A Semantically-Enabled Provenance-Aware Water Quality Portal

Jin Guang Zheng[1], Ping Wang[1], Evan W. Patton[1], Timothy Lebo[1], Joanne S. Luciano[1] Deborah L. McGuinness[1]
[1] Tetherless World Constellation, Rensselaer Polytechnic Institute
{zhengj3, wangp5, pattoe, lebot, jluciano, dlm}@cs.rpi.edu

***Abstract*— Environmental informatics applications often analyze data collected from various sources. Both data collection and data analysis benefit from expert knowledge. However, if applications are to be used by a broader range of users with less expert knowledge, applications will need to include a deeper understanding of the data used and analysis performed. We present the *Tetherless World Constellation Semantic Water Quality Portal* as both a water quality portal application and as an example of a semantic approach to environmental informatics applications. The portal integrates water data from multiple sources and captures the semantics of the data in a simple water quality ontology. Portal users can identify polluted water sources and polluting facilities according to multiple regulation perspectives and geographic constraints by using visualizations of semantically-enabled queries. Further, knowledge provenance is encoded in the data capture and integration services to enable provenance-based queries and reasoning capability.**

Keywords—Semantic Web, Visualization, Semantic Environmental Informatics, Water Quality Portal

I. Introduction

Water quality has been a major concern for environmental scientists and local citizens who understand the important role that clean water plays in our lives and the health of our planet. Polluted water sources, the kinds of pollutants, and those responsible for the pollution need to be discovered so that corrective and preventative measures can be undertaken. To monitor and control water quality, government agencies [1,2] regularly collect water quality data about pollutants and establish regulations to define pollution in terms of acceptable levels of pollutants. With the amount of data collected, it can be complex and time consuming for trained professionals and citizens to discover polluted water sources. Beyond discovery, citizens and professionals may also want to gain an overall insight into a region's water sources by viewing trends of pollutant levels for many pollutants and many source locations.

Motivating Example: Children in the same neighborhood start to get sick. Their primary symptom is diarrhea and some parents suspect that impure drinking water is causing the problem. They contact authorities and request that the water sources be tested. The authorities collect samples, run tests, and log results. In addition to the data collected, the authorities get regulation data from other reputable agencies, such as the Environmental Protection Agency (EPA), U.S. Geological Survey (USGS), and state agencies. Data include recent sample analysis from routine monitoring operations that record where and when pollutant levels violate appropriate guidelines. Once the analysis is performed, authorities notify the community of the results and recommend actions to make. Specialized domain knowledge is required to collect these data and perform analyses to identify pollutants and polluted water sources.

In 2009, a small county in Rhode Island experienced such a scenario[1,2]. According to citizens, the dissemination of information relating to potential health hazards was too slow. This real life scenario motivated us to develop the *Tetherless World Constellation Semantic Water Quality Portal* (TWC-SWQP). SWQP is a portal enabled by semantic web technologies and can be used to identify polluted water sources, pollutants, and possible sources of pollution. In the remainder of this paper, we describe our design and implementation, highlight the benefits of our semantic approach, and discuss the potential impact of this approach for water quality informatics applications and other similar informatics needs.

II. Methods

A. SWQP System Architecture and Components

The system architecture of the TWC SWQP is illustrated in Fig. 1. The system comprises five major components: (a) ontology, (b) data conversion, (c) reasoning, (d) visualization and (e) provenance.

Ontology Component: There are two types of ontologies in the SWQP: the core ontology and the regulation ontology. The core TWC Water ontology[3] consists of 18 classes, 4 object properties, and 10 data properties. It extends existing best practice ontologies, including SWEET [3] and OWL-Time [4]. The core ontology models domain objects (e.g. water sources, facilities, measurements, and pollutants) as classes, and includes terms for relevant pollution concepts. For example, a polluted water source is modeled as the intersection of water source and something that has a pollutant measurement outside of an allowable a range. The application can use the core ontology to conclude "any water source that has a measurement outside of its allowable range" is a polluted water source. Further, it can discover pollution with respect to any particular pollutant such as arsenic. Subsequently, we can identify a polluted water source with respect to a particular pollutant,

