A Web Query System for Heterogeneous Government Data

Nancy Wiegand and Naijun Zhou
University of Wisconsin
550 Babcock Drive
Madison, WI 53706
wiegand@cs.wisc.edu, nzhou@wisc.edu

Isabel F. Cruz and William Sunna
Department of Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607
[ifc,wsunna]@cs.uic.edu

Abstract

This paper describes a Web-based query system for semantically heterogeneous government-produced data. Geospatial Web-based information systems and portals are currently being developed by various levels of government along with the GIS community. Typically, these sites provide data discovery and download capabilities but do not include the ability to pose DBMS type queries.

One of the main problems in querying distributed government data sources is the difference in semantics used by various jurisdictions. We extend work in schema integration by focusing on resolving semantics at the value level in addition to the schema or attribute level. We illustrate our method using land use data, but the method works for any domain having a heterogeneous set of values. Our work starts from an XML Web-based DBMS and adds functionality to accommodate heterogeneous data between jurisdictions. Our ontology and query rewrite systems use mappings to enable querying across distributed heterogeneous data. This paper describes our current system.

1. Introduction

This paper presents an application of ontologies and database technology to query government-produced data over the Web. We give an overview of our system and address the problem of mediating between highly heterogeneous information developed independently by many different government units. In this project, we focus on geospatial data, in general, and on diverse land use coding systems, in particular. We extend work on semantic integration to the *value level* to be able to map between heterogeneous values in land use attributes.

Currently, Web sites are being developed by many levels of government to serve geospatial data. A new example from the Federal Government is Geospatial One-Stop [GOS03], which was initiated to have all data accessible from one portal. Prior examples of geospatial clearinghouses and portals at the national level include the Alexandria Digital Library [ADL] and sites available from the Federal Geographic Data Committee (FGDC) Web site [FGDCa]. State level examples include the Wisconsin Land Information System [WLISsite] and the Wisconsin Land Information Clearinghouse [WiscLinc]. In addition, there are many geospatial Web sites being produced at the county and local levels.

These Web sites, however, are limited or lacking in their ability to allow Database Management System (DBMS) type querying over the distributed government data sources. Querying in these sites is limited to a few metadata fields for users to locate data sets for download. Geospatial Web sites that do allow DBMS querying on source data provide an explicit connection over the Web to a backend DBMS containing the data. Otherwise, to query a data source, a user has to download it and put it into a local GIS or DBMS. Instead, we are working to enhance these sites by supporting full-fledged DBMS type querying over distributed Web geospatial data sources.

We began our work in the context of the statewide Wisconsin Land Information System (WLIS). However, the work can be applied to Geospatial One-Stop or other geospatial Web sites. WLIS, for example, will allow central access to distributed data sets that remain under local government control and reside on local and county servers across the State [WLIS99]. The GIS community was active in promoting the WLIS concept, and much of the data is geospatial. The problem is that, without state-mandated standards, data sets produced by local and county governments are highly heterogeneous. As one moves across jurisdictional boundaries, database schemas vary, and the definitions and acceptable values of attributes change significantly. In fact, most of the effort required to integrate diverse geospatial data lies with the nonspatial attributes and not with geometric conversions.

This paper discusses our prototype for a geospatial Web-based query system with a focus on resolving semantic heterogeneities between the attribute data of GIS data sets. Section 2 considers characteristics of geospatial data for designing a distributed Web query system. Section 3 discusses the semantic heterogeneity problem in general, and Section 4 presents our specific application problem of land use codes. Our existing working system is described Section 5 using land use data examples. Section 6 provides a summary.

2. Needs of a Geospatial Distributed Web Query System

In this section, we discuss basic data model characteristics needed for a Web-based distributed query system for geospatial data. Although any one characteristic is not necessarily unique for geospatial data, traditional DBMS system architectures do not accommodate these characteristics, and criteria for emerging Web systems are still an area of research. We found that a geospatial Web query system needs to recognize the following differences for geospatial data:

- Heterogeneity in data content
- Inherent data set organization, such as by theme, jurisdiction, and spatial extent
- Existence of metadata
- Spatial functionality

Heterogeneity in data sources is not a new problem (See Section 3). However, we are particularly concerned with semantic heterogeneity at the value level (Sections 3.3 and 4). For example, coding systems used to describe land use vary by jurisdiction but need to be integrated for comprehensive planning purposes involving multiple jurisdictions.

