Searching for Geospatial Government-Produced Data

Nancy Wiegand
University of Wisconsin-Madison
550 Babcock Drive
Madison, WI 53706
USA 608 263-5534
wiegand@cs.wisc.edu

ABSTRACT
This poster investigates problems and solutions for searching the Web for geospatial data produced by various levels of government. Currently, geospatial data to meet particular needs are difficult to find. We suggest using Internet DBMS and Semantic Web technology to improve search techniques over metadata by using context-oriented search phrases, full query expressions, and semantic mappings. Furthermore, automatic searching for data is possible using an Internet DBMS that ranges over Web published metadata and consults semantic mapping expressions of models and terms.

Categories and Subject Descriptors
H.2.5 [**Database Management**]: Heterogeneous Databases. I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods - *representation languages*.

General Terms
Management, Design, Human Factors, Standardization.

Keywords
Geospatial, Semantic and model heterogeneity, Internet technology, DBMS, Semantic Web.

1. INTRODUCTION
Finding government-produced geospatial data is problematic. Although large amounts of data are produced, potential users have a difficult time finding data sets they need. This is true despite the fact that many geospatial portals or clearinghouses are being developed by local, state, and federal government agencies [5]. Because of the difficulties of knowing what data are available and where they might be stored (i.e., which portal), many data seekers currently use a general Internet search engine to search for geospatial data. There is a need for better methods to effectively disseminate geospatial data. Furthermore, an ultimate goal would be to have a Web agent or service automatically locate needed data sets. This possibility is becoming technically feasible using new Web technologies.

Digital Government Conference'05, May 1–2, 2005, Atlanta, GA, USA.

Current geospatial Web portals, such as Geospatial One-Stop [1], require data source publishers to register with the portal by completing minimal metadata forms that are kept by a central authority and used to process user requests. A limitation of this method is that there are many geospatial clearinghouses and portals, and data producers may not register at all (or any) of them. Also, more detailed information describing the data sets, which may be found, for example, in associated FGDC metadata files, may not be available for search criteria.

Metadata is a core feature in searching for geospatial data sources and enabling interoperability. For example, metadata is crucial in the triumvirate involving the NSDI, the National Map, and Geospatial One-Stop [6]. FGDC metadata, in particular, is required for National Map data.

A portal for a public commons of geographic data has been proposed to help solve the problem of data not being easily available to potential users [4]. The designers of this approach recognize the difficulties of creating full metadata, which typically is only done by trained professionals, and they note there are other metadata formats in common use besides FGDC or its replacement ISO 19115, such as Dublin Core. They promote a menu drop down list for easy creation of minimal metadata. However, they note that for maximum use some data producers should provide comprehensive metadata.

A global vision of a spatial data infrastructure (GSDI) has also been put forth [3]. Among other technical specifications being developed to allow national and global sharing of information, *semantic registries* are proposed. That is, although FGDC metadata tags, for example, are standardized, the content for metadata fields may contain regional-specific or otherwise heterogeneous terms, which limits success in search requests. Also, interoperability standards from the Open Geospatial Consortium (OGC) do not specify data content and structure for nonspatial attributes. In [3], it is suggested that data models and semantic models are also published with data sources, along with, if possible, mappings to national schemas and definitions. OWL is suggested as a mechanism to create a Semantic Web for the geospatial community. In summary, the issues addressed in this paper include the problem of where to publish or find data, the usefulness of full descriptive metadata, the problem of semantic heterogeneity found in metadata content, and the need to alleviate manual searching for data.

2. DISCUSSION OF SOLUTIONS
Although another area of study would be to use data source content to help locate data sources, solutions discussed here focus on the use of metadata, where the metadata exists separately from the source data. Potential solutions involve more in-depth querying of metadata using DBMS query methods, the use of

semantic resolution technologies, and the automation of locating data sources by a Web agent.

2.1 Internet DBMSs

We first discuss the use of Internet DBMS technology to enable more advanced search capabilities. Assume here that geospatial metadata is published on the Web separately from the actual data sources so that it can be crawled and searched itself. In this scenario, however, current Internet search engines are limited to simple keyword searches. However, new technology in the form of Internet XML DBMSs, such as Niagara [2], allows *text-in-context* searching. For example, to find only complete FGDC data sets, rather than typing "complete" as a keyword, a constraint on the standard FGDC XML tags can be used such as 'metadata.idinfo.status.progress *contains* "complete"'. This returns the URLs of a limited set of FGDC metadata files.

