and provides tagging functionality to the end user (in particular to taggers), with the Tagging Services to access the TaToo Business tier. Portlets are defined as Web components, managed by a portlet container, that process requests and generate dynamic content. Portals are configurable through different portlets to offer the presentation layer to the end user.

The concrete realisation of the Tagging Portlet may depend on the meta-information schema (tags) to be supported as well as on specific user requirements. At the beginning the Tagging Portlet will be realised as a generic tagging tool providing basic user interface and functionality.

3.3.1 Open versus closed tagging

As already stated, tags are important as they provide useful information on resources, and how to search, use, access, or share resources.

In general, a tag can be any kind of term, in a certain natural language, from a certain domain. It can express knowledge about resource subject or type, resource location, intended use, or any other concerning detail. It can be a single word or a sentence. Obviously, tags highly depend on the tagger providing them (e.g. user's knowledge, application domain, interest in the resource, etc.).

Depending on the characteristic of the set of available tags, two types of tagging can be identified:

- Open tagging, when tags can be freely defined by the taggers;
- Closed tagging, when the taggers is forced to use a specific set of tags.

Normally, when taggers are allowed to use no matter what tags (open tagging), they are really free to express themselves. However, this opens problems related with the harmonisation of the semantics associated with the tags (ontology matching). It might happen that tags could be considered different even when they have the same meaning.

In order to reduce the number of tags in an open tagging system, two solutions can be mutually adopted:

- Stemming algorithms to be used to eliminate tags provided as plurals or using upper case letters ('tattoo', 'tattoos', 'TaTToo' will be considered the same tag even if different).
- Vocabulary to be used to eliminate problems relying on synonymy and / or polysemy

Tagging resources through a fixed set of terms (closed tagging) limits the tagger, but has several advantages. It allows:

- Searching for resources basing on a set of well known terms from a certain domain of interest, so increasing the hit rate;
- Searching for resources basing on different tags types expressing different classification criteria (e.g. searching a book by author, title, or other);
- Sharing personal vision about the tagged resources in a way other taggers (in the same domain) can easily understand;
- Classification of resources made by a huge amount of people creating a kind of community knowledge;
- Connecting different resources.

Focusing on the improvement of the discovery process, TaToo will adopt the closed tagging approach allowing taggers to use concepts only from shared domain ontologies.

3.4 Search and Discovery Overview

In general, the process of discovery and / or search can be defined as the retrieval by a system of a set of resources that satisfy a need expressed by a user. Despite there is no commonly accepted distinction between search and discovery, in the scope of TaToo, search is the process of expressing the need for information by the user, and discovery is the process performed by the system to retrieve the results.

Search and discovery components take part in the three architecture Building Blocks User Components, TaToo Public Services, and TaToo Core Components and the architecture's Presentation Tier, Service Tier and Business Tier respectively. All Building Blocks and Tiers are detailed in (Dihé 2011) and (Dihé/Schlobinski 2010).

The User Component providing access to the system will be a web portal, specified, designed, and implemented as a set of portlets. The Search Portlets allow users to query the TaToo System in different ways. Two search portlets have been defined: A Simple Search Portlet and a browsable Hierarchy Portlet. A results portlet shows the user the relevant resource annotations for a given query.

3.5 Discovery Service

The Discovery Service (ParienteLobo/Fuentes Lopes 2010) exposes the public discovery functionality to User Components. It supports (semantic) search and discovery of annotated resources. The discovery process is query driven, allowing the user to select certain terms from an ontology that are then used for the semantic search. The Discovery Services itself does not implement any business logic.

In principle, the Discovery Service interacts with the Clearinghouse and the search User Component. It receives search requests in a specific format from the User Component and transforms them into a message suitable for the Clearinghouse. Information exchanged with the Search Portlet and the Clearinghouse includes:

- Query containing search terms selected from an ontology.
- Information about search preferences (personalisation, use of inference, etc).
- Search results (resource ID and some descriptive meta-information about the resource).

The Discovery Service is invoked by a lightweight User Component to perform a semantic search. It takes the input provided by the User Component and transforms it into a request for the Clearinghouse. The Clearinghouse performs the search and returns some results to the Discovery Service. The Discovery Service transforms the results provided to facilitate the work of the User Component.