Geospatial data have various inherent data set organizations. These base organizations are often used in geospatial portals as search criteria. For example, GIS data sets tend to be based on a theme (also called a data category or channel) such as land use, transportation, hydrography, etc. In addition, data sets are produced by or describe government jurisdictions. Jurisdictions range from cities, towns, and villages to counties, regions of various types, the state, and the federal level. A spatial extent is inherent in geospatial data even without explicit coordinates. Geospatial are also characterized by type of data such as raster or vector and by temporal aspects.

Another feature of geospatial data is the existence of separate metadata files that describe the data source. These files provide more information than typical DBMS schemas and can be queried separately. A common format for metadata is the FGDC's Content Standard for Digital Geospatial Metadata (CSDGM), oftentimes referred to as the FGDC standard [FGDCb]. A Web DBMS for geospatial data needs to handle separate metadata files and their associations with the actual data. Challenges regarding metadata have been recognized by [MD99].

A last characteristic of GIS data is the explicit spatial extent, which may be represented using spatial coordinates at the data set level and/or the record level. There is a need for spatial indexes and functionality such as buffering and overlay. Ideally, a Web-based query system would also include spatial queries and other capabilities, but that is not the focus here.

3. Related Work on Heterogeneity

This section discusses selected prior work on heterogeneous data sources. We also discuss the level of semantics at which we are currently working, which is the value level.

3.1 GIS Interoperability

3.1.1 Semantic Interoperability

Types of interoperability relevant to GIS were delineated in [Bis98], and the difficulty of semantics was noted. Bishr's highest level of interoperability proposes seamless communication between remote GISs without prior knowledge of their semantics.

To help realize GIS interoperability, the OpenGIS Consortium [OGC] was formed in 1994. Later, OGC extended XML to develop the Geography Markup Language [GML] to further help achieve geospatial interoperability. GML uses a standard markup for spatial representations although variations exist. Topology has recently been included, but the initial approach is feature-based in which marked up elements for nonspatial properties are included with the spatial properties to describe each entity/feature.

However, in GML, the representation and meaning of the *nonspatial* elements do not follow a standard and may be described in any manner. Therefore, GML does not solve semantic problems for nonspatial properties (just as XML does not). In this regard, OGC recommends the formation of *information communities* that share a data dictionary and metadata schema.

Ontologies have been proposed as a solution for semantic integration [Fen01]. An ontology is a shared and machine-executable conceptual model in a specific domain of interest [BFM02]. Creating ontologies automatically is being explored, for example, in [Mal02]. Ontology driven GIS has been proposed for geographic information integration, especially between GIS and remote sensing systems [FEA02]. They recognize the difficulties of creating and imposing standards, which we also found regarding a statewide land information system. Instead, they assume communities will commit to common ontologies, which will be used to characterize and locate information sources.

An XML query system using ontologies was developed in [TW00]. In their system, query processing using their special similarity operator includes semantically similar terms obtained from a graph they developed using WordNet. Use of the similarity operator produces ranked results for the similar words. In our application, however, the land use code mappings we need cannot be found in a general collection such as WordNet. Also, we need to present the user with precise semantic nuance information instead of retrieval relevance rankings. As a result, we needed to develop a precise ontology mapping method.

3.1.2 Semantic Standards for Land-Related Data and GIS

An example of work being done on national standards related to land parcel data is the National Integrated Land System [NILS]. NILS is an effort between the BLM and the USDA Forest Service to manage cadastral records and land parcel information specifically for use in geographic information systems.

A new XML standard related to land information exchange for land development and transportation professionals is [LandXML]. LandXML is being used for civil engineering applications, for example.

However, until additional standards are developed or information communities are formed, many GIS data sets will continue to be independently developed and heterogeneous. Furthermore, legacy data sets may never be converted to a new standard.

3.2 Prior work in DBMS heterogeneity

Heterogeneity in Database Management Systems is not a new topic; see for example [BBE98]. Various types of heterogeneity have been identified. *Syntactic* heterogeneity can be solved by mapping between data models, such as between a relational model and an object-oriented model. *Schematic* heterogeneity is handled by schema integration and is important when there are semantically similar objects of interest. *Semantic* heterogeneity concerns discrepancies in the meaning, interpretation, and intended use of the same or related data [SL90].