[1] http://www.epa.gov/

[2] http://www.usgs.gov/

[3] http://purl.org/twc/ontology/swqp/core

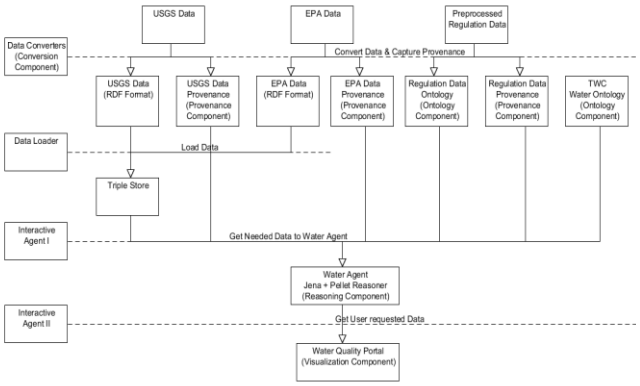

Figure 1. SWQP System Architecture and Workflow

given an existing constraint. For example, the portal can identify water sources that are polluted with arsenic, given the rule that arsenic concentrations value greater than 0.01 mg/L are poisonous.

The regulation ontologies[4] model the federal and state water quality regulations for different regions. For example, in California, the state regulation defines 0.01 mg/L as the limit for Arsenic. Because regions differ in their ecology and each state is responsible for its own regulations, the number of pollution concepts (pollutants and limits) and properties vary.

Portions of the core TWC Water ontology and Regulation Ontologies are illustrated in Fig. 2.1 and Fig. 2.2.

Data Conversion Component: We use two software tools to convert data into Resource Description Framework (RDF) [5] representations: the open source tool csv2rdf4lod[5] and an ad-hoc converter we developed for SWQP. The general-purpose csv2rdf4lod tool converts tabular data into well-structured RDF according to declarative parameters encoded in RDF [6]. To convert SWQP data, we wrote several conversion parameters to

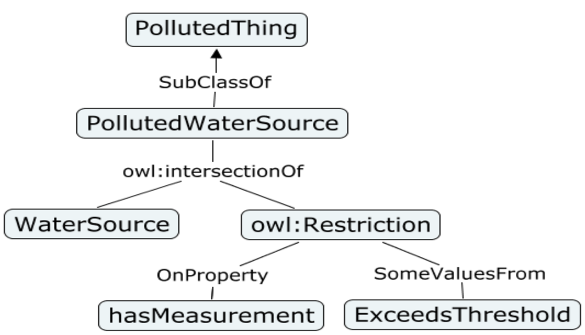

Figure 2.1 Portion of the TWC Water Ontology.

[4] e.g., http://purl.org/twc/ontology/swqp/region/ny and http://purl.org/twc/ontology/swqp/region/ri; others are listed at http://purl.org/twc/ontology/swqp/region/

[5] http://purl.org/twc/id/software/csv2rdf4lod

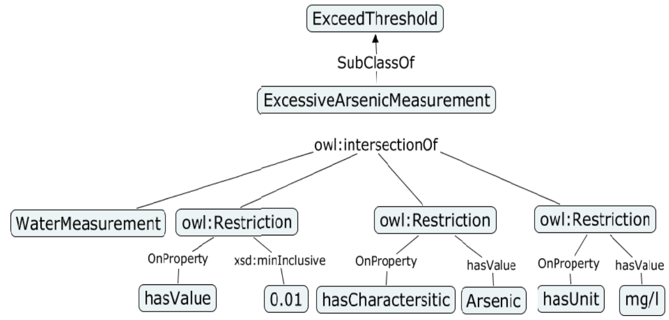

Figure 2.2. Portion of EPA Regulation Ontology.

map properties of the raw data to those in our ontologies. For example, we map the column heading “CharacteristicName” to the TWC water ontology property `hasCharacteristic` so that when data are converted to RDF, all values under the column “CharacteristicName” are associated using the `hasCharacteristic` property. One advantage of using the csv2rdf4lod tool is the provenance it captures as we convert the data, which we discuss below.

To construct OWL 2 [7] constraints that align with rules and properties in our TWC water ontology, we wrote ad hoc converters to extract regulation data from HTML web pages. Data and ontologies supporting the SWQP were loaded into OpenLink Virtuoso 6 open source community edition, which provides a SPARQL [8] endpoint[6].

Reasoning Component: We utilize the Pellet OWL Reasoner [9] together with the Jena Semantic Web Framework [10] to reason over the data and ontologies in order to identify polluting facilities and polluted water sources. Using the core ontology, we model water quality determinations such as “any water source that has a measurement that exceeds a regulation threshold, is to be considered a polluted water source”; using the regulation ontology, we model regulation criteria data, which are region-specific, e.g. California water regulation stipulates: “the threshold for Arsenic is 0.01 mg/L”. Combining the above two statements, the reasoning component asserts that any water source that has a concentrattion of aresenic greater than 0.01 mg/L is a polluted water source. At run time, the reasoning component invokes Jena to load the water quality data, the regulation ontology, and the core ontology. Then, Pellet is invoked to classify water sources as polluted or unpolluted from measurements from water samples and their water sources using the regulations as the criteria. The results of this operation can then be queried and visualized.