The service specification of discovery is comprised of the following interfaces: the discovery interface, that includes all operations related to search and discovery functionality; the ontology retrieval interface, that includes all operations related to the retrieval of ontologies to support discovery functionality;

3.6 Search and Discovery Tools – Portlets

In the following we give a short overview on two search tools as User Components of TaToo:

- the Simple Search Portlet and
- the Hierarchical Search Portlet.

Within the TaToo User Components, the Simple as well as the Hierarchical Search Portlet are components within the web portal. More information on the TaToo Web Portal can be found in (Dihé/Schlobinski 2010).

3.6.1 Simple Search Portlet

The Simple Search Portlet allows a user to search resources regardless of their type. The Simple Search Portlet provides a GUI suitable for expressing the most relevant queries for a given domain in accordance to a given search profile. When the user clicks on the search button, the Simple Search Portlet contacts the TaToo Discovery Service to retrieve the resources relevant to the user's search. To display this information, the Simple Search Portlet redirects the results to a TaToo User Component able to show results (e.g. Result Presentation Portlet).

The Simple Search Portlet does not perform free text search, every search is always domain based and search profile compliant. The Simple Search Portlet does not allow the user to express an arbitrary query on an arbitrary RDF graph but only a predefined subset of them. This subset is defined in the search profile. The Simple Search Portlet will offer a limited set of graphical components for data displaying and / or entering. The graphical component used to display a given input is determined in the search profile. If special graphics components for a specific domain are needed, it is advisable to develop a specific user component for that domain and not use the Simple Search Portlet. The Simple Search Portlet is unable to show results for itself.

The GUI consist of three areas, a domain selection area, an area for property – value pair selection, and an area for search options selection. The domain selection area allows the user to select a domain of interest. The area for property - value pair selection allows the user to establish what resources he seeks to re-

cover by setting properties and values for those properties that a resource must have to be of interest. It will consist of an area to add new properties and an area that will display the already added properties. The area for search options selection allows the user to select search options. This area will consist of a check indicating whether inference will be used during the performed search.

On search button pressed, all resources considered relevant to the given query will be retrieved by the system and presented to the user through the configured Result Presentation Portlet.

3.6.2 Hierarchical Search Portlet

The Hierarchical Search Portlet allows a user to browse a category tree, displaying the name of the category and the number of resources that, for a given property previously set, has the category as value. When the user clicks on a particular category, the Hierarchical Search Portlet contacts the TaToo Discovery Server to retrieve the resources belonging to this category and its related metadata. To display this information, the Hierarchical Search Portlet redirects the results to a TaToo User Component able to show results.

The Hierarchical Search Portlet does not display hierarchies in a different format than the tree. The Hierarchical Search Portlet is unable to show results by itself.

The GUI will consist of two areas, a domain selection area and a category tree area. The domain selection area allows the user to select the domain of interest. The category tree area shows to the user a set of hierarchically organized categories. These categories will be retrieved by the component depending on its configuration and the selected domain ontology. For each category, the component displays the number of resources related to the category with the property specified in its configuration. On click over a category, all resources related to the selected category by the configured property would be retrieved and showed in the configured Result Presentation Portlet.

4 Semantic Case

TaToo encourages the adoption of the Minimum Environmental Resource Model (MERM) (PareienteLobo 2011) and its extensions as resource annotation model. MERM is defined as the largest common denominator between a set of heterogeneous description formalisms related to a common environmental resource. MERM can be refined to be used in different fields of the environmental domain by adding properties, modifying existing properties, adding new resource classes, etc. The MERM is formalized as an OWL ontology so TaToo resource annotations will be encoded as ontology instances stored as RDF triples. This structuring of the resource annotation greatly influences the subsequent search.

As resource annotations are always going to have a structure, TaToo queries are always going to be, from a strict point of view, structured queries. However, although structured queries are used, TaToo should allow the user to query the system without a thorough knowledge of structure of the resources annotated. TaToo will provide at least two search alternatives, a general or simple search, which allows users to search resources regardless of their structure, and a complex or advanced search, which allows users an intensive use of the structure of the resource to improve their searches performance.