Clio is a tool for managing schema integration [HMH01, MHH+01]. Because of the difficulty to fully automate semantic decisions, Clio combines user intervention with automated techniques. An interesting feature of Clio is that it provides actual data values for review to help with mapping decisions.

Schema integration to solve schematic heterogeneity has taken two general approaches: Global-As-View (GAV) and Local-As-View (LAV). In GAV, each entity in a global schema is associated with a view over the data sources. Contrary to this, in LAV, a source is defined as a view over the global schema. In our work, we follow the LAV approach by defining an ontology for a master schema, which is then mapped to local attributes and values. We use this approach because of the large number of diverse local data sources in our application and the need to remain extensible for additional data sets. Also, using this method, we can present the user with master terms to use for posing queries ranging over the many diverse sources.

3.3 DBMS Heterogeneity at the Value Level

Prior research on schema integration has recognized semantic problems at the value level. By the value level, we mean the values in the domain of an attribute. [BBE98] presents three representational differences for value level conflicts: differences in the expression (e.g., 4.0 vs. A), differences in units (e.g., miles vs. kilometers), and differences in precision (for example, cardinality differences, e.g., low, medium, high vs. a range with five choices). These types of representational differences can mostly be resolved using straightforward formulas or algorithms.

However, we found an additional type of value level heterogeneity in our application, especially with land use codes. Land use coding systems include categories such as Agriculture, Commercial, and Residential. Land use coding systems are often hierarchical with each category being subdivided into more specific land uses.

As above, there are differences in the expression of land use codes (e.g., Agriculture codes beginning with “A” vs. beginning with “9”). And, units and precision vary (e.g., hectares vs. acres and six subcategories for Agriculture vs. eleven). But, in addition, the meanings of codes between different local data sources cannot be directly compared. For example, a coding scheme for one jurisdiction in which the Commercial category is divided into “Commercial Sales” and “Commercial Services” cannot easily be compared to another code scheme divided into “Commercial Intensive” and “Commercial Nonintensive”. More examples are given in the next section.

4. Value Level Semantic Problems for Land Use Data

One of the most important themes in WLIS is land use data. Wisconsin, similar to other states, has passed *Smart Growth* legislation requiring local governments to develop land use plans. Of the many types of data sets needed to create a land use plan, such as transportation, environmental corridors, wetlands, etc., land use data is of prime importance.

Land use codes are oftentimes found in GIS data sets. The data may be parcel based in which each polygon forms a parcel or may be such that each polygon covers an area having a particular land use code. In either case, there is great variability in what other fields exist in a data set containing a land use field.

The largest problem, however, is that city and county governments develop their own coding systems for land use. Historically, a standard coding system was never imposed, and individual communities preferred to develop land use codes that more closely represented the particular land uses in their jurisdictional area. A multi-dimensional coding system called the Land Based Classification System (LBCS) [EN99] was developed by the American Planning Association to help provide a standard. However, to date, it has not been widely adopted.

Table 1 shows example land use codes in Wisconsin. The synonyms (*Lucode*, *Tag*, *Lu1*, and *Lu_4_4*) for the land use attribute are more easily resolved than determining whether the code descriptions share common definitions. Value level semantic resolution is needed because the descriptions are not exact matches; each description slightly varies from one another. For example, the 8110 code of the City of Madison makes no distinction between cropland and farm buildings, whereas the Dane County Regional Planning Commission (RPC) has a separate code for farm buildings. Eau Claire County's most specific code that would include cropland is at the general agriculture level which also includes various other subcategories.

Table 1. Heterogeneity in land use codes

Planning Authority	Attribute Identifier	Land Use Code	Description of Code
Dane County	Lucode	91	Cropland/Pasture
Racine County (SEWRPC)	Tag	811 815	Cropland Pasture and Other Ag
Eau Claire County	Lu1	AA	General Agriculture
City of Madison	Lu_4_4	8110	Farms

Comprehensive planning between neighboring communities requires mapping between coding systems. An example query needed as part of a land use decision may be to "*Find all lands with a particular land use over a watershed spanning several counties*."

One solution to resolving differences in meanings of land use codes is to only compare codes at a general level, such as commercial, residential, or agriculture and not try to resolve subcategories. Even at this general level, a system to automatically make comparisons between diverse coding systems is extremely valuable for comprehensive planning because, currently, efforts to resolve codes are done by hand. We developed a method, however, to keep comparisons between the meanings of subcategories for codes as precise as needed.

