INTEROP NETWORK TO SUPPORT GEOSPATIAL DATA SEMANTIC INTEROPERABILITY

Nancy Wiegand
Senior Scientist
Space Science and Engineering
University of Wisconsin-Madison
Madison, WI 53706
wiegand@cs.wisc.edu

ABSTRACT

This work describes a new NSF funded INTEROP project organized by the Spatial Ontology Community of Practice (SOCoP, www.socop.org). Geospatial data, which give location information for geographic features, are pervasive across many disciplines and fundamental for diverse applications, such as economic development, natural resources, environmental protection, and emergency response. But, re-using geospatial data remains difficult because of its heterogeneity. Although progress has been made by initiatives such as the National Spatial Data Infrastructure (NSDI) and Geospatial One-Stop to distribute data, it is now necessary to address issues of standards harmonization and agreement on the meanings of relevant concepts in diverse data sets due to different community views. The lack of semantic interoperability has been recognized as a stumbling block to collaboration. As a solution, well designed, formal ontologies, relying on agreements between communities on unambiguous representation of concepts and relationships for a given problem area, can be used as an analytic tool to bridge community gaps. The purpose of this INTEROP project is to create opportunities for the geospatial community to work on semantic agreements and use semantic mappings and other emerging Semantic Web technologies. An initial INTEROP Network is based on the existing SOCoP network that has a core group of active participants from academia, government, and industry. We now want to broaden the Network to cover various domains and include more participants. One of the purposes of this paper is to include the ASPRS community and gather additional use cases.

KEYWORDS: semantics, geospatial data, ontology, interoperability

INTRODUCTION

The purpose of this paper is to describe a new National Science Foundation (NSF) grant to work on semantic interoperability for geospatial data and engage the ASPRS community in the work. For example, we are seeking use cases for semantic problems relevant to ASPRS. We are also looking for people in the ASPRS community to participate as informal members of a network.

The project is funded by the interdisciplinary INTEROP program of NSF, and the work is being conducted by the Spatial Ontology Community of Practice (SOCoP, www.socop.org). SOCoP is a national level group of practitioners, academic researchers, federal agency workers, and industry representatives. The group was formed in 2006 based on recognition of the need for semantic interoperability for geospatial data and the potential of emerging semantic technologies, ontologies, and formal representations as solutions. The purpose of SOCoP is "to foster collaboration among researchers, technologists, and users of spatial knowledge representations and reasoning, towards the development of spatial ontologies for use by all in the Semantic Web" (www.socop.org).

Geospatial data include location information for geographic features and are pervasive across many disciplines and fundamental for diverse applications, such as economic development, natural resources, environmental protection, and emergency response. But, re-using geospatial data remains difficult because of its semantic heterogeneity. Although progress has been made by initiatives such as the National Spatial Data Infrastructure (NSDI) and Geospatial One-Stop to distribute data, it is now necessary to address issues of standards harmonization and agreement on the meanings of relevant concepts in diverse data sets due to different community views. The lack of semantic interoperability has been recognized as a stumbling block to needed collaboration. As a solution, well designed, formal ontologies, relying on agreements between communities on unambiguous representation of concepts and relationships for a given problem area, can be used as an analytic tool to bridge community gaps. The INTEROP project intends to create opportunities for the geospatial community to address semantic problems and create solutions.

A goal of the INTEROP project is to create a large Network of people to contribute use cases of semantic heterogeneity in geospatial data and work collaboratively to create solutions. SOCoP forms the initial Network, but one of the purposes of this paper is to reach out to broaden the Network to cover various domains and include more participants. The Network serves as an umbrella over more specific domains that will be linked through the Network. We describe the project further and invite participation from the ASPRS community.

MOTIVATION

This project is concerned with semantic heterogeneity. Although geospatial data can be heterogeneous in file formats, attribute organization, and coordinate systems, those types of differences can be resolved, for example, using coordinate conversion routines or ETL (extract, transform, and load) techniques. But, the use of different terms and conceptual notions of a domain still poses significant challenges. The proposed solution is to create formal knowledge bases or ontologies to describe a domain and use these to create semantic agreements and resolve semantic differences. Such knowledge bases hold terms that are generally accepted in a community. These terms can potentially be used in data sets created in the future and also be used to resolve existing heterogeneity in legacy data through mappings from ontology terms to local terms.

