Jun 19th, 10:40 AM - 12:20 PM

# Ontology Mapping in Semantic Time Series Processing in Climate Change Prediction

Bojan Božić
*Austrian Institute of Technology*, bojan.bozic@ait.ac.at

Jan Peters-Anders
*Austrian Institute of Technology*, jan.peters-anders@ait.ac.at

Gerald Schimak
*Austrian Institute of Technology*, gerald.schimak@ait.ac.at

*International Environmental Modelling and Software Society (iEMSs)*
*7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,*
*Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)*
*http://www.iemss.org/society/index.php/iemss-2014-proceedings*

# Ontology Mapping in Semantic Time Series Processing and Climate Change Prediction

**Bojan Božić, Jan Peters-Anders and Gerald Schimak**

*Austrian Institute of Technology, Donau-City-Straße 1, Vienna, Austria (bojan.bozic@ait.ac.at, jan.peters-anders@ait.ac.at, gerald.schimak@ait.ac.at)*

**Abstract:** In today's time series processing there is more and more a need for addressing diverse user groups interested in a specific domain with appropriate user tailored time series data. The complexity of time series (e.g. involved data from different data sources and/or domains, visualization and representation, etc.) is growing rapidly. As a consequence, it means that users need to find a path through the jungle of time series data. After we have presented our concepts for semantic time series filtering and enrichment of time series with meta-information and annotations (Božić et al., 2012), we are now going to present a validation of these methods in the specific use case of climate change prediction. In this specific use case also called validation scenario, we demonstrate this approach in the Climate Twins application, which is a prediction model for geo-regions based on temperature and precipitation. The use case shows the how to select the right time series data and how to provide the right resources to the right user groups. In order to reach this goal, the idea has been to produce a domain specific ontology for a dedicated user group and to use it for the definition of basic discovery criteria of environmental respectively climate change related resources. The validation has been performed during the TaToo project, where a show case been developed to demonstrate the advantages of semantic enrichment for climate change prediction data using the Climate Twins application.

**Keywords:** Semantic Web; Time Series Processing; Climate Change; Ontologies; Metadata.

## 1 Introduction

The TaToo FP7 project[1] had the goal to develop tagging tools which are based on a semantic discovery framework. This means that, during the project, tools have been developed for users in order to share information as well as to discover new information relevant for their field of expertise. One of the several validation scenarios of the project was the “Climate Change Twin Regions - Discovery Platform” (short: Climate Twins). One of the results of the project was the development of an ontology for this platform which can be used to connect the platform to a knowledge base and, further on, to discover new resources which are semantically relevant. Additionally, it should serve as a basis for reasoning in the climate change area in order to improve the user's experience and generate connections between users and data (commonly represented as time series). In our paper, we show how users and time series data can be unified in certain groups of interest and therefore benefit by receiving customized data and sharing common topics and interests.

The remainder of the paper is organized as follows: Section 2 explains the presented validation scenario from the TaToo project. We present parts of the thematic background and the Web portal of the scenario. In Section 3 we show how the semantic time series processing approach has been validated by the Climate Twins ontology from the TaToo project. Section 4 summarizes relevant work in the fields of Semantic Web, Time Series Processing, and Community Building. Finally, Section 5 presents our conclusions and future work.

[1] http://www.tatoo-fp7.eu/tatooweb/

## 2 Validation Scenario Description

The Climate Twins application has the goal to show the impact of climate change and to allow adaptation for its users by finding model regions where the future climate of a certain Point-of-Interest (POI) is similar to the current climate of another location ((Ungar et al., 2011) and (Schimak et al., 2011)). These regions are called Climate Twins because they have a similar climate. The POI region which is subject to changes in the future can provide adaptation by knowing what the impacts of the future climate to the region will be. The tool for the search of Climate Twins regions is implemented as a web user interface which allows exploration of climate change effects on two maps. In the TaToo project, tools are used to improve the accuracy of the search for Climate Twins as well as the integration of further resources (Rizzoli et al., 2010). This means that TaToo tools have enabled the application to add tags, annotate resources, reuse tags of other users, discover and retrieve further resources for the selected region (documents, web sites, web services, etc.).

### 2.1 Ontology

Although similarity is based only on precipitation and temperature, there is a complex matching method which depends on quantification between data vectors and provides values which can be combined with other indicators as well (Božić et al., 2012). This makes it also possible to differentiate between more or less similar regions as results of a query. The ontology of the validation scenario, which has been developed during the TaToo project and used further on for ontology mapping, can best be explained by looking at the screenshots of the core application.