The problem of resolving land use codes is a significant one. Not only are the ramifications important within a state, they are important for planning efforts between states or even between countries, such as between the U.S. border and Mexico [GW00].

5. Method for an Ontology-Based Web Query System

Although we are illustrating our method using land use codes, we are not limited to that domain. Instead, the method can be generalized to capture semantic differences for any attribute and its values.

5.1 Internet XML DBMS and Ontology System Overview

Our goal of providing full DBMS querying over distributed geospatial data sources, such as found in Geospatial One-Stop or WLIS, can be partially met using an XML Internet query system such as Niagara [NDM+01]. Niagara's search and query engines allow querying distributed Web data marked up in XML.

Similar to HTML systems, Niagara's crawler searches the Web and indexes data, here XML data. But, contrary to an HTML engine, the Niagara search engine allows *text-in-context* searching in addition to keyword searching. That is, rather than just typing in a keyword (e.g., "cropland"), one can add context to limit the search by typing, for example, "landusecode *contains* cropland". Furthermore, the Niagara query engine processes full XML-QL [DFF+98] queries (Java version). Among other contributions, Niagara has an "IN*" capability which allows a query to range over the entire Web, eliminating the need to specify data sources in a query.

Although a system such as Niagara has great potential for Web-based querying, it also has limitations. A significant limitation is that it cannot handle the heterogeneity that exists among conceptually similar data sources that are represented differently. Furthermore, it is not designed to take advantage of the unique characteristics of geospatial data.

Figure 1 shows our system that incorporates semantic integration into the Niagara architecture. Our modifications include the metadata and ontology indexes, a query rewrite system in an ontology component, and enhanced output displays.

As described further in the following sections, we provide a custom GUI, and we modified the XML-QL query language to support querying through an ontology. Briefly, our custom GUI captures the user query along with appropriate metadata to locate data sources. Our ontology system locates data sources and their semantic agreement files to perform look-ups for query re-writing. We also enhance results with spatial displays and semantic information.

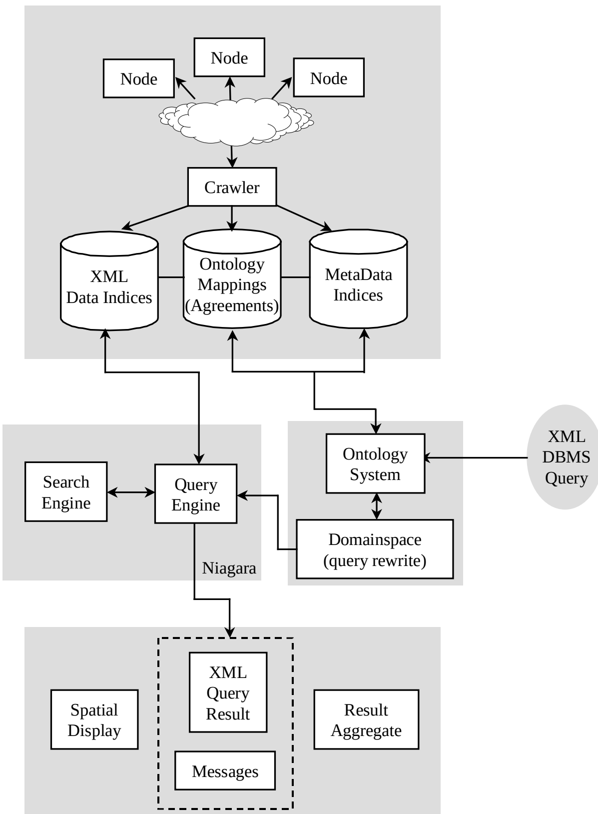

Figure 1. System Architecture

5.2 XML Data

We assume geospatial data will be marked up in XML. For our test data, we use a feature-based approach that includes the spatial coordinates for each XML feature. To do this, we converted each ArcView GIS [ESRI] land use data set to an XML file by adding XML tags to the nonspatial .dbf file. We combined the .dbf and .shp files by embedding the spatial coordinates for each feature. We also calculated and included the feature's bounding box extent. We currently use this information for our MapObjects [ESRI] spatial display of the results. Figure 2 shows sample XML data for a land parcel in Eau Claire County having a land use code of AR, farm residence.

```
<EauClaireLandUseData>
   <LandParcel>
      <area> 1704995.587470 </area>
      <perimeter> 5223.944820 </perimeter>
      <id> 258 </id>
      <lu1> AR </lu1>
      …
      <town> Wilson </town>
      <xycoords> 477229.236863, …</xycoords>
      <extent> 477229.236, , , 295553.507 </extent>
   </LandParcel>
   ….
</EauClaireLandUseData>
```

