ROLE OF ONTOLOGIES IN CREATING HYDROLOGIC METADATA

Luis Bermudez[1], Michael Piasecki[2]

ABSTRACT

Recent developments and discussions on nationwide scales increasingly stress the need for semantic interoperability among communities due to the lack of specific domain descriptions of the data being processed. These shortcomings are largely based on the fact that each community typically only focuses on its specific needs with little or no attention paid to making these community-specific data descriptions part of a much bigger data realm. We state that part of the problem arises because the formalizations of the available metadata schemas are general, difficult to implement and inflexible to be extended. In this paper we attempt to show that through the use of the Ontology Web Language (OWL) and the creation of domain specific ontologies some of these shortcomings can be overcome. This is demonstrated by creating two examples of knowledge inference to reason from a domain ontology. In the first one we created an ontology for the US Geological Survey Hydrologic Units, where logical inference is use to extract a list of desire watershed names that a user could select to fill a metadata element. In the second example we show the possibilities of extending a metadata element and restricting it to accept values from another distributed resource, such as the Global Change Master Directory (GCMD). We conclude that knowledge representation systems like OWL provide a more flexible platform and possibility to reuse distributed resources, simplifying the process of creating metadata schemas for the hydrologic community.

1 INTRODUCTION

Metadata is needed to facilitate sharing of data among communities, minimizing duplication, reducing costs and facilitating efficient analysis and decision making. (Commission on Geosciences Environment and Resource CGER 1995). To create metadata, a user or a system needs a set of metadata elements that guides the creation process. These elements are typically arranged in a metadata model, catalog, or schema and are summarized in a specification document, or standard, that is published by the creating entity, like the International Standard Organization (19115:2003 Geographic Metadata Standard), Federal Geographic Data Committee, FGDC, or a specific community that creates its own standard like the Ecological Markup Language, EML.

---

[1] Graduate Student, Drexel University, Department of Civil Architectural & Environmental Engineering, 3141 Chestnut Street, Philadelphia, PA 19104, USA, Phone: (215) 895-1391; FAX: +(215) 895-1363, E-mail : leb27@drexel.edu

[2] Professor, Drexel University, Department of Civil Architectural & Environmental Engineering, 3141 Chestnut Street, Philadelphia, PA 19104, USA, Phone: (215) 895-1391; FAX: +(215) 895-1363, E-mail : m29@drexel.edu

For Hydrology, standard descriptions for gage stations, watersheds, well pumping observations and other hydrologic data are not explicitly available. However, by reusing, restricting or creating new elements from existing metadata models it should be possible to create a hydrologic metadata model to fit the needs of the hydrologic community. For hydrology related metadata models can be deduced from conceptual models that are described using the Unified Modeling Language (UML) or the Extensible Markup Language, XML, schemas. For example ISO's Geographic Metadata (ISO 2003) is published as diagrams using UML, while EML is specified using an XML schema.

Current formalizations of metadata models are difficult to use because they are very general and complex (Elmargarmid and Pu 1990; Stocks and Quinn 2002; Helly, Koppers et al. 2003), which translates into a lack of support by commercial software packages. For example, up to now there is no software that would allow extending a metadata model, or a web based application to create instances of ISO metadata.

Ideally, domain experts should be able to edit a metadata model and extend it to fit their own needs by reusing Web resources. For example creating a new concept, or a property, or restricting a property to have certain values or cardinalities of a resource available in the web. We explore a novel way to create metadata conceptual models and extend them, reusing distributed resources in the WWW through the use of the Ontology Web Language (OWL).

We found that OWL has richer expressions capabilities than object models in UML and XML schemas to create conceptual models. This is because OWL is a language that supports description logics while the other two do not. Using OWL it is possible to create logical statements like inverse, transitive, symmetric and functional relations. For example, defining that a *watershed* has only one possible *outlet* location is not possible in UML or XML schemas. UML and XML schemas will simply declare that a *watershed* has a property *outlet* with a cardinality of one. If a *watershed* is described in two different XML documents or UML instances, and the *outlet* location differs in each document, the instances will pass the XML schema and UML test. In contrast, in OWL we could declare that *outlet* is a functional property and the instances will not pass a semantic validation test.