Visualization Component: This component is responsible for mashing up and representing the data collected from various sources. We support two types of visualizations: (1) map visualization that displays the sources of the water pollution in the context of geographic regions and (2) trend

[6] http://sparql.tw.rpi.edu/virtuoso/sparql

Figure 3. Map Visualization. The results of applying the EPA federal water regulations on the region with zip code 02888 is visualized on a Google Map.

visualization that depicts pollution levels over time with respect to a particular water source or facility:

- Map visualization: This component gets the reasoning results for a user query from the back-end reasoner and displays the results on a Google Map. We use different icons to distinguish between clean and polluted water sources, and between clean and polluting facilities. Fig. 3 shows an example map visualization. The user may select the data sources to be queried, the regulations to apply, or the types of water sites and pollutants he or she is interested in. The results of applying the EPA federal water regulation on the region with the zip code 02888 (Warwick, RI) is visualized in this example. Two polluted water sources and eight polluting facilities are indicated with icons. SWQP provides additional information in the pop up window by including the names of pollutants, the measured values, the limit values, and the water measurement time.
- Trend visualization: This visualization retrieves water quality data related to a selected water site or facility by querying the triple store and displays the water quality data as a time series using Protovis visualization toolkit. Fig. 4 shows the phosphorus measurements from 2007 to 2010 in green and the regulation defined limit in blue. Note that the data show one violation in 2009 (in red) and no subsequent violations.

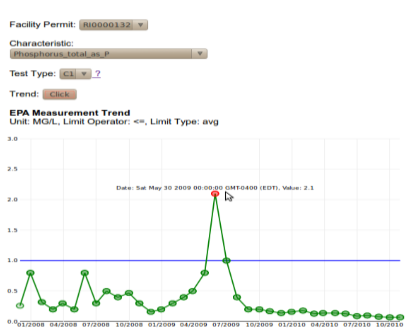

Figure 4. Trend Visualization. The phosphorus measurements from 2007 to 2010 and the regulation defined limit for the selected facility are visualized.

Provenance Component: SWQP supports two levels of provenance:

- Data source level provenance: When data are retrieved and converted to RDF, we store provenance information these processes using the Proof Markup Language (PML) [11]. We capture the data sources (i.e., the url of the original data files), the user retrieving the files, and when it was retrieved. These metadata are used to support provenance-based data queries, when the user selects the data sources he/she trusts by selecting the values under the Data Source facet in the map visulization (see Fig. 3). The portal will then use only data from the selected sources.
- Application level provenance: Besides provenance information about data sources, we also capture provenance information during the data conversion, loading, and reasoning steps and offer the intermediate results produced via the web. SWQP uses these information to explain to the user why a water source is marked as polluted or a facility is marked as a polluting facility. To access these explanations, the user can simply select a polluted water source in the map visualization (see Fig. 3). Along with these explanations, supporting provenance information for these explanations can also be accessed by clicking on the question marks in the pop up window.

B. *Source Data*

The data sources incorporated into SWQP span several government agencies, including the EPA and USGS, and federal and state regulation agencies.

EPA Data: We obtain permit compliance and enforcement status of facilities regulated by the National Pollutant Discharge Elimination System (NPDES) under the Clean Water Act (CWA) [7] from ICIS-NPDES [8], an EPA system. The compliance and enforcement status of facilities contains measurements of pollutants in the water discharged by the facilities, and also the threshold values for up to five test types for each pollutant. Three test types (C1, C2, C3) use concentration-based limits, while the other two (Q1, Q2) use quantity-based or mass-based limits.

USGS Data: We also fetch the National Water Information System[9] (NWIS) water quality data provided by USGS. The NWIS water quality data provides measurements of substances contained in water samples collected at USGS data-collection stations.

Regulation Data: The water portal makes use of water regulations, which are lists of pollutants and their maximum contaminant level[10] (MCLs). The national level drinking water

[7] http://www.epa.gov/agriculture/lcwa.html

[8] http://www.epa-echo.gov/echo/compliance_report_water_icp.html

[9] http://waterdata.usgs.gov/nwis

[10] http://water.epa.gov/drink/contaminants/

regulations from EPA, and the state drinking water regulations for California, Massachusetts, New York, and Rhode Island have been encoded and incorporated into SWQP.