Figure 2. XML Source Data

5.3 Ontology Method

Given the semantic differences between the domains for land use coding systems, it is not possible to resolve discrepancies to the level of precision we need using fully automatic methods. For example, algorithms that involve matching either will not work or will not capture the needed semantics. That is, "cropland" will not match "agriculture" although cropland is a subset of agriculture. Other types of automatic matching methods involve rankings of results using probabilities. We had some success in automatically matching subcategories by modifying the Naïve Bayes classifier [Zho03], but conditions were limited. Furthermore, in an application involving land use planning, it is not sufficient or appropriate to return answers with attached probabilities. A land use planner needs to know as precisely as possible where and how much cropland exists over a multi-jurisdictional area, for example.

Therefore, we follow the approach of using ontologies as a solution for semantic integration. We pre-map a global ontology, that includes value-level information for heterogeneous domains, to local ontologies. Use of a global ontology is also beneficial for the user attempting to query heterogeneous data sets because the user can select from provided ontology terms.

Because many data sets on the Web, including land use data, are in a relational table format, we use a table-based ontology DTD (Figure 3) [CRS+02]. A unique feature of our approach is that the ontology DTD includes an "attrvalue" element to hold ontology domain values for heterogeneous domains. The attrvalue element can itself have nested attrvalue elements to describe any level of subcategories for codes.

```
<!ELEMENT database    (table+)>
<!ELEMENT table       (tuple+)>
<!ELEMENT tuple       (attrname+)>
<!ELEMENT attrname    (attrvalue*)>
<!ELEMENT attrvalue   (attrvalue*)>

<!ATTLIST database   id CDATA #REQUIRED>
<!ATTLIST table      id CDATA #REQUIRED>
<!ATTLIST tuple      id CDATA #REQUIRED>
<!ATTLIST attrname   id CDATA #REQUIRED>
<!ATTLIST attrvalue  id CDATA #REQUIRED>
```

Figure 3. Ontology DTD

We present a subset of the values for the global ontology for the land use code attribute (Figure 4). Another unique feature to our approach, in addition to providing a method to allow value level representations, is that the ontology does not specify a *standard* set of land use code values, which can be a contentious issue. Instead, it is a potential composite of all values that might be needed by the user in a query. We also allow multiple categorizations to be included. In this way, our solution solves the problem of a domain being categorized differently in various data sources. For example, in Figure 4, the Commercial code includes *both* the Scale and Function subcategories.

Allowing multiple categorizations in the global ontology also allows additional types of ontology categories to accommodate totally different types of mappings for special uses. For example, to study nonpoint water runoff, all local codes could also be mapped to two main global ontology categories, *Impermeable* and *Permeable*, in addition to typical codes such as Agriculture, Commercial, and so on.

```
<attrname id= "land_use_code">
 <attrvalue id= "Commercial">
  <attrvalue id= "Commercial-Scale>
    <attrvalue id= "Commercial-Scale-Intense"/>
    <attrvalue id= "Commercial-Scale-NonIntense"/>
  </attrvalue>
  <attrvalue id= "Commercial-Function>
    <attrvalue id= "Commercial-Function-Sales"/>
    < attrvalue id= "Commercial-Function-Service"/>
  </attrvalue>
      …
 <attrvalue id= "Other"/>
 </attrvalue>
 <attrvalue code= "Residential">
      …
 </attrvalue>
 …
</attrname>
```

Figure 4. Ontology for Land Use Code Values

5.4 XML Agreement Files

We developed a tool to map the global ontology to each local schema and the values of its land use coding system. The tool first performs semi-automatic alignment over ontology trees using a deduction algorithm [CR03, CSC04]. For the remaining mappings, a local domain expert chooses the cardinality of the mappings as Exact, Superset, Subset, Approximate, or No Mapping (Figure 5).