In OWL it is possible to easily extend conceptual models distributed in the Web, reusing previously created resources, due the capabilities of the Resource Description Framework, RDF, to link concepts across the WWW. UML does not accept this, because it breaks the principle of modularization (Baclawski 2002) and XML schemas are not well suited for this purpose.(Hunter and Lagoze 2001; Gil and Ratnakar 2002; Hendler 2002). This is the reason why other related XML technologies have appeared to deal with this issue, like XPath and XLink. Also, representing restrictions in OWL is much more flexible because it allows multiplicity of restrictions on properties in a way that it does not affect the membership of objects in a class. (Baclawski 2002).

We present a definition of ontologies and conceptual models and then show with two examples the role of OWL as a conceptual schema that allows to create more flexible metadata models for hydrologic communities.

2 ONTOLOGIES

In computer science an ontology is an explicit and formal specification of mental abstractions, which conforms to a community agreement about a domain and design for a specific purpose. (Gruber 1993). It is different from the term Ontology (first letter in upper case) used in Philosophy to describe the existing things in the world. (Fonseca 2001). Different abstractions, specifications and agreements exist among communities, so different domain ontologies could exist, while only a single Ontology is possible.

An ontology provides the structure of the controlled vocabulary, similar to a dictionary or a thesaurus. The vocabulary agreed to by a community is the expression of concepts (mental abstractions) of their domain. Since a concept could be expressed in different ways and differ in meaning from one person to another, the controlled vocabulary helps to solve semantic incompatibilities. (Bishr 1998; Harvey, Kuhn et al. 1999; Sheth 1999; Hadzilakos, Halaris et al. 2000). For example, when conceptualizing the observation of the water level in a river US Geological Survey (USGS) refers to it as s*tage* while the National Oceanic and Atmospheric Administration (NOAA) refers to it as g*age height*. Also *stage* could be a hydrologic parameter but also a place for performing arts.

A formal specification of a vocabulary could be given in different ways, such as a plain list of words, a dictionary, a taxonomy, an Entity-Relational (ER) diagram, an Object Model in Unified Modeling Language (UML) diagram, an XML schema and possible many others. What makes a controlled vocabulary being an ontology is that in an ontology the concepts are defined explicitly by creating classes or entities. A class or entity is created using a mental abstraction, which could be a classification, an aggregation or a generalization (Batini, Ceri et al. 1992). For example, a list of terms such as: *USA, Germany*, and *Colombia* do not represent any explicit conceptual relation until an explicit class *Country* is abstracted to classify them. In addition to this requirement an ontology needs to conform to strict hierarchical subclass relationships between the classes (McGuinness 2003). Also, classes have properties and relations among them as shown in Figure 1.

In the small ontology example presented, the classes *BodyOfWater*, *River* and *Lake* are shown explicitly as boxes with the name of the class in bold in the first row. Properties are presented in the second and third rows. The property *connectsTo* applies to all the classes that are inherited from *BodyOfWater*, while *length* and *area* apply only to the local classes *River* and *Lake* respectively. Figure 1 is one of the many possible representations of an ontology. A given domain ontology should be understandable to members of a community and members of other communities, by describing it in a formal manner. A formal way to express ontologies is the Ontology Web Language (OWL).