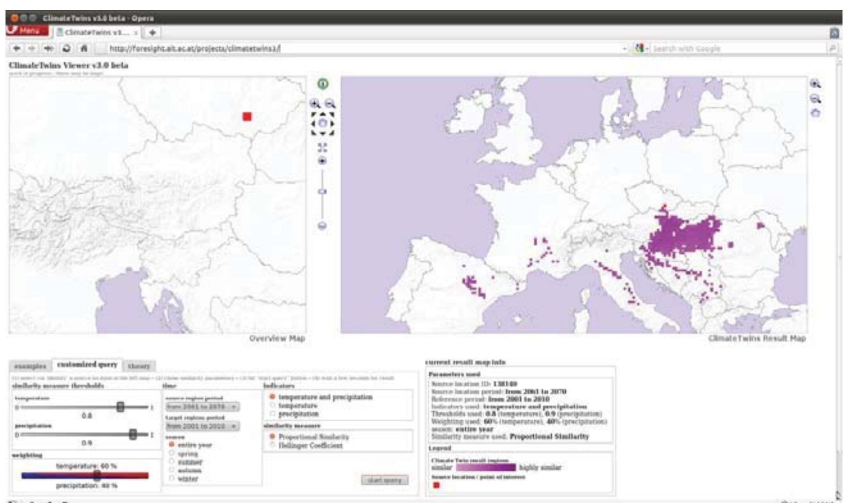


**Figure 1.** Screenshot of the Climate Twins application (Ungar et al., 2011).

Figure 1 shows a screenshot of the Climate Twins application. The concepts of the application, such as temperature and precipitation have been used as a basis for the development of the ontology.

Figure 2 shows an enlarged view of the Climate Twins control panel, and thus gives us a better perspective on the basis for the classes of the domain ontology. The ontology defines the concepts of temperature, precipitation, weighting, time as defined in the time series as well, and other indicators and parameters.

### 2.2 Ontology Mapping Approach

In our ontology mapping approach, we use the domain ontology of the validation scenario and connect it to a semantic time series processing bridge ontology (as introduced in Section 3). Connecting means that

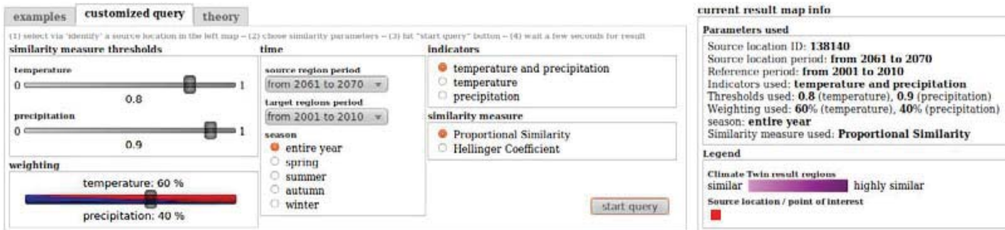


**Figure 2.** Enlarged screenshot of the Climate Twins control panel (Ungar et al., 2011).

relations are defined between the classes of the domain ontology and the classes of the bridge ontology.

## 3 Semantic Time Series Processing Validation

The validation of our Semantic Time Series Processing approach is done by the application of the Climate Twins ontology to a certain use case for the generation of groups which unite users and time series sharing the same domain of interest. We perform the following steps which also describe the workflow of the use case:

1. *Definition of users:* Creation of sample users from different fields (as described below) and different interests (required views on the data).
2. *Definition of time series:* Creation (import) of different time series from the climate area (related to prediction of climate change).
3. *Mapping of user (domain) and time series ontology:* Showcase of how the "Climate Twins" domain ontology is mapped to the bridge ontology (the bridge ontology approach is described in Xu et al. (2004)).
4. *Reasoning examples:* Showcase of how new RDF triples are inferred as a result of mapping the domain ontology to the bridge.
5. *Group generation:* Generation of groups with time series and users sharing similar topics and interests.

### 3.1 Ontology

The ontology of the Validation Scenario 1 "Climate Twins" (AIT ontology) is a result of the TaToo project. It has been developed to enrich the data model of possible applications based on the Climate Twins technology, and represents single components of the field of application which are represented as classes of the ontology. The Semantic Web for Earth and Environmental Terminology (SWEET)[2] and the Clean Energy Info Portal (reegle)[3] have been used as a basis for developing the "Climate Twins" domain ontology.

