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

Lost in Translation - Mediating between distributed environmental resources.

Patrick Maué, Marcell Roth [a]

[a] *Institute for Geoinformatics (ifgi), University of Münster (WWU), Münster, Germany (firstname.lastname@uni-muenster.de)*

Abstract: Implementing distributed environmental modelling infrastructures supporting domain experts to discover, compose, share, and execute environmental models is subject of the ENVISION (Environmental Services Infrastructure With Ontologies) project. In this paper we focus on one specific aspect of distributed environmental models: how can we open up these infrastructures to include the wide range of data formats in the environmental modelling domain into the Web service compositions? We present the annotation pattern as solution to provide extensible resource descriptions based on flexible ontology languages. Model References link between different representations of resource descriptions, while Domain References link to terms in the common domain taxonomies. The use of shared ontologies minimizes the risk of semantic conflicts due to misinterpretations, and supports the mediation between distributed environmental resources.

Keywords: Environmental services, Semantic Annotations, Ontologies, Environmental Models

1 INTRODUCTION

The Model Web, initially suggested by Geller and Melton [2008], refers to a dynamic modelling infrastructure building on distributed services. Individual services are developed and maintained independently by researchers, policy makers, or public organizations. It embraces the idea of a Digital Earth which visualizes and relates historical, real-time, and forecasted data sets on a virtual globe. Models[1] are loosely coupled: its algorithms are separated from input data sources or front-end visualisation. Exposing models as Web services supports chaining of individual models, the comparison of different kinds of models, and their (ideally seamless) integration into end-user applications [Maué et al., 2011]. End users, ranging from novice users to peer scientists, are able to re-compose, adapt, re-configure, or execute the models and finally interact with the results.

The concept of the Model Web cuts across research on environmental modelling, geographic information science (GIScience), and distributed and service-oriented architectures. The fields overlap, but through time different standards and data formats have emerged. The World Wide Web Consortium (W3C) drafted generic standards for describing, accessing, and composing Web services. Supported by the European INSPIRE directive and national mapping agencies, the idea of Spatial Data Infrastructures (SDI) has been proposed in the GIScience community. SDIs built on services compliant to the standards by the Open Geospatial Consortium (OGC), which are not always compatible

[1] The term "model" is ambiguous. Whenever it is used in this article, it refers to environmental computer models. A model is our view always the "full package", including the input data source, model algorithms, and output data.

to the W3C standards [Tu and Abdelguerfi, 2006]. The environmental modelling community is less focusing on services. They heavily utilize sophisticated data formats which support the inherent complexity of multi-dimensional data. Building a distributed environmental services infrastructure across domains relies on tools which bridge the different formats and standards to come up with compositions of Web services which integrate SDI Web services and environmental data files.

In the research project ENVISION[2], we have developed a portal-based platform for the discovery, annotation, and composition of environmental services and models [Maué and Roman, 2011]. The implemented components are tested and validated with the help of different scenarios, with the following being the focus in this article. Evaluating response measures to oil spill disasters in the Norwegian Sea relies on models capable of predicting the drift of the oil slick. Response measures are either mechanical (e.g., skimming), thermal (e.g., in-situ burning) or chemical (e.g., dispersants). How to mitigate the effects of the spill depends on factors such as location, upcoming weather conditions, or the sensitivity of local wildlife. Public authorities have to react swiftly. A wrong decision can have a potentially devastating effect on the local environment and economy. Precision of the models is critical. Sophisticated algorithms have been developed which take various aspects into account when predicting the drift's speed, expansion, and direction. This includes real-time sensor data from buoys measuring the current, wind speed, wave height, or the salinity. Bathymetric data representing the sea depth is required to model the three-dimensional expansion of the oil slick below the surface. Data about the shoreline enriched with information about its resilience, or data representing wildlife sanctuaries, is considered to compute the potential impact of the oil spill. The core algorithms build on OSCAR, the Oil Spill Contingency and Response model, developed by Reed [1995] at SINTEF's institute for Marine Environmental Technology[3]. The input data sources are provided by various organizations such as the Norwegian Mapping Agency[4] or the Norwegian Meteorological Institute[5]. Three organizations, three different communities: the modelling algorithms have been exposed as W3C compliant Web service. The weather forecasts are published files in the GRIB (GRIdded Binary) file format, which is typically used in the meteorologic community. NMA provides their data as Web services compliant to the OGC standards.