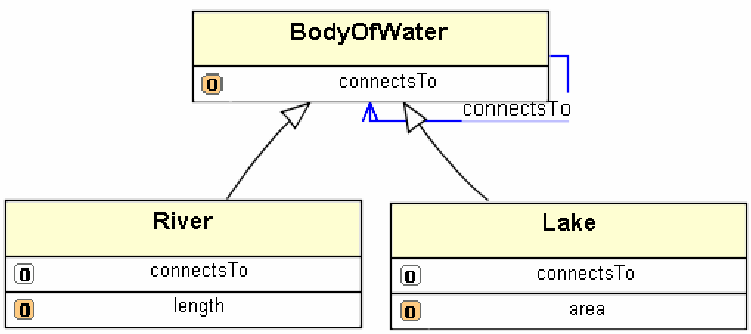

Figure 1. Small ontology example

Ontologies provide the mechanism to create the necessary classes and properties in similar way as object models. Ontologies in OWL, supports logical statements like inverse, transitive, symmetric and functional relations, that allows richer semantic declarations for creating control vocabulary that could be used in metadata schemas.

3 CONCEPTUAL MODELS

The development of the metadata model is similar to the development of an information system. It starts with the specification requirements, that answers the question why the metadata is going to be created and how it is going to be used. The requirements are presented as a list of possible elements. Some elements could then be put together under entities and related to each other to facilitate the understating of the model. The rearrangement of elements, creating classes for entities, presenting elements as properties of classes and the relating classes to each other is what is called a conceptual model.

A concept is a mental abstraction of a real world object. Concepts are related to each other via statements like *isA*, *isPartOf* or *isMemberOf* (Batini, Ceri et al. 1992) and contain some characteristics called properties. A set of statements is a conceptual model that helps domain experts to express formally a system or domain.

Conceptual Models are formalized in diagrams, like Entity-Relational (ER) diagrams, Unified Modeling Language (UML) and in ontologies like OWL. UML is the current standard of the Object Management Group (OMG) to create models, and is also used by ISO and the OpenGIS consortium to share their conceptual models. OWL is a recommendation of W3C to specify ontologies and while it shares many similarities it has some advantages over UML.

Differences and similarities between UML and DAML+OIL and RDF models are discussed by (Baclawski 2002). Since OWL is very similar to DAML+OIL, most of the analysis done by Baclawski applies also to OWL. Both OWL and UML allow explicit declarations of classes and properties, generalization relations, datatypes, restriction of properties, and declaring container for classes. However, a *property* in UML and OWL are very different. In UML, properties are binary relations whilst in OWL, properties could have complex domain and ranges and could be restricted multiple times in different classes. This allows flexible extensions in OWL, not possible in UML.

4 OWL: RICH SEMANTIC DECLARATIONS

Description logics (DL) allow declaring logical statements that UML is not able to express, such as inverse, transitive, symmetric and functional relationships. Description logics are used to build intelligent applications that allow a system to reason, and make deductions based on explicit representation of knowledge. For the creation of metadata we use these OWL-DL capabilities to identify resources that can be used to create dynamic user interfaces for creation of metadata instances. Also, these resources are used to validate the semantics of metadata instances.

Metadata models declare elements and the domain values of these elements. We create a statement to assert that an element could have a set of finite values. This statement is an assertion that makes use of the inverse, transitive or symmetric logical expressions.

Suppose that the element *MD_Identifier* of the ISO 11915 standard is defined by a hydrologic community to permit only names of watersheds located in a particular region. Using UML, it would be necessary to create the exact list of the watershed codes or names in an enumeration or codelist. In contrast, using OWL it is only necessary to declare a statement, that refers to an already existing hydrologic-units-ontology. Then, a knowledge system, will be able to infer the values and use them as required.

Figure 2, shows an ontology for the hydrologic unit system used by the US Geological Survey. The hydrologic unit system is a hierarchical classification of nested large-to-smaller watersheds within a certain region. Based on this ontology a system that handles knowledge inference uses this ontology to get values to either populate a predefined list in an input form or to validate the data semantically. A statement could be something like:

"the property *MD_Identifier* of *EX_Geographic_Description* is restricted to allow all cataloging units that are part of the *Subregion Delaware* ".

In this particular case, the cataloging units *Schuylkill* and *Lehigh* will appear on the list, while hydrologic units that are part of the Potomac should not be included. *Schuylkill* will appear because *Schuylkill* is part of the *Lower Delaware*, and *Lower Delaware* is declared to be part of *Delaware*. Since *Is Part Of* is declared to be a transitive property, the system will infer all that *Schuylkill* is also part of *Delaware*.