As the annotations are encoded as instances of an ontology and stored in a knowledge base as RDF, there are two main ways to make queries:

- interacting with the knowledge base through a dedicated API.
- using a suitable query language.

Interacting with the knowledge base through a dedicated API is the most basic way to access its content. This is a flexible approach since it allows to arbitrary interaction with the content. However, this approach also offers significant disadvantages. The main disadvantage is that different knowledge bases (or semantic repositories) can provide different APIs, so it is difficult to implement generic and interoperable discovery processes following this approach. Currently there are initiatives that aim to improve this situation such as RDF2GO. RDF2GO is an abstraction over triple stores that allows developers to program against

RDF2GO interfaces and choose or change the implementation of the knowledge base or triple store later easily. However, although there are currently RDF2GO implementations for more widespread knowledge bases (Jena, Sesame, OWLIM), RDF2GO cannot be considered a standard, albeit a de facto one.

The other principal means of querying a knowledge base is by using a suitable query language. These languages provide a common syntax and semantics, so that all queries made in those languages are always interpreted in the same way, regardless of the underlying knowledge base. The clearest example of this would be the SQL query language for relational databases. SPARQL is the query language for RDF repositories recommended by the W3C. SPARQL allows users to query on different RDF repositories using a common language. SPARQL queries describe patterns of graphs that are matched against a repository. The result of a SPARQL query can be a result set or a RDF graph. SPARQL allows performing set operations over its results (union, intersection, etc).

However, from a general viewpoint, and from the particular TaToo viewpoint, both using a suitable query language and using a dedicated API share a common problem: they involve too much complexity, technological knowledge and knowledge of information structures to be useful to an end user. This is so because the purpose of both approaches is very general, namely to retrieve information on any graph of directed and named edges. Moreover, as we have argued above, in the scope of TaToo, resource annotations respond to a pre-defined structure, so the discovery process will only be performed on a very limited set of possible graph structures, so the expressiveness that allows us to pursue the mentioned approaches may not be necessary except in very specific cases.

Although the TaToo Core Components will interact with the Semantic Processor using SPARQL or a dedicated API the objective of the TaToo search tools is to enable users to query the system in a user-friendly way, hiding the aforementioned complexity. When defining these tools, the general strategy to follow will be to anticipate the user's information needs and build graphical user interfaces that allow the user to perform the most common queries to the system. Obviously, the tools designed following this strategy will, more or less, restrict the capabilities of the user to query the system. This effect can be mitigated by publishing a special GUI that allows experienced user to express queries in a more general way, such as SPARQL endpoint.

This will allow introducing new domains and having a configurable Graphical User Interfaces (GUI) by simply, making some configurations, without need for further development. This also provides an advantage when developing user interfaces for TaToo Validation Scenarios. Of course, in addition to these configurable interfaces, the TaToo architecture allows the development of specific interfaces that can meet special needs of different domains, resource types, etc. For more and further details on components dealing with Semantics see also (Avellino 2011).

5 Climate Twin Scenario

TaToo's Validation Scenario 1 "Climate Twins" (one of the three TaToo Validation Scenarios) is a web application that tries to bridge the gap between scientific climate change model outputs and the scientists' problem of transporting climate change implications to a broader public audience. To allow "real world insights" about future climate impact and appropriate adaptation, one can look at model regions, where the current climate appears similar to an expected future climate of a point of interest (POI). We call such region pairs with similar climate conditions (at different times) "Climate Twins" (Ungar/Peters-Anders 2011). From these (remote) current Climate Twin region parts one can learn "hands on" how future climate impacts may be experienced in the POI and how to adapt there to the changing climate conditions, expected in the future.

The idea is to identify regions whose current climate conditions show high similarity to the expected future climate in the POI. The Climate Twins search tool is a web-based graphical user interface (GUI) allowing for exploration of climate change effects based on maps of current and future climate (Figure 3).

Climate Twins has been chosen as a validation scenario for TaToo because it delivers valuable information for both: the public and stakeholders interested in future climate changes in their regions and should be traceable in the TaToo repository to be explored by interested people from various domains. On the other hand Climate Twins itself is interested in using TaToo services: Getting a picture of climatologically similar regions is just a first step – a lookup of information on the Twin Regions (i.e. metadata) will improve the application drastically (see below).