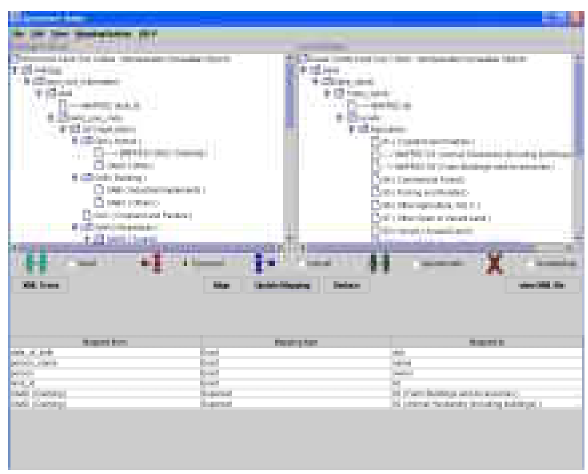

Figure 5. Tool to Create an Agreement File

The tool automatically generates an *agreement* file in XML for each local mapping. XML tags are used within the agreement file to record the semantics of the mappings (Figure 6). For example, the attribute value "OAAD" (Ontology Agriculture) maps using superset to two local values. The second example resolves the global Multi-Family code (ORMF) to the general Residential code when there is no mapping, i.e., the local coding system does not have a multi-family code. Information from the agreement files is used to generate subqueries sent into Niagara, as described in the next section.

```
<attrvalue id="OAAD" mapping="Superset">
        <localvalue>91</localvalue>
        <localvalue>92</localvalue>
</attrvalue>

<attrvalue id="ORMF" mapping="NoMapping">
         <resolve_to>Residential</resolve_to>
</attrvalue>
```

Figure 6. Fragment of an Agreement File

5.5 Class of Queries –GeoQueries

This section describes the class of queries we address and our query re-writing techniques. We give an example with a typical query in a land use application, "*Find all cropland in Dane, Racine, and Eau Claire counties*". This type of query is different from a typical DBMS query because more than one data source is identified, but there is no join. Instead, the same predicate is applied over multiple data sources. Because the data sources represent geographic areas, we call this type of query a *GeoQuery*.

We developed a new XML query language construct to accommodate a *GeoQuery*. We introduce a GeoSpace statement into the XML-QL query language (Figure 7). A GeoSpace can be considered to be an agent to record and generate queries specifically for each data source.

```
Query ::= Geospace WherePart ConstructPart
                   |    WherePart ConstructPart ;
Geospace ::= GEOSPACE id = SourceList;
SourceList ::= SourceList, Source | Source;
Source ::= string;
GEOSPACE ::= string;
 id ::= string;
```

Figure 7. BNF for a GeoSpace Statement

The GeoSpace statement (Figure 8) includes a variable, e.g., "*Area*", to hold the list of URLs for the data sources needed in the query. The variable is then used in the body of the query as a qualifier for generic ontology terms.

```
GEOSPACE   Area =“www.co.wi.us/Dane.xml,
                         www.co.wi.us/Racine.xml,
                         www.co.wi.us/EauClaire.xml”
WHERE <$*>
            <Area:LandUseCode> “agriculture” </>
            </>  ELEMENT_AS $a
CONSTRUCT $a
```

Figure 8. GeoSpace in an XML-QL Query

From the formal representation of the XML-QL query, subqueries are produced in local terms for each data source using our query rewrite system. That is, the formal query is rewritten for each URL in the GeoSpace list using native terms found in the agreement file for that source. Each resultant subquery is sent into Niagara. An example subquery pertaining to Dane County is shown in Figure 9. Because *cropland* does not map exactly to Dane County's *cropland-pasture* code, the user will receive this information in the results.

```
WHERE <$*>
            <Lucode> “91” </Lucode>
            </>  ELEMENT_AS $a
IN  www.co.dane.wi.us/Dane.xml
CONSTRUCT $a
```

Figure 9. A Generated Subquery

5.6 User Interface

Figure 10 shows the user interface to our system, which is modeled after the interface to Geospatial One-Stop. Compared to Geospatial One-Stop, we have an additional link that allows a user to *query* data sources (*Query Data Content*). When the user clicks on the link, he/she will then select a data category (theme) along with the geography over which the query will range. The latter can be done spatially or by selecting jurisdiction types and names. If the user selects the Land Use category, a land use ontology will be presented along with potential values for land use codes. The user composes a query by selecting ontology terms. The system then automatically processes the query over the various data sources corresponding to the chosen geography.