With the ENVISION platform, specialists compose new models from existing Web services providing access to data sources and processing functionalities. But first, the user has to find the appropriate Web services. While being an expert in his own domain, he might have difficulties using appropriate terminology to discover all required input data sources. A "shoreline" might also be a "coastline", in our case it is in fact the "LandWaterBoundary". This is even more difficult for the model algorithms, which traditionally have cryptic names, undocumented input parameters, and rarely yield information about the simulated processes in the real world. Allocating all required input data sources requires some method to ensure syntactic, structural, and semantic interoperability. Connecting the services relies on means to map the appropriate data sources to the input parameters. But in particular for binary data sets this can be challenging. In the next Sec. 2 we explain how Semantic Web technologies can address the various challenges arising when crossing the information communities. It also introduces the simple data and service model vocabularies developed in ENVISION, and how discovery and mediation is enabled through semantic annotations. Sec. 3 presents the various implemented components to create and maintain the vocabularies and annotations.

[2]The acronym ENVISION stands for "Environmental services infrastructure with ontologies".
[3]See `http://www.sintef.no/home/Materials-and-Chemistry/Marine-Environmental-Technology/`
[4]NMA, see `http://www.statkart.no/`
[5]met.no, see `http://www.met.no/`

2 MAKING THE SEMANTICS OF ENVIRONMENTAL RESOURCES EXPLICIT

Environmental models are typically stand-alone applications, sometimes implemented decades ago and gradually improved over the time to accommodate for the growing knowledge we gain about our environment. The modularisation and transformation of the individual components to Web services is difficult, for complex and highly integrated computer models it might not be feasible. Relocating the models into the Web is also problematic in terms of licensing, the complexity of calibrating the models, or the scientist's fear of loosing control over the model execution. But the benefits such as the improved re-usability, the opportunity of peer scientists to re-execute and validate the model results, and savings in terms of used resources can in certain cases out-weight these drawbacks. These benefits only apply if the potential for semantic conflicts is considered.

2.1 Why semantics matters

Kuhn [2005] explains that the question for semantic interoperability results from the migration of geospatial tools and data sets into the Web. Semantic heterogeneities are usually not an issue within one information community. But conflicting interpretations of common terminology might even exist within small organizations. Semantic conflicts emerge in communication processes. Reasons can be manifold: semantic heterogeneities are commonly cited. Differences in language, culture, and knowledge in the field of interest are typical. Most are due to information asymmetry. Lacking necessarily information leads to misinterpretation (e.g., we have to assume that the water level is measured in meters if no further documentation is provided). Having semantic interoperability requires to minimize the potential of semantic conflicts, i.e., by making the interpretation of data semantics as explicit as possible. Shared vocabularies enriched with axioms constraining interpretations serve as descriptions of one particular perspective on reality. If this perspective is reflected in the data, semantic annotations make this link explicit. Semantically annotated resources are used in the ENVISION project to support information retrieval, mediation in the creation of the workflows, and dynamic adaptation of the workflows before execution.

The model predicting the oil drift is deployed as generic Web service compliant with the W3C standards. It computes the oil concentration both on the surface and in the water column. The output format is NetCDF, a binary machine-independent format for representing time series with three or four dimensions[6] It expects as input shoreline data served by an OGC Web Feature Service. It is no coincidence that these two services have potential for semantic conflicts. The Web Feature Service is considered to be compliant with the European INSPIRE directive, but even international legislation cannot ensure precise and extensive meta-data. Missing information is common. One reason is the lack of flexibility in the XML schemas used for the meta-data documents. Semantic annotations are simple extensions to the XML documents which link the original descriptions with flexible and extensible vocabularies capturing the data semantics. In this case we have conflicting terminology, lack of detail, and lack of descriptions about the inner relationships between the attributes. The feature "LandWaterBoundary" has three attributes: its geometry, "origin", and "waterLevelCategory" (as defined by the INSPIRE Thematic Working Group Hydrography [2009])

Conflicting terminology: A searching user, but also the expert specifying the expected input for the oil drift prediction service, might use terms like "shoreline", or "coast-

[6] A specific subset of the NetCDF standard has been recently adapted by the OGC.

line" to describe the feature of interest. Vocabularies which make such synonyms explicit are required to apply typical information retrieval techniques such as query expansion.