It should be pointed out that other concepts that could be used in populating metadata, should also be declared in an ontology. These include geographic locations (e.g. name of stations), instruments (e.g. in-situ devices, remote sensors) and properties of observed phenomena (e.g. stage, precipitation intensity).

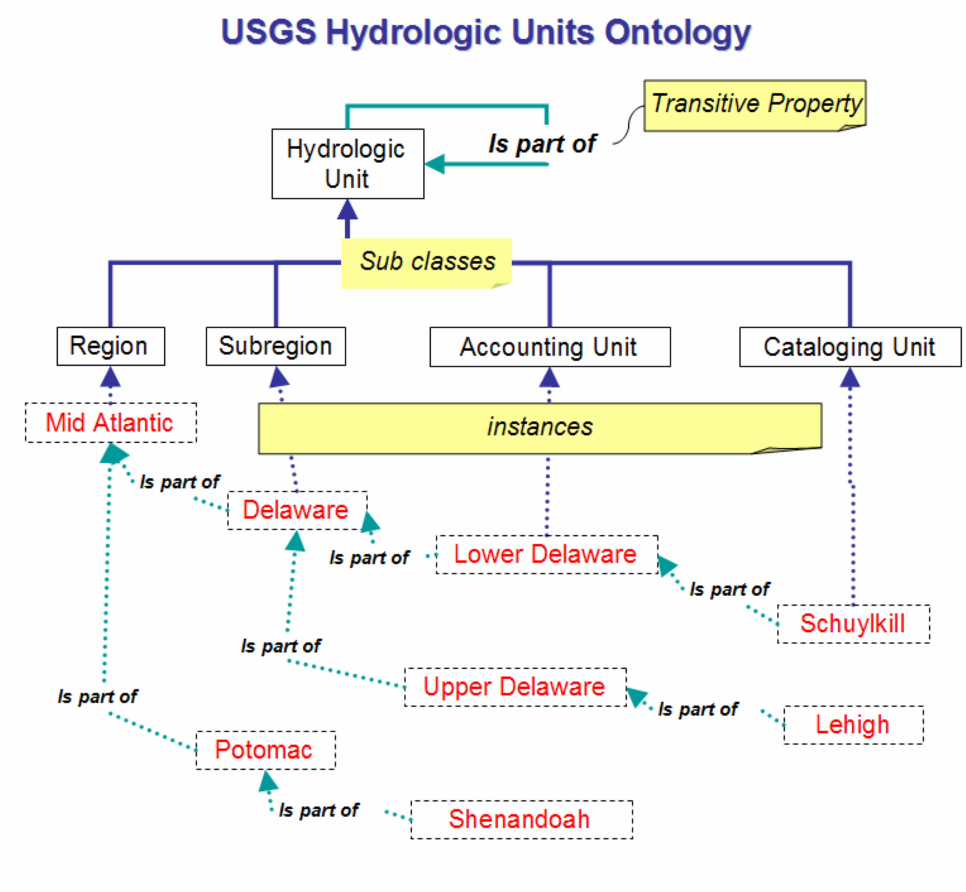

Figure 2 USGS Hydrologic Units Ontology

5 FLEXIBLE EXTENSIONS OF METADATA PROPERTIES WITH OWL

Representing restrictions in OWL is much more flexible than UML because it allows multiplicity of restrictions on properties in way that it does not affect the membership of objects in a class. (Baclawski 2002). This is done indirectly by stating that the class that is restricting the property is a subclass of a class called *restriction*. Nonetheless, we use this OWL feature to restrict metadata elements with success.