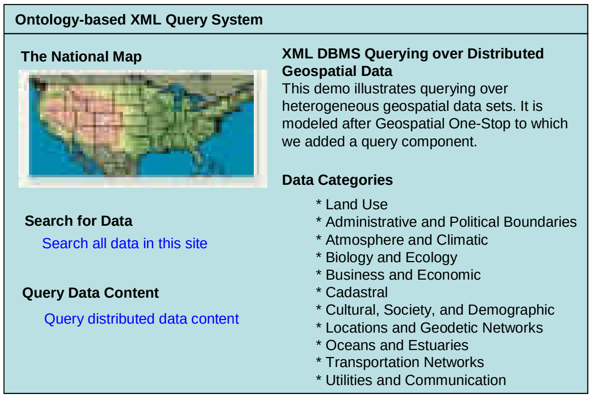

Figure 10. User Interface

5.7 Storage and Lookup Using Minimal Metadata

To be able to identify data sources involved in a query, we categorize data sets using minimal metadata consisting of theme, jurisdiction type, jurisdiction name, and spatial extent. Our interface is designed to obtain this information from the user so our integration system can identify the data sets relevant for a query. We create metadata for data sets not already having associated FGDC metadata files and store the information in metadata indexes.

5.8 Spatial Display and Semantic Output Messages

We post-process Niagara's results to group result elements by data source. We also perform client-side aggregations to display statistics on the results, and we use MapObjects [ESRI] for spatial displays.

Also, it is very important that a semantic query system inform the user of the semantic mapping resolutions used to generate subqueries. For example, if a user wanted "Multi-family" lands from a jurisdiction in which "Multi-family" mapping was resolved to the general "Residential" code, the user should be informed that the output includes all residential lands, i.e., a superset of information. Our output includes the mapping type and the local codes to which the ontology code was mapped. A short description of our demo is given in [WZC+04].

6. Summary and Conclusions

Data produced by local governments, in particular, tend to be highly heterogeneous. However, government operations, such as land use planning that affect a multi-jurisdictional area, require data integration.

Many levels of government are producing geospatial Web sites, such as the Federal Geospatial One-Stop and the Wisconsin Land Information System. Our goal is to provide query support to these sites. To solve the problem of semantic heterogeneity among independently developed government data sources, we developed an ontology and query rewrite system on top of an XML Web DBMS.

We extended work in schema integration by focusing our efforts on resolving differences at the value level. This was needed to be able to query across the many different land use coding systems in use in Wisconsin. However, our method is applicable to heterogeneous domains of any attribute.

We developed a tool to aid local experts in mapping an ontology to their local schemas. The tool automatically generates an XML agreement file in which semantic relationships are stored using XML tags. To formally represent a typical type of query in our application, in which the same predicate is applied over multiple data sets, we extended XML-QL with a GeoSpace statement.

7. Acknowledgement

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

8. References

[ADL] Alexandria Digital Library project, http://webclient.alexandria.ucsb.edu/.

[Bis98] Y. Bishr, "Overcoming the Semantic and Other Barriers to GIS Interoperability", *International Journal of Geographical Information Science*, Vol. 12, No. 4, 1998, pp. 299-314.

[BBE98] A. Bouguettaya, B. Benatallah, and A. Elmagarmid, *Interconnecting Heterogeneous Information Systems*, Kluwer Academic Publishers, 1998.

[BFM02] Bussler, C.; Fensel, D.; and Maedche, A. 2002. "A Conceptual Architecture for Semantic Web Enabled

Web Services", SIGMOD Record, Vol. 31, No. 4, pp. 24-29.

[CR03] Cruz, I.F. and Rajendran, A. 2003. "Exploring a New Approach to the Alignment of Ontologies", Semantic Web Technologies for Searching and Retrieving Scientific Data Workshop, International Semantic Web Conference, Sanibel Island, Florida, pp. 7-12.

[CRS+02] I.F. Cruz, A. Rajendran, W. Sunna, and N. Wiegand, 2002. "Handling Semantic Heterogeneities Using Declarative Agreements", In Proceedings of ACM GIS, pp. 168-174.

[CSC04] I.F. Cruz, W. Sunna, and A. Chaudhry, 2004. "Ontology Alignment for Real-World Applications", In Proceedings of The National Conference on Digital Government Research, dg.o2004.

[DFF+98] A. Deutsch, M. Fernandez, D. Florescu, A. Levy, and D. Suciu, "XML-QL: A Query Language for XML", 1998, http://www.w3.org/TR/NOTE-xml-ql/.