Figure 3 shows a simplified overview of the most important classes of the ontology and their properties. A building has a thermal process which defines how the temperature inside the building is regulated (defined by the `hasThermalProcess` property) and it uses a certain type of energy (defined by the `usesEnergy` property). The climate model used for a certain scenario has a certain reliability (property `producesReliability`) and can have a proportional similarity or Hellinger coefficient (`hasProportionalSimilarity` and `hasHellingerCoefficient`). The model can have an observed meteorological phenomenon (`observedMeteoPhenomena` property) and its subclass precipitation has a value (`hasPrecipitationValue`) and measure type (`hasMeasureType`). The precipitation value is defined by the precipitation value expression which has a minimum (`hasPrecipitationIntervalMin`) and a maximum

[2]http://sweet.jpl.nasa.gov/
[3]http://www.reegle.info/

(`hasPrecipitationIntervalMax`) for its interval as well as the exact value (`hasExactPrecipitationValue`). Furthermore, the temperature expression is defined by the type (`hasTemperatureType`), minimum (`hasTemperatureIntervalMin`) and maximum (`hasTemperatureIntervalMax`) thresholds, and exact temperature (`hasExactTemperature`).

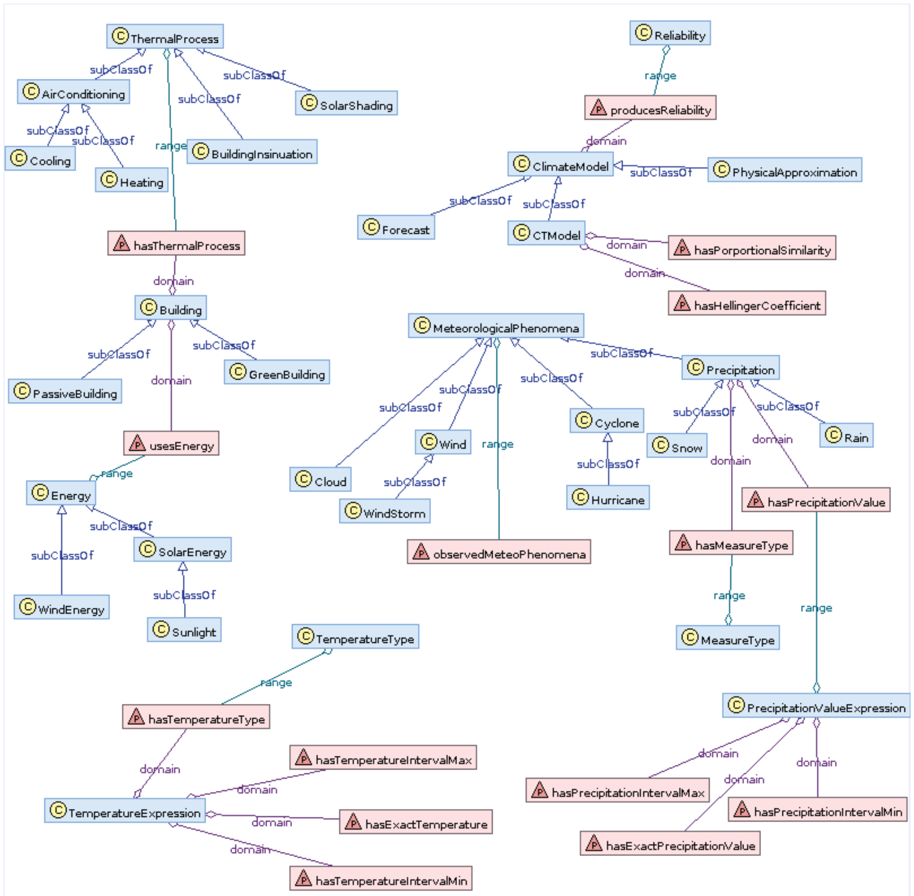


**Figure 3.** A simplified overview of the most important parts of the ontology.

## 3.2 Users

The most relevant user types in our Climate Twins Use Case are the following:

- *Politicians* and people working in public authorities who are interested in public issues linked to (the change of) climate conditions at a regional or local level. Such issues concern several policy areas like spatial planning, housing, agriculture and forestry, water and energy supply, etc.
- *Business managers* in industries which are climate-sensitive. This applies to all industries where renewable resources matter much in the production of their goods (e.g. food production, hydro energy). It also applies to industries where the climate is an important framework condition (e.g. tourism, construction).
- *Scientists* who work in fields of research related to the issues mentioned above and contribute through the results of their research to climate adaptations.
- *Non-professional users* who are only interested in information in trends and time series data of a certain field.