In OWL restrictions can be applied on properties declaring a different cardinality or a different range. Suppose that the element *iso:keyword* should be restricted by a hydrologic community to have all values related to surface water from a web catalog of scientific keywords. For example, such a catalog could be the Global Change Master Directory (GCMD), which should also be expressed as an ontology. Figure 3 shows a class named *MD_Keywords_EXT*, which is a subclass of *iso:MD_Keywords.* It extends the property *iso:keyword*, but it also restricts it to allow *allValuesFrom gcmd:Surface_Water*.

It is important to note that the logical reading of this statement is "all individuals that have values for the property *iso:keyword* of type *gcmd:Surface_Water* are of type *MD_Keywords_EXT*. A system to collect hydrologic metadata for the above example could make sure that all the individuals meet the requirement of having the element *iso:keyword* to be either *gcmd:Discharge* or *gcmd:Stage_Height* or any other value of type *gcmd:Surface_Water*..

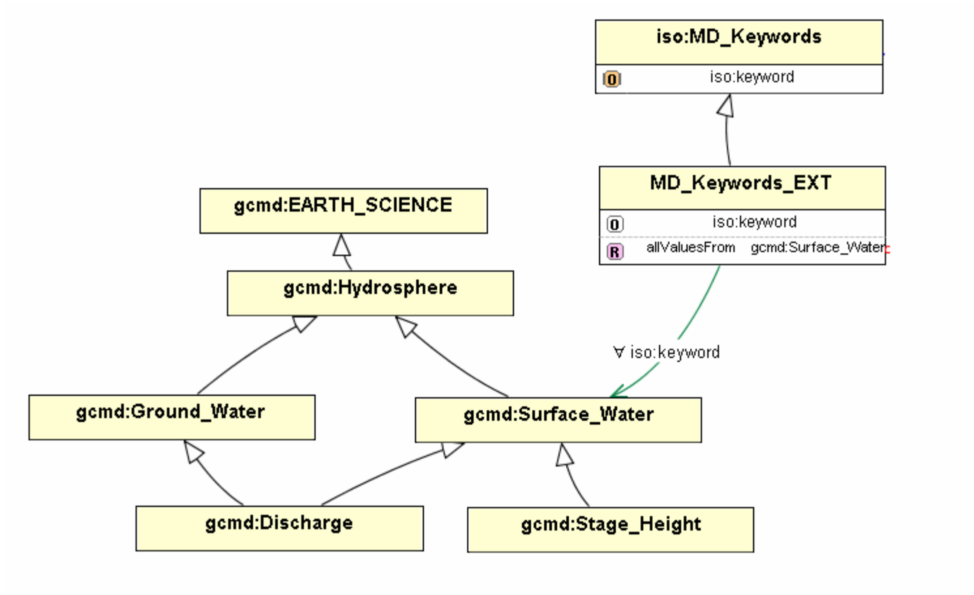

Figure 3 Extension of iso:keywords

6 SUMMARY

This paper outlines the role that ontologies can play for the creation of interoperable metadata sets for a specific realm, in this case the hydrologic community. We stated that knowledge representations systems, like OWL, permit a larger degree of flexibility for creating metadata models than conventional conceptual models expressed in UML or XML schemas. Creating a logical statement was sufficient to restrict metadata elements to a finite set of values from a control vocabulary. This could be achieved due to the richer semantics declarations possible in ontologies and the capabilities of knowledge inference of description logics. Also, we showed that in OWL models it is possible to apply, indirectly, restrictions on properties so that they could conform to specific needs of hydrologic communities.

For ontology-driven information system to work, a consensus among communities is also needed. OWL is a promising new technology that is opening up far reaching possibilities for the Semantic Web World (Berners-Lee, Hendler et al. 2001) which carries the promise of much improved human- machine interactions. Scientific communities like the hydrologic community are poised to take advantage of this technology to solve semantic problems and share metadata models on a broader scope reaching across communities.

ACKNOWLEDGEMENTS

This work has been supported by the National Ocean Partnership Program, NOPP, through NASA grant number NAG13-0040, and by the National Science Foundation, NSF, Geoscience program. We would also like to acknowledge the valuable contributions by the members of smileConsult, GmbH in Hannover, Germany as well as the many discussions we had with individuals from groups like the openGIS consortium JENA, RDF-logic, and Stanford University (makers of PROTÉGÉ).