[ESRI] Environmental Systems Research Institute (ESRI), http://www.esri.com.

[EN99] J. Everett and C. Ngo, "Land-Based Classification Standards—Federal Role." In Proceedings of the American Planning Association National Planning Conference, 1999, www.asu.edu/caed/proceedings99/LBCS/EVERETT.HTM

[FGDCa] Federal Geographic Data Committee (FGDC), http://www.fgdc.gov/clearinghouse/clearinghouse.html.

[FGDCb] Federal Geographic Data Committee (FGDC), CSDGM Metadata, www.fgdc.gov/metadata/metadata.html.

[Fen01] D. Fensel, *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag, Berlin, 2001.

[FEA02] F. Fonseca, M. Egenhofer, P. Agouris, and G. Camara, "Using Ontologies for Integrated Geographic Information Systems", Transactions in GIS, 6(3), 2002.

[GW00] P. Ganster and R. Wright, Eds. "San Diego-Tijuana International Border Area Planning Atlas", Institute for Regional Studies of the Californias, San Diego State University, 2000, http://www-rohan.sdsu.edu/~irsc/atlas/atlashom.html.

[GML] Geography Markup Language (GML) 2.0, OpenGIS® Implementation Specification, 20 February 2001, OGC Document Number: 01-029, http://opengis.net/gml/01-029/GML2.html.

[GOS03] Geospatial One-Stop, http://www.geodata.gov and http://www.geo-one-stop.gov, 2003.

[HMH01] M. Hernandez, R. Miller, and L. Haas, "Clio: A Semi-automatic Tool for Schema Mapping", Demo, In Proceedings of the 2001 ACM SIGMOD International Conference of the Management of Data, Santa Barbara, CA, May, 2001, p. 607.

[LandXML] LandXML.org, http://www.landxml.org/.

[MD99] D. Maier and L. Delcambre, "Superimposed Information for the Internet", In ACM SIGMOD Workshop on The Web and Databases (WebDB '99), Philadelphia, 1999, pp.1-9.

[Mal02] R. Malyankar, "Vocabulary Development for Markup Languages – A Case Study with Maritime Information", WWW2002, Honolulu, Hawaii, May 2002.

[MHH+01] R. Miller, M. Hernandez, L. Haas, L. Yan, C. Ho, R. Fagin, and L. Popa, "The Clio Project: Managing Heterogeneity", SIGMOD Record, Vol. 30, No. 1, March 2001, pp. 78-83.

[NILS] National Integrated Land System (NILS), http://www.blm.gov/nils/.

[NDM+01] J. Naughton, D. DeWitt, D. Maier, and others. "The Niagara Internet Query System", IEEE Data Engineering Bulletin, Vol. 24, No. 2, 2001, pp. 27-33.

[OGC] OpenGIS Consortium, Inc. (OGC), http://www.opengis.org.

[SL90] A.P. Sheth and J.A. Larson, "Federated Database Systems and Managing Distributed, Heterogeneous, and Autonomous Databases, ACM Computing Surveys, 1990, 22(3):183-226.

[TW00] A. Theobald and G. Weikum, "The Index-based XXL Search Engine for Querying XML Data with Relevance Ranking, in Proceedings of EDBT 2000, Jensen, C.S. (ed.), 2000, pp. 477-495.

[WZC+02] N. Wiegand, N. Zhou, I.F. Cruz, and A. Rajendran, "Querying Heterogeneous GIS Land Use Data Over the Web", In Proceedings of GIScience 2002 Abstracts, M. Egenhofer and D. Mark (Eds.), Boulder CO, September 2002, pp. 207-210.

[WebView] Wisconsin Department of Natural Resources (DNR) WebView, http://maps.dnr.state.wi.us/dnrwebview.

[WLIS99] Wisconsin Land Council Technical Working Group, Wisconsin Land Information System Technical Report, 1999.

[WLISsite] Wisconsin Land Information System Core Node Pilot Project, http://wlis.dnr.state.wi.us/wlis.

[WiscLinc] Wisconsin Land Information Clearinghouse, http://wisclinc.state.wi.us.

[WZC+04] N. Wiegand, N. Zhou, I.F. Cruz, and W. Sunna, 2004. "Ontology-Based Geospatial XML Query System", In Proceedings of The National Conference on Digital Government Research, dg.o2004.