### 3.3 Example

Our example scenario covers a user who is a politician. This user works for the ministry of environment and is interested in time series data about the climate change. Furthermore, the user loads the Climate Twins ontology as her domain ontology to the system, which enables her to set “Climate Change” as her topic of interest. She needs to get data which supports her decisions to implement regulations and make investments in certain environmental fields.

The following RDF triples are most relevant for further processing to generate data of interest for the user (the classes and properties with the prefix `ct` - for climate twins - are defined in the domain ontology and the individuals – e.g. `:user1` – are defined by the Web portal, after creation of the user):

```
:user1 rdf:type ct:User .
:user1 ct:hasRole ct:Politician .
ct:Politician ct:hasEmployer ct:Government .
:user1 ct:worksFor ct:MinistryOfEnvironment .
:user1 ct:hasTopic ct:ClimateChange .
```

Since the user has the topic “Climate Change” selected from her domain ontology, she gets the list of time series which have this topic as a result.

The resulting time series are retrieved based on annotations from previous users or reasoning. In our case the time series are retrieved based on the annotation of the `Topic` subclass `ClimateChange`. All time series with this topic are retrieved from the system. This is only the starting point of the scenario. To experience the real work in the scenario we need to have a look at the ontology mapping and reasoning processing steps of our workflow.

As a first step towards group generation, the relevant parts of the domain ontology are mapped to the bridge ontology of our system. Listing 1 shows the merged parts of the two ontologies, where the domain ontology parts are marked in blue and the bridge ontology parts in red:

**Listing 1.** Mapping of domain ontology and bridge ontology.

```
(1)
ct:User owl:sameAs bridge:User .
ct:Topic owl:sameAs bridge:Topic .

(2)
ct:Government rdfs:subClassOf bridge:Institution .

(3)
ct:Building rdfs:subClassOf bridge:Topic .
ct:Energy rdfs:subClassOf bridge:Topic .
ct:Weather rdfs:subClassOf bridge:Topic .

(4)
ct:Reliability rdfs:subClassOf bridge:Subject .
ct:ClimateAdaptation rdfs:subClassOf bridge:Subject .
ct:ClimateMitigation rdfs:subClassOf bridge:Subject .

(5)
ct:TimeSeries owl:sameAs bridge:TimeSeries .
ct:TemperatureExpression rdfs:subClassOf bridge:Property .
ct:PrecipitationValueExpression rdfs:subClassOf bridge:Property .
ct:SpatialExpression rdfs:subClassOf bridge:Property .
ct:TemporalExpression rdfs:subClassOf bridge:Property .
```

In order to better explain the ontology mappings, we have split the mapped concepts in 5 parts (paragraphs).

The first part defines which classes of the domain ontology are equal to which classes of the bridge ontology. Part (2) defines subclasses for a bridge class. In parts (3) and (4) new subclasses for `bridge:Topic` and `bridge:Subject` are defined, which enables a user to be assigned to new topics (interests) and to create new types of subjects in order to add new resources to a topic. Finally, part (5) defines the concept of a time series for the Climate Twins domain ontology.

After the ontology mapping step, the domain and bridge ontologies can be used as one single ontology and form the basis for the reasoning process. Therefore, the following rules, which have been defined previously in Description Logic, need to be applied by a reasoner (Pellet):

- Rule 1 - $topic : bridge{:}Topic \models \forall bridge{:}hasTopic.user1 \sqcup \forall bridge{:}hasTopic.\ user2 \sqcup \forall bridge{:}hasTopic.user3$ - Extraction of relevant topics for users.
- Rule 2 - $ts : bridge{:}TimeSeries \models \forall bridge{:}hasTopic.topic$ - Retrieval of time series with relevant topics and subtopics.
- Rule 3 - $bridge{:}hasTopic.user \equiv bridge{:}hasTopic.group \Rightarrow (user, group) : bridge{:}hasUser$ - Assignment of users to groups.
- Rule 4 - $bridge{:}hasTopic.ts \equiv bridge{:}hasTopic.group \Rightarrow (ts, group) : bridge{:}hasTimeSeries$ - Assignment of time series to groups.

Rule 1 defines the extraction of relevant topics from user definitions. This means that we have a look at which topics the observed users are related to and collect them in a list. Of course, not only the topics themselves are taken into account, but also all subtopics (subclasses of the according topic class).