Lack of Detail: The meaning of the attribute "origin" remains cryptic to users unfamiliar with the data. In this case the code states if one particular part of the shoreline is either man-made (e.g., a quay or sea defence) or natural. The attribute "waterLevelCategory" indicates the water-level which this boundary is valid for. Possible values are, for example, "high water" and "low water". External documentation is required to understand these properties, but this might not always be available.

Dependencies: The boundary itself is defined as a one-dimensional line, it yields no information about the geographic entities it separates. But the attribute "waterLevelCategory" refers to the water level of the sea which is delineated by this boundary. The "origin", on the other hand, refers to the land. While this relation may seem obvious, it can be useful to make it explicit to support more complex queries.

Understanding the data helps to limit errors in the model and misinterpretation of the model results. Semantic annotations make the meaning of the properties explicit.

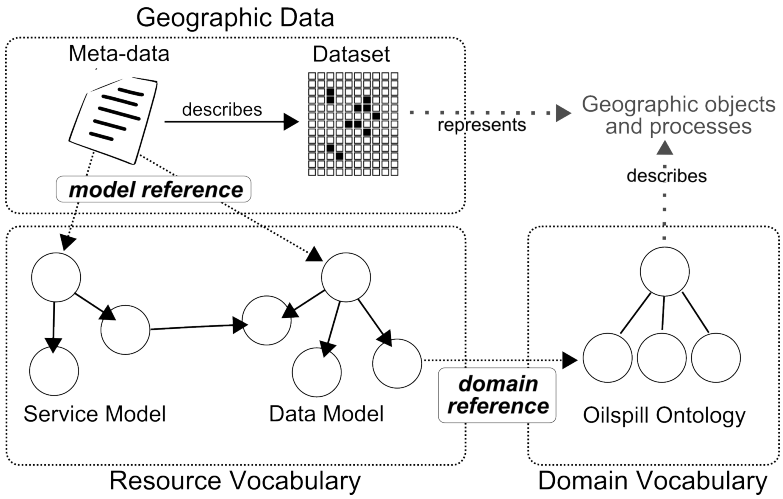

Figure 1: The Semantic Annotation Pattern. Meta-data describing external properties of a dataset are enriched with links pointing to service and data model vocabularies. A rule-based annotation approach aligns these vocabularies to domain ontologies.

2.2 The semantic annotation pattern

The ENVISION project refines and adapts the rule-based semantic annotation pattern originally developed by Klien [2007] and Grčar et al. [2009]. The application of semantic annotations for geospatial Web services have been discussed in detail in Maué et al. [2009]. Figure 1 illustrates the annotations pattern in ENVISION. The shoreline data set, served by an OGC Web Feature Service, is associated with some meta-data. It represents in this case the observed boundary between water and land. It comprises service-specific functional (i.e., where and how to access the service, restrictions of use, supported file formats and projections, ...) and non-functional information (i.e., contact information, quality of service parameters, ...). The meta-data additionally includes

details about the data model (i.e., the provided feature types and their attributes). The meda-data is typically based on some XML-dialect which allows for extending existing elements with model references pointing to a resource vocabulary. The model reference is, following the concept of the W3C recommendation for Semantic annotations of WSDL documents and XML Schema by Kopecký et al. [2007], linking between different representations of the same information. The resource vocabulary simply re-models the service- and data-specific qualities of the geographic data. But while the meta-data is encoded in a domain-specific language, the resource vocabulary is based on a common and extensible ontology language. The resource vocabulary is extended with rules which link the individual elements in the service and data model with their appropriate concepts from domain ontologies (which are ideally grounded in top-level ontologies).