7 REFERENCES

Baclawski, K., Kokar, M., Kogut, P., Hart, L., Smith, J., Letkowski, Jerzy, Emery, Pat (2002). "Extending the Unified Modeling Language for ontology development." Software System Model **1**: 1-15.

Batini, C., S. Ceri and S. B. Navathe (1992). Conceptual Database Design. Redwood City, California, The Benjamin/Cummings publishing Company, Inc.

Berners-Lee, T., J. Hendler and O. Lassila (2001). "The Semantic Web." Scientific American **184**(5): 34-43.

Bishr, Y. (1998). "Overcoming the semantic and other barriers to GIS interoperability." Geographic Information Science **12**(4): 299-314.

Commission on Geosciences Environment and Resource CGER (1995). A Data Foundation for The National Spatial Data Infrastructure. Washington, D.C., National Academy Press.

Elmargarmid, A. and C. Pu (1990). "Guest Editors' Introduction to the Special Issue on Heterogeneous Databases." ACM Computing Surveys **22**: 175-178.

Fonseca, F. T. (2001). ONTOLOGY-DRIVEN GEOGRAPHIC INFORMATION SYSTEMS. Spatial Information Science and Engineering. Maine, The University of Maine.

GCMD. Global Change Master Directory. Retrieved January 06, 2004, from http://gcmd.gsfc.nasa.gov/Resources/valids/gcmd_parameters.html.

Gil, Y. and V. Ratnakar (2002). TRELLIS: An Interactive Tool for Capturing Information Analysis and Decision Making. A. Gómez-Pérez and V. Richard Benjamins(eds.). Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web : 13th International Conference, EKAW 2002, Siguenza, Spain, Springer-Verlag Heidelberg. p. 37 - 42.

Gruber, T. (1993). "A Translation Approach to Portable Ontology Specification." Knowledge Acquisition **5**(2): 199-220.

Hadzilakos, T., G. Halaris, M. Kavouras, M. Kokla, G. Panopoulos, I. Paraschakis, T. Sellis, L. Tsoulos and M. Zervakis (2000). ""Interoperability and Definition of a National Standard for Geospatial Data: The Case of the Hellenic Cadastre",." International Journal of Applied Earth Observations and Geoinformation **2**(2): 120-128.

Harvey, F., W. Kuhn, H. Pundt and Y. Bishr (1999). "Semantic interoperability: A central issue for sharing geographic information." The Annals of Regional Science **33**(2): 213-232.

Helly, J., A. A. P. Koppers and H. Staudigel (2003). "Scalable models of data sharing in Earth sciences." Geochem. Geophys. Geosyst **4(1), 1010, doi:10.1029/2002GC000318**.

Hendler, J. (2002). "XML and the Semantic Web." XML Journal **October**.

Hunter, J. and C. Lagoze (2001). Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles The Tenth International World Wide Web Conference, Hong Kong, ACM Press, May 1 - 5 2001. 457-466.

ISO (2003). "Geographic Information - Metadata."

McGuinness, D. L. (2003). Ontologies Come Age. In D. Fensel, J. Hendler, H. Lieberman and W. Wahlster (ed.)^(eds.) Spinning the Semantic Web. London, England, The MIT Press.

Sheth, A. P. (1999). Changing focus on interoperability in information systems: from system, syntax, structures to semantics. In M. F. Goodchild, M. J. Egenhofer, R. Fegeas and C. Cottman.(ed.)^(eds.) Interoperating geographic information systems. Boston, Kluwer Academic Publishers**:** 5-29.

Stocks, K. and J. Quinn (2002). Data technologies: Geospatial data integration. W. Michener and P. Tooby(eds.). Scalable Information Networks for the Environment (SINE). Report of an NSF-sponsored workshop, San Diego Supercomputer Center, Oct. 29-31 2001. pp 23-29.