Rule 2 defines the retrieval of all time series with relevant topics and subtopics. This is achieved by generation of SPARQL queries and their application to the knowledge base which, in the next step, retrieves all IDs of the time series. These IDs can be used in order to retrieve the real time series from the time series database.

Rule 3 defines the creation of new groups based on extracted user topics and the assignment of the respective users to these groups. This is done by simply creating new instances of groups and extending them by triples with the properties `hasTopic` (for the topics) and `hasUser` for the users which need to be assigned.

Finally, in Rule 4, we also assign the right time series (with the same topic or subtopic) to the groups.

Another relevant part of the process is shifting users and time series between groups based on ratings and annotations. This means that we perform a periodic check of ratings and annotations of every user and time series. The check has the goal to evaluate whether a user or time series is in the right group. This is achieved by investigating which users have provided ratings and annotations for another user or time series. The critical point is reached when a user or time series receives a better rating and more annotations from users of another group, which initiates a shift of the element to the other group.

## 4 Related Work

A connection between semantic approaches and environmental modelling is drawn in Villa et al. (2009), which is relevant in order to connect models and ontologies and thus extend the process of meta-data generation. The principles behind ontology mapping methods, including our approach of using a bridge ontology, are presented in Choi et al. (2006), while Madhavan et al. (2002) deal with reasoning in mapped domain ontologies as is also used in our paper to infer new triples for group generation. Another related research field is Semantic Sensor Web as introduced in Sheth et al. (2008) where distributed environmental sensors are described, which are annotated with semantic metadata and made interoperable to provide contextual information for situational knowledge.

## 5 Conclusion

The goal of our paper was to validate the approach of semantic time series processing. We concentrated especially on the usage of a bridge ontology in order to map domain ontologies and to set rules for reasoning and inference of new meta-data. We used the Climate Twins ontology which has been developed in the TaToo FP7 project to demonstrate the validation in form of a use case.

The workflow of the validation process was the definition of a user, retrieval of time series data, mapping of single ontology concepts between bridge ontology and domain ontology, resoning over a bridge and domain ontology with new RDF triples as results, and generation of a group of interest. Based on this workflow we have shown how our approach is able to improve the area of climate change prediction with semantic time series processing.

In our future work the aim is to further develop domain ontologies and integrate new domains as well as to perform advanced reasoning methods, develop further rules for reasoning, and therefore improve meta-data generation in various fields which depend on time series data.

## Acknowledgments

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

## References

Božić, B., Peters-Anders, J. and Schimak, G.: 2012, Visualization and Filtering of Semantically Enriched Environmental Time Series, *Proceedings of 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*.

Choi, N., Song, I. and Han, H.: 2006, A Survey on Ontology Mapping, *ACM Sigmod Record* **35**(3), 34–41.

Madhavan, J., Bernstein, P. A., Domingos, P. and Halevy, A. Y.: 2002, Representing and Reasoning about Mappings Between Domain Models, *AAAI/IAAI* **2002**, 80–86.

Rizzoli, A. E., Schimak, G., Donatelli, M., Hřebíček, J., Avellino, G., Mon, J. et al.: 2010, TaToo: Tagging Environmental Resources on the Web by Semantic Annotations, *Proceedings of International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software Modelling for Environments Sake, Fifth Biennial Meeting, Ottawa, Canada*.

Schimak, G., Božić, B., Kaufmann, A., Peters-Anders, J., Dihé, P. and Pariente Lobo, T.: 2011, The TaToo Semantic Case – Requirements, impacts and applications, *Proceedings of the 25th EnviroInfo* pp. 832–847.

Sheth, A., Henson, C. and Sahoo, S. S.: 2008, Semantic sensor web, *Internet Computing, IEEE* **12**(4), 78–83.

Ungar, J., Peters-Anders, J. and Loibl, W.: 2011, Climate Twins - An Attempt to Quantify Climatological Similarities, *ISESS 2011*, TaToo Project, TaToo Project, Brno (CZ).
**URL:** *http://www.springerlink.com/content/14722218777588pj/*

Villa, F., Athanasiadis, I. N. and Rizzoli, A. E.: 2009, Modelling with Knowledge: A Review of Emerging Semantic Approaches to Environmental Modelling, *Environmental Modelling & Software* **24**(5), 577–587.

Xu, B., Wang, P., Lu, J., Li, Y. and Kang, D.: 2004, Bridge ontology and its role in semantic annotation, *Cyberworlds, 2004 International Conference on*, IEEE, pp. 329–334.