C. System Workflow

We now present how the components described in the previous sections work together to identify polluted water sources, polluting facilities, and pollutants. Fig. 1 also shows the system workflow. SWQP first downloads data from USGS, EPA, and state regulation agencies for conversion into RDF using the Data Conversion component. During the conversion process, data level provenance information for the downloaded and converted data is captured. Next, SWQP loads the converted data into a triple store. When the user accesses the front-end interface of SWQP and issues a request, the request is sent to the back-end reasoning component. The reasoning component then loads the TWC Water Ontology, appropriate regulation ontologies (Ontology component), appropriate water or facility data and performs analysis. After the reasoning component completes its analysis, the results are sent to the visualization component for user presentation.

III. TECHNIQUES RESULTS

In this section, we discusses how semantic web technologies can serve as useful technologies for solving problems in the domain of water quality investigation from the following aspects: semantic data integration, semantic reasoning, and provenance support.

A. Semantic Data Integration helps SWQP integrate data from various sources, and eases future data integration.

SWQP integrates data from various sources, including EPA, USGS, and state governments. Researchers who need to analyze such heterogeneous data face two major challenges. First, raw data from multiple sources are stored in different formats, e.g. CSV, HTML, TXT, which makes it difficult to integrate and query the data. In addition, the semantics of the raw data are often not machine-accessible, i.e. they cannot be handled by a computer program. Furthermore, the semantics of the water quality data are not explicitly encoded in the data files, but are instead described in help pages on web sites, although not in a machine-understandable format.

SWQP uses a semantic approach to address these problems. We designed ontologies to model the domain of water quality investigation and explicitly encode the semantics of the data. Then, data from different sources were converted into RDF triples compatible with the ontologies. In this way, we achieved consistent and machine accessible semantics for the converted data. In our case, we load the data into a triple store and retrieve data required by users' queries with SPARQL

Another benefit of semantic data integration is ease of future extensibility. If one wanted to import more heterogeneous data using other technologies, it would be difficult to describe and store the data because schemas are typically difficult to alter once implemented. With an ontology-based approach, extensions can be made with ontology expansions such as adding equivalencies or new properties and classes. Similarly key information, such as regulation limits, becomes clear and easily accessible. For example, once the data was encoded in the regulation ontology, we could easily generate a comparison[11] of the federal and state limits for different pollutants.

B. Reasoning supported by semantic technologies enables SWQP to perform automatic analysis on water quality.

Not only do environmental scientists collect the water data from various sources, they also conduct all sorts of analysis over the collected data. Such analysis tasks are often time consuming, since data can be large due to large spatial regions or long time spans. Furthermore, some of the analysis tasks can be complex. For example, to identify if a water source is polluted, we need to compare all measurements of all pollutants with their corresponding limits in the adopted water regulations.

Aiming at saving environmental scientists' time and effort, we utilize semantic technologies to automate the water data analysis. Using the ontologies and reasoning components, SWQP is able to automatically identify polluting facilities and polluted water sources, as well as the corresponding water measurements that violate the regulations. This is done through semantic reasoning provided by the Pellet reasoner.

Often, scientists who are interested in environmental studies are not experts in writing complex queries to query the data they need. For example, to query polluted water sources without reasoning, we need to write complex queries as shown in (1), which compares all measurements from a water source against all regulations. However, with automatically reasoned results, scientists can simply query polluted water sources and their related information as shown in (2).

```
SELECT * WHERE {
  ?watersource   twcwater:hasMeasurement ?measurement.
  ?measurement twcwater:hasValue            ?value;
                 twcwater:hasCharacteristic ?charactericsitc;
                 twcwater:hasUnit            ?unit.                    (1)
  ?regulation    twcwater:hasValue            ?limit;
                 twcwater:hasCharacteristic  ?characteristic;
                 twcwater:hasUnit             ?unit.
  ?watersource geo:lat    ?lat;   geo:long ?long.
  FILTER( ?value > limit )
}
SELECT * WHERE {
  ?watersource rdf:type twcwater:pollutedWaterSource.
                geo:lat    ?lat;                                       (2)
                geo:long ?long.
}
```

C. Provenance information encoded using semantic web technology supports transparency and trust.

The primary purpose of SWQP is to discover polluted water sources and polluting facilities in areas a user finds interesting. However, SWQP responses may not be trusted by some users if there is no mechanism that provides the option to examine how the responses are obtained. As pointed out in [12], knowledge provenance, which includes source identification, source authoritativeness, and a supporting graph, can be used to provide explanations. These "explanations" help users

[11] http://tw2.tw.rpi.edu/zhengj3/water/reg_comp.html

understand where responses come from, and what they depend on, thus allowing users to determine for themselves whether they trust the responses they received.