Let us consider the semantic annotation of the property "origin". We assume that the following domain concepts exist: SHORELINE, GEOGRAPHICREGION, and ORIGIN. A rule which links the concept FEATURETYPE_LANDWATERBOUNDARY and its property *Property_origin* in the resource vocabulary could therefore state:

```
FeatureType_LandWaterBoundary(r1) and
Property_origin(r2) and
hasFeatureProperty(r1,r2) and
ShoreLine(d1) and
GeographicRegion(d2) and
Origin(d3) and
separates(d1,d2) and
hasOrigin(d2,d3)
 implies
  domainReference(r1,d1) and
  domainReference(r2,d3)
```

It says that the property origin (r2) of the feature type LandWaterBoundary (r1) represents the origin (d3) of the geographic region (d2) which is separated by the Shoreline (d1). Reasoning engines make use of these rules to resolve queries. In this case, a query for a Web service which serves boundary data (since ShoreLine is modelled as sub-concept of Boundary) would, for example, be resolved through this semantic annotation.

3 IMPLEMENTATION

Various tools, libraries, and vocabularies have been implemented in the ENVISION project to support the semantic annotations discussed before. The resource vocabulary builds on the Resource Description Framework (RDF), a language particularly useful to express graph-based models such as ontologies. The Service Model is based on the existing vocabularies for Procedure- or Resource-oriented Service Models (POSM/ROSM) [Krummenacher et al., 2010]. Service-type specific extensions exist to support the various service interface standards from the OGC, i.e., the Web Feature Service (WFS), the Sensor Observation Service (SOS), the Web Processing Service (WPS), and more. The data model vocabulary instantiate concepts from ontologies representing the OGC data standards, such as Geography Markup Language (GML), or Observation & Measurements (O&M). To provide, for example, machine-readable meta-data for sensor data provided by a SOS requires concepts representing observation data based on the O&M model. The domain ontologies are based on the same language. But while the data model vocabularies capture the inner-relationships of the data, the domain ontologies are supposed to capture one particular interpretation of a domain. The domain references separate between the local application-specific vocabularies describing one particular data set, and the globally shared domain vocabularies representing the terminology accepted within one information community. The semantic annotations are not limited to specific technologies. Various ontology languages exist. The Web Ontology Language

(OWL) is well accepted and is supported by a wide range of tools and libraries. But it lacks expressiveness and native support for rules. In ENVISION we instead use the Web Service Modelling Language WSML [Roman et al., 2006].

The service and data model vocabularies are the result of an automatic translation. The ENVISION Resource API[7] is a library which manages all resources required for the composition of environmental models. Besides the translation is also supports the publication of Web services to OGC-compliant catalogues, the deployment of compositions to a runtime engine, or the import of new resources into the user's collection. The translation to the service descriptions is achieved by the Service Model Translator (SMT). It produces the introduced service/data model vocabularies for a set of OGC Web services by fetching the original service description. It also creates W3C-compliant service descriptions (using the WSDL standard) for each service to enable support for the compositions. The Resource API also cross-links all resources (original and generated), and registers the updated original service description to the Semantic Annotations Proxy (SAPR)[8]. The proxy is discussed in detail in Maué et al. [2012].

The rules linking service-specific elements with concepts in domain ontologies have to be manually created. In the ENVISION project we are developing a graph-based annotation editor. The user can load a resource model into the view, search for appropriate concepts in the ontology, and then link the individual elements in the resource with these domain concepts. Searching for the concepts in ontologies can be challenging. The domain ontologies are therefore enriched with text automatically fetched from the Web. A user can then, for example, type in the keyword "coastline" and will also find the concept SHORELINE. This approach also enables multilingual discovery, searching with the German word "uferlinie" would yield the same result.

The annotations are stored as rules in the service model. The capabilities file, which has been registered to the semantic annotations proxy, includes pointers to its RDF-based service model. After the semantic annotation process, the resource is published to a common OGC-compliant catalogue service. An adaptor filters out the model references to the service model, and loads the service model into its knowledge base. All following search requests are then handled by the adaptor. If the discovery query includes a semantic query (specified with the same semantic annotation editor), the query as well as all currently registered service models are loaded into a reasoning algorithm. A query, for example, asking for any services which serve data representing shorelines will then return also the service providing the feature type "LandWaterBoundary". But also more complex queries are supported: a user might want to search for geometries delineating land or artificial coastal regions.