Furthermore, an Internet DBMS provides full query expressions using an XML query language, allowing conjunctions and functions, for example. The following query in XML-QL ranges over metadata to find the URL for land use data for Dane County:

```
WHERE <$*>                                    (matches any pattern)
    <Theme_Keyword> "Land Use"          </Theme_Keyword>
    <Place_Keyword> "Dane County, WI" </Place_Keyword>
    <URL_Source>  $u  </URL_Source>
IN *                                   (ranges over whole data space)
CONSTRUCT $u
```

Use of full query power and full metadata, instead of minimal metadata, will also help searching within existing portals. For example, one could find the data source with the greatest accuracy among a set of data sources by finding the one with the minimum value for horizontal accuracy.

2.2 Semantic Resolution

Semantic resolution can be done by mapping between various forms of metadata and between content terms. For example. OWL mappings can compare the schemas of FGDC, ISO, Dublin Core, and user-created metadata. Then, a user query referring to the FGDC tag "Theme_Keyword" can be re-written using ISO or Dublin Core equivalents to be able to re-target or broaden the search over additional metadata formats.

OWL can also be used to map between heterogeneous data values used within metadata descriptions. Although a standard dictionary for geographic information does not currently exist, various efforts are underway which could result in a global ontology of terms [4]. Local terms could then be mapped to a global ontology.

An example OWL mapping is shown in Figure 1 in which "swamp" is mapped to the global resource term "wetland", but "bayou" is not considered an equivalent term We propose that additional semantic mapping types are needed to supplement OWL's current types.

The use of semantic registries and mappings would be valuable within existing geospatial portals [3]. Another architecture has been proposed that uses RDF, ontologies, and PostgreSQL to find geospatial data [5]. However, this paper uses forthcoming Internet DBMSs to suggest automatic searching for data as presented next.

```
<owl:Class rdf:ID="LouisianaTerms">
  <rdfs:comment>Mapping ontology to LA</rdfs:comment>
  <owl:oneOf rdf:parseType="Collection">
     <LouisianaTerms rdf:ID="swamp"/>
          <owl:sameAs rdf:resource="#Wetland"/>
     </ LouisianaTerms >
     < LouisianaTerms  rdf:ID="bayou"/>
          <owl:differentFrom rdf:resource="#Wetland"/>
     </ LouisianaTerms >
   </owl:oneOf>
</owl:Class>
```

Figure 1. OWL mapping between an ontology and local terms

2.3 Automatic Searching for Data

Given formal representations of mappings between models and between values, a Web application can be written to automatically locate data sources for a particular type of application such as land use planning or emergency response [7]. This method assumes metadata are published on the Web and that an Internet DBMS is used as a Web Service component. For example, similar types of data are needed for land use planning for each locality, enabling template queries to be created that just need the addition of geographic area or place. For example, if a user specifies Dane County, WI, a land use planning Web application can create queries to automatically locate the many different data sources needed. The Web application would use semantic lookups to range over many metadata formats and diverse content terms.

3. ACKNOWLEDGMENT

This work was partially supported by the Digital Government Program of NSF, Grant No. 091489.

4. REFERENCES

[1] Geospatial One-Stop, http://www.geodata.gov.

[2] Naughton, J., DeWitt, D., Maier, D., and others. The Niagara Internet query system. IEEE Data Engineering Bulletin, Vol. 24, No. 2, 2001, 27-33.

[3] Nebert, D. Technical baseline for accessing a virtual global SDI. Presented at GSDI-7, Jan. 29-Feb. 6, 2004, Bangalore, India, http://www.fgdc.gov/publications/papers/attloves.pdf.

[4] Onsrud, H., Camara, G., Campbell, J., and Chakravarthy, N. Public commons of geographic data: research and development challenges. In *Proceedings of the third international conference on Geographic Information Science* (Adelphi, MD, USA October 20-23, 2004). Springer, LNCS 3234, 223-238.

[5] Reitsma, F. Spatial data accessibility and the semantic Web. http://gest.umbc.edu/student_opp/2003_gssp_reports/FemkeReistsma.pdf.

[6] Shin, S. and Landis, E. Enabling data interoperability through metadata. Presented at NSF, May 23-27, 2004, http://www.fgdc.gov/publications/papers/attyszd.pdf.

[7] Wiegand, N. Querying metadata over the web to locate geospatial data. In *Proceedings of the third international conference on Geographic Information Science* (Adelphi, MD, USA October 20-23, 2004). NCGIA, 224-226.