Prior solutions to avoid or resolve semantic differences were to create and enforce standards or resolve differences on an application by application basis. Although standards can be a good solution, they may be difficult to create, and, if very general terms are used as standard terms, they might not fully cover the nuances and needs of local data. In any case, legacy data may not conform to the standards. And, as to 'one-of' solutions on an application by application basis, after enough unique 'one-of' solutions are made, the need is recognized for a more universal solution. Creating and then consulting a comprehensive knowledge base for a domain, i.e., an ontology, results in a re-usable solution across many applications.

There are various scenarios in which semantic heterogeneity needs to be addressed. One scenario is in searching for geospatial data. In addition to keyword search, along with associated synonyms, the vision is that search would be expanded or made more precise by including ontology information, such as superclass terms, subclass terms, or terms with other types of relationships to the initial search term. Ontology-based search could be part of Spatial Data Infrastructures (e.g., Hochmair, 2005). Another scenario is querying data. An example is querying over local data sets to get an aggregated result when each local data set uses different terms and a different classification of terms. This happens with land use coding systems, for example, and other such classification systems (e.g., zoning, wetlands, land cover). In these examples, setting standard terms may not work because such terms would likely be at a high level of description (e.g., agriculture) and lose the needed local detail (e.g., orchard).

One of the purposes of the SOCoP INTEROP grant is to create the knowledge bases that can be used for enhanced search for geospatial data and also for resolution of semantic heterogeneity in querying. This effort is in the vein of work regarding the Semantic Web (Berners-Lee et al., 2001) and the development of other semantic technologies.

SEMANTIC TECHNOLOGIES

The vision of the Semantic Web can be thought of as two-fold, although both aspects involve semantics. One aspect is to create a Linked Web of Data, also called Linking Open Data (LOD) or a linked data cloud. The other aspect is creating the semantic knowledge bases called ontologies.

Linked Open Data

One of the visions of the Semantic Web is to have data linked across the Web. Figure 1 shows the classic example of a linked data cloud. This linked open data targets the Life Sciences area.