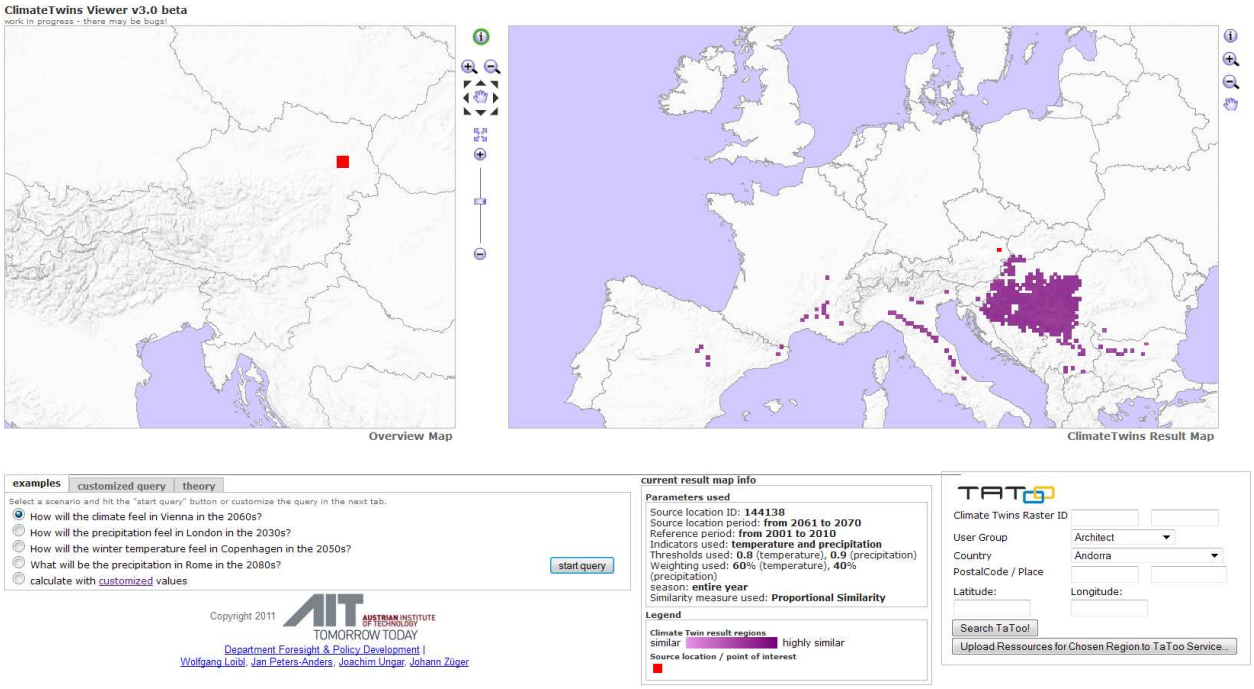

Figure 3 Climate Twins Web Application

Climate Twins is unique among the two other TaToo validation scenarios in the way that it does not exploit data sources per se, but delivers its data in forms of maps (not in form of a standardised protocol e.g.) and uses HTML web pages to transport the graphical representations of the model data behind the application. The uniqueness of the validation scenario lies also in the way Climate Twins will make use of the TaToo services: By using the TaToo repository as a "pool" of information on certain aspects of the Twin Regions. E.g.:

- An architect trying to explore future implications of climate changes in order to cope better with future climatological conditions might want to know more about the way houses are built in the "twin region" the system has identified
- a tourist manager might be interested in the touristic possibilities of the twin regions,
- a planer in the way how the infrastructure is organised in the twin region.

It is obvious that Climate Twins (like other application in the future) will be "enriched" by the TaToo services through an improvement of the application's functionality and the degree of information it can pro-

vide, in this way fostering the process of awareness rising concerning climatological changes in the future. In the case of Climate Twins there will be an additional frame placed in the GUI of the application (see Figure 3) which will hold a portlet (or proprietary code making use of a future TaToo API). This portlet/code will be connected to the TaToo repository and will make it possible (in the Climate Twin case) to do a TaToo search based on a geographical location, a user group and a question the user is interested in (e.g. the orientation of houses in a twin region). This semantic search approach will/shall lead to much better search results than in e.g. a syntactic search in a standard web search engine like Google.