4 CONCLUSIONS

In this short paper we could only provide a very brief introduction in the methodology for semantic annotations implemented in the ENVISION project. Semantic annotations are required to reduce the possibility of mis-interpretations during discovery and mediation between different environmental services. The presented pattern based on rules supports more sophisticated annotations which can also model the dependencies and further details of certain elements in the data. The whole approach is building on a bottom-up philosophy. The goal of ENVISION is not to develop an integrated solution with all tools and ontologies bundled in one application. Instead, we provide a set of

[7] All components are FOSS and can be downloaded from ENVISION's open source project available at `http://kenai.com/projects/envision`.
[8] SAPR is a free service available at `http://semantic-proxy.appspot.com`.

libraries which can (but don't have to be) used in the platform. The choice of domain ontologies and resource vocabularies enable the integration. We didn't cover the actual composition and execution of environmental services. The presented tools also translate the original service descriptions into WSDL documents, which are re-used in the compositions. While this is straight-forward for OGC-compliant Web services, it is challenging for environmental data sets typically published as downloadable files.

ACKNOWLEDGMENTS

The presented research has been funded by the European project ENVISION(FP7-249120).

REFERENCES

Geller, G. N. and F. Melton. Looking Forward: Applying an Ecological Model Web to assess impacts of climate change. *Biodiversity*, 9(3&4), 2008.

Grčar, M., E. Klien, and B. Novak. Using Term-Matching Algorithms for the Annotation of Geo-services. In Berendt, B. et al., editors, *Knowledge Discovery Enhanced with Semantic and Social Information*, volume 220 of *Studies in Computational Intelligence*, pages 127–143. Springer Berlin/Heidelberg, 2009.

INSPIRE Thematic Working Group Hydrography. D2.8.1.8 INSPIRE Data Specification on Hydrography - Guidelines. Technical report, 2009.

Klien, E. A Rule-Based Strategy for the Semantic Annotation of Geodata. *Transactions in GIS*, 11(3):437–452, June 2007.

Kopecký, J., T. Vitvar, C. Bournez, and J. Farrell. SAWSDL: Semantic Annotations for WSDL and XML Schema. *IEEE Internet Computing*, 11(6):60–67, November 2007.

Krummenacher, R., B. Norton, A. Marte, A. Berre, A. Gómez-Pérez, K. Tutschku, and D. Fensel. Towards Linked Open Services and Processes. In Berre, A. J. et al., editors, *Proceedings of Future Internet - FIS 2010*, volume 6369 of *Lecture Notes in Computer Science*, pages 68–77, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

Kuhn, W. Geospatial Semantics: Why, of What, and How? *Journal on Data Semantics III*, 3534:1–24, 2005.

Maué, P., H. Michels, and M. Roth. Injecting semantic annotations into (geospatial) Web service descriptions. *Semantic Web Journal (SWJ)*, 0(0):1–10, 2012.

Maué, P. and D. Roman. The ENVISION Environmental Portal and Services Infrastructure. In Hrebícek;, J. et al., editors, *Environmental Software Systems. Frameworks of eEnvironment*, volume 359 of *IFIP Advances in Information and Communication Technology*, pages 280–294, Brno, Czech Republic, 2011. Springer Berlin/Heidelberg.

Maué, P., S. Schade, and P. Duchesne. Semantic Annotations in OGC Standards. OGC discussion paper, Open Geospatial Consortium (OGC), July 2009.

Maué, P., C. Stasch, G. Athanasopoulos, and L. Gerharz. Geospatial Standards for Web-enabled Environmental Models. *International Journal of Spatial Data Infrastructures Research (IJSDIR)*, 6(2011):145–167, 2011.

Reed, M. Quantitative analysis of alternate oil spill response strategies using OSCAR. *Spill Science & Technology Bulletin*, 2(1):67–74, March 1995.

Roman, D., J. De Bruijn, A. Mocan, H. Lausen, J. Domingue, C. Bussler, and D. Fensel. WWW: WSMO, WSML, and WSMX in a nutshell. In Shi, Z. et al., editors, *The Semantic Web - ASWC 2006*, volume 4185 of *Lecture Notes in Computer Science*, pages 516–522, Beijing, China, 2006. Springer Berlin/Heidelberg.

Tu, S. T. S. and M. Abdelguerfi. Web Services for Geographic Information Systems. *IEEE Internet Computing*, 10(5):13–15, 2006.