SWQP supports provenance at two levels. Data source provenance supports water quality and regulation source queries. Application level provenance allows the user to examine all the manipulations the data go through.

Our portal also demonstrates that provenance can also be used to enable additional types of questions. Users can apply any regulations they want to investigate possible pollution. For example, one might choose a “what if” scenario, for example, to apply a stricter regulation (e.g. from another state) to a local water source. If for example, Rhode Island regulations are applied to water quality data for zip code 02888, 13 polluted water sites are identified. When California regulations are applied, 16 polluted water sites are identified (shown in Fig. 5). Using California criteria on this same region, the indicated number of polluted water sites increases by 23% compared to the number indicated using RI regulation criteria. If we compare the results of using California criteria with using federal regulations, the number of polluted sites grows by 700%.

SWQP brings together seemingly disparate regulatory and measurement data from multiple sources and, through automated classification and visualization, it can present the data to non-expert users. It provides basic tools to enable users to evaluate exploratory hypotheses. The availability and integration of data are critical to the portal’s ability to rapidly disseminate information to the public. With tools such as SWQP, the public could review historical water quality data quickly. Further, citizen scientists could provide their own sample collection and testing data along with its provenance. Although citizen-scientist findings may not be as reliable as experts’, they may be timelier and lead authorities to more appropriate testing and validation.

IV. DISCUSSION

Environmental informatics research often benefits from domain knowledge. For example, water quality research requires domain knowledge concerning pollutants, thresholds for pollution, and pollutant test options. Applications that aim to integrate and disseminate water quality data to support analyses related to pollution need to capture and interpret domain knowledge such as sufficient conditions for determining water pollution states and events. Our work is the first we know of that uses a semantic approach to a provenance-aware water quality portal. Other works focus on facilitating water quality management [13, 14] and wastewater treatment [15] via knowledge sharing and reuse. Chen [13] proposed a prototype system that integrates water quality data from multiple sources and retrieves data using semantic relationships among data. Chau [14] presented an ontology-based KM system (KMS) that can be integrated into the numerical flow and water quality modeling to provide assistance on the selection of a model and its pertinent parameters. OntoWEDSS [15] is an environmental decision-support system for wastewater management, which augments classic rule-based and case-based reasoning with a domain ontology. SWQP differs from these projects in that it supports provenance based query and data visualization. Moreover, SWQP is built upon standard semantic technologies (e.g. OWL, SPARQL, Pellet, Virtuoso) and thus can be easily replicated or expanded.

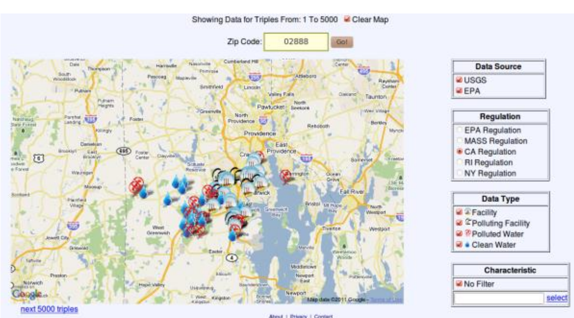

Figure 5. Applying California regulation data to RI water quality Data

SWQP can be expanded in several ways. 1) We can expand SWQP to support all 50 US states. Water quality data can be obtained from EPA and USGS websites. Then, SWQP can identify water pollutions in all the states according to the federal water regulation (or other state regulations we have already encoded such as CA and RI). It is similarly not difficult to obtain the remaining state regulations using either our existing ad-hoc converters or potentially new converters if the data is in different forms. 2) We can quickly add interesting applications to SWQP by integrating data from other sources, e.g. weather and flood forecasts. Flood conditions can exacerbate pollution impacts when pollutant control strategies fail due to floods or when a polluted water source is mingled with a non-polluted water source. If weather conditions suggest anticipated flood regions, SWQP can identify polluting facilities near the flood zone and potentially identify risks and suggest compensating strategies. Another direction is to model