6 Impact and Conclusion

Today a large number of environmental information sources are available and are going to be published.. However, to implement the vision of the Single Information Space for the Environment in Europe (SISE) it is not enough just to publish resources. Environmental information must be discoverable, and it must be „understandable in different contexts in order to be used effectively by parties (e.g. users, communities, etc) of various thematic domains.

But currently the deployment of a Single Information Space for the Environment in Europe is inhibited by the lack of a mechanism allowing data and service users to provide annotations that would add value to these resources. A middleware infrastructure as provided by TaToo and described in (Dihé 2011) is needed to fill this gap between environmental resources and end users. This framework needs to facilitate the life-cycle utility of environmental information from its collection and persistent storage to its discovery and purpose-oriented exploitation.

TaToo takes into account nearly all the needs and requirements as expressed in the different dimensions (i.e. User Dimension, Content Dimension, Data Dimension, Modelling& Decision Support Dimension, Information and Service Dimension) for SISE but specifically the support of discovery of relevant resources; the facilitation, integration and access to data sources; offering of metadata and incorporation of security.

TaToo's open approach allows to contribute not only to a single European Information Space for Europe, but also to the requirements of sharing information as expressed by many communities, not only in the environmental domains. Our deep hope is that TaToo can really contribute to mitigate the burden of creating "meaningful" (i.e. semantically enhanced) meta-information and communities taking part in SISE will take over our approach and benefit.

7 Acknowledgment

TaToo is a Collaborative Project (contract number 247893) co-funded by the Information Society and Media DG of the European Commission within the RTD activities of the Thematic Priority Information Society Technologies.

Bibliography

Božić B., Schimak G. (2010): Requirements document – V2, Deliverable 2.3.1 of TaToo Project, Public Document available at: www.tatoo-project.eu

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; ISBN 978-3-642-22284-9; ISSN 1868-4238; www.springerlink.com.

Dihé P., Schlobinski S. (2010): TaToo Semantic Service Environment and Framework Architecture – V1, Deliverable 3.1.1 of TaToo Project, Public Document available at: www.tatoo-project.eu.

Gruber, 1993 A translation approach to portable ontology specifications, Knowledge Acquisition 5 (2): 199-220, June 1993

ParienteLobo T., Fuentes Lopez J.M. (2010): TaToo Semantic Discovery Tools and Services – V1, Deliverable 3.2.1 of TaToo Project, Public Document available at: www.tatoo-project.eu.

ParienteLobo 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; ISBN 978-3-642-22284-9; ISSN 1868-4238; www.springerlink.com.

Petronzio L., Avellino G. (2010): TaToo Tagging Tools and Services – V1, Deliverable 3.3.1 of TaToo Project, Public Document available at: www.tatoo-project.eu.

Schimak, G., Usländer T., Havlik D., Argent R.: (2008) "Creating robust sensor networks - architecture and infrastructure" at: International Congress on International Environmental Modelling and Software (iEMSs 2008), Barcelona;07.07.2008 - 10.07.2008; in: "Proceedings of the International Congress on International Environmental Modelling and Software (iEMSs 2008)", Vol. 3 (2008), ISBN: 978-84-7653-074-0; 13 S.

Ungar J., Peters-Anders J., Loibl W., (2011): Climate Twins – An Attempt to Quantify Climatilogical Similarities 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. 429-436, ISBN 978-3-642-22284-9; ISSN 1868-4238; www.springerlink.com.

Usländer, T. (2007), Reference Model for the ORECHESTRA Architecture v.2.1, OGC “Best practices” paper o. 07-097, http://www.opengeospatial.org/standards/bp, (2007)

Avellino G., ParienteLobo T., Schimak, G., Rizzoli A.E., Dihé P. (2011): Applying semantics in the environmental domain: The TaToo project approach; ENVIP Workshop, EnvrioInfo 2011 Conference, ISPRA, Italy, ENVIP Workshop publication (to be published)