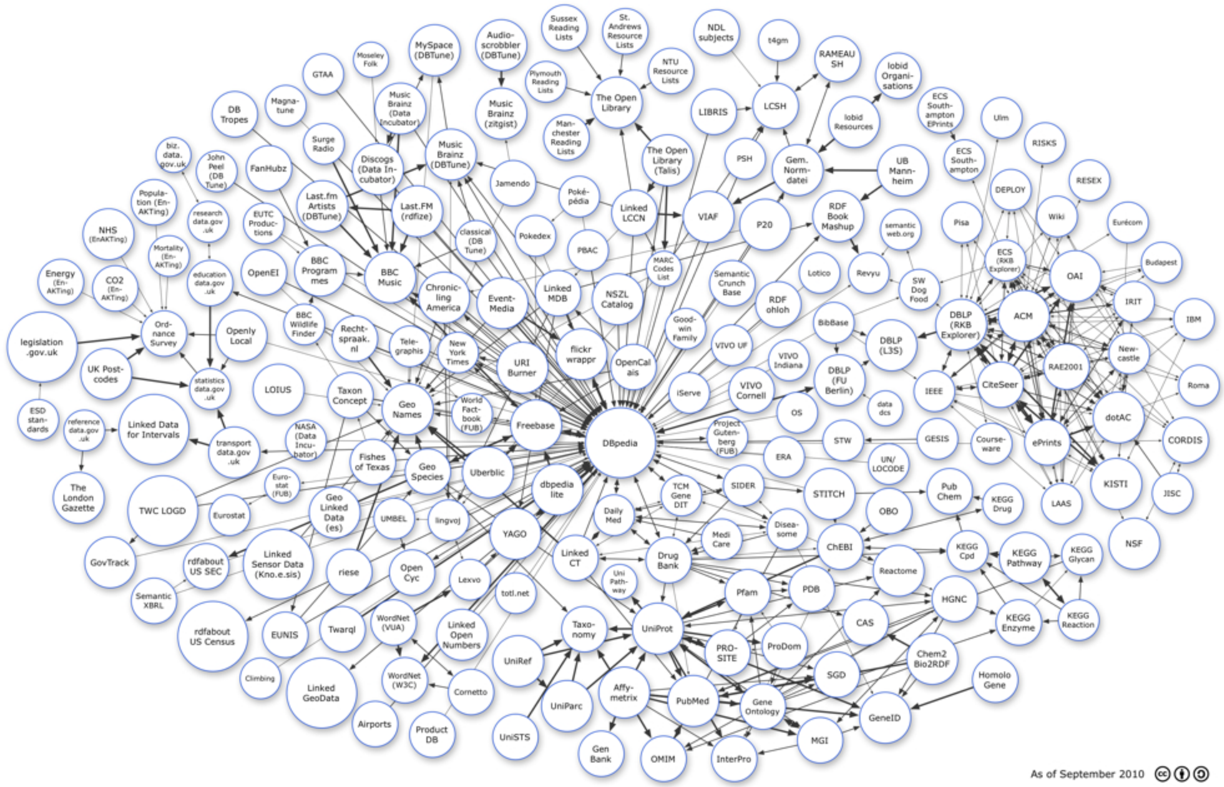

Figure 1. Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/.

Each of the circles in Figure 1 represents a knowledge base for a separate domain. Each knowledge base consists of information represented as triples using the Resource Description Framework (RDF) (Beckett and McBride 2004). RDF is an enhanced form of XML, and an RDF triple consists of a subject, predicate, and object. Examples are river_x *flowsInto* lake_y or pond *typeOf* water_body. If represented in graph form, the subject and object are nodes connected by an arc which describes the relationship between them. Because the subject of one triple could be the object of other triples, as well as being the subject of other triples, an interconnected graph is formed. Further, each component of a triple (i.e., subject, predicate, or object) is identified by a URI (Uniform Resource Identifier). URIs are similar to URLs for Web pages and uniquely identify an entity, such as a person, term, or concept. Knowledge bases representing a domain are formed from interconnected triples resulting from matched URIs, such as DBpedia or PubMed in Figure 1.

Further, if another knowledge base consisting of a set of RDF triples exists on the Web to describe another domain and it uses the same URIs to represent the same entities as other knowledge bases of triples, even more linking is achieved. For example, many other knowledge bases are using the URIs established by DBpedia. This creates a larger graph and enables further relationships to be explored as more and more data are linked. If identical URIs do not exist but concepts between graphs are the same, the URIs can be declared as 'sameAs' to still achieve the desired connections. Again, Figure 1 shows the classic diagram of a linked data cloud. This linked information on proteins, genomics, and pharmaceuticals allows Life Sciences data to be combined and explored in ways that were difficult before. There is the possibility of discovering new knowledge from these extended links.

Figure 1 also has a few nodes relevant to geospatial data such as GeoNames, Census data, and Linked Geodata. More geospatial data will be put into RDF creating a linked data cloud for the geospatial area.

Ontologies

A second aspect of the vision for a Semantic Web is to create knowledge bases or ontologies that describe a domain more fully. The nodes in Figure 1 may already be ontologies, or they may just be collections of data converted to RDF without modeling a full domain. Formal ontologies are represented using an ontology language, such as OWL

(McGuinness and van Harmelen 2009). OWL stands for Web Ontology Language, with the letters reversed. To represent ontologies, OWL uses description logic and adds structure and constructs to the RDF language. OWL provides schema-type information in addition to allowing specification of constraints and relationships such as transitive.

An ontology contains terms, concepts, and relationships that describe a domain. An ontology is often organized into classes and subclasses but differs from a taxonomy in that other kinds of relationships are also included. For example, Figure 2 graphically depicts that 'river' is a subclass of 'water body' along with 'ocean' and 'lake'. It also shows partonomy and a potentially recursive relationship.

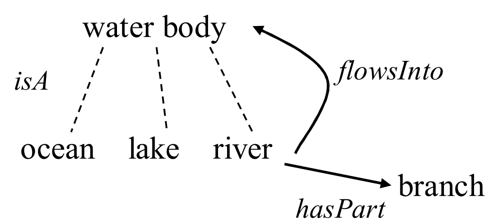

Figure 2. A small ontology.

An ontology expressed in a formal ontology language, such as OWL, is machine readable. This allows automated inference of subsumption (subclassing) by reasoners. Further, an OWL ontology can be queried using the SPARQL query language (Prud'hommeaux and Seaborne 2008). Uses for an ontology include:

- Description of terms, concepts, and relationships in a domain
- Organization of information
- Formation of a knowledge base
- Possibility to use ontology terms as standard terms
- Use of an ontology for enabling semantic interoperability

The last bullet describes situations in which the ontology is consulted when encountering diverse terms. That is, mappings can be done between ontology terms and local terms to resolve differences between local databases. Those mappings are then referred to during query processing, enabling local data to be queried in its local terms but yet be part of a combined answer. The initial vision for our project is to create such crosswalks or mappings. The ultimate vision for our project as well as the Semantic Web in general, however, is to create automatic or semi-automatic methods to resolve semantic heterogeneity using ontologies.

INTEROP GRANT

The initial people on the INTEROP project, in alphabetical order, are: Gary Berg-Cross, Knowledge Strategies; Mike Dean, BBN; Dave Kolas, BBN; John Moeller, formerly from Northrop Grumman Corporation, now at JJMoeller and Associates LLC; Nancy Wiegand, University of Wisconsin-Madison; James Wilson, James Madison University; Peter Yim, CIM Engineering, Inc.; and Naijun Zhou, University of Maryland, College Park. The management of the grant is being done at UW-Madison by N. Wiegand.

The tasks for the INTEROP grant include the following:

- Create a geospatial ontology repository
- Establish Web-based collaboration methods
- Conduct workshops/meetings, in-person or virtual
- Develop prototypes or demos
- Produce an educational component
- Conduct basic research in geospatial data interoperability

For a geospatial ontology repository, we are building on work already being developed by members of our project team for the Open Ontology Repository (OOR), http://oor-01.cim3.net/ontologies. The OOR community is developing a generic portal environment for storing, searching, and editing ontologies, along with associated tools related to working with ontologies, such as alignment tools. Currently, OOR has many ontologies, and we will be gathering and adding existing and new geospatial ontologies.

As to Web-based collaboration methods, we have a Wiki site at http://www.socop.org. This site is hosted by CIM3 as a Collaborative Work Environment (CWE) which allows linking to other pages. It also provides storage and access for documents and slide shows, and we have started putting some educational material here. The CWE also allows the hosting of virtual workshops and provides recordings of them. We will also have workshops as part of existing conference series or stand-alone or for the purposes of disseminating information about semantic technologies, gathering use cases, and working on ontologies.

We also will be developing prototypes or demos using various semantic representations and technologies to illustrate the use of ontologies and semantics for geospatial users. Further, we will be developing an educational component with in-depth material on geospatial ontologies and their use. We intend to collaborate with the University Consortium of GIScience (UCGIS) and their Body of Knowledge (BoK) project. In addition, we will pursue basic research in semantics for geospatial data interoperability.

We will also coordinate with existing standards groups, such as the Open Geospatial Consortium (OGC), to work on semantic issues for various geospatial domains. We intend to be an 'umbrella' over different groups or domains.

SUMMARY

This paper presented information on a new project that is working on semantic interoperability for geospatial data. The project furthers the development and use of technologies that are emerging as part of the overall Semantic Web vision. The focus in this project, however, is on the geospatial area.

We are in the process of broadening our Network to include interest and participation from the ASPRS community. We are also seeking additional use cases in an effort to consider and include multiple kinds of geospatial data for semantic needs.

ACKNOWLEDGMENTS

This work was supported by the INTEROP Program of the National Science Foundation, grant No. 0955816.

REFERENCES

Beckett, D. and B. McBride, 2004. RDF/XML Syntax Specification (Revised), W3C Recommendation 10 February 2004, http://www.w3.org/TR/rdf-syntax-grammar/.

Berners-Lee, T., J. Hendler, and O. Lassila, 2001. The Semantic Web. *Scientific American*, May, 29-37.

Hochmair, H., 2005. Ontology Matching for Spatial Data Retrieval from Internet Portals. *Proceedings of Geospatial Semantics.* Lecture Notes in Computer Science, 166-182.

McGuinness, D.L., F. van Harmelen, 2009. OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004, http://www.w3.org/TR/owl-features/, and OWL 2 Web Ontology Language Document Overview, W3C Recommendation 27 October 2009, http://www.w3.org/TR/owl2-overview/.

Prud'hommeaux, E. and A. Seaborne, 2008. SPARQL Query Language for RDF, W3C Recommendation 15 January 2008, http://www.w3.org/TR/rdf-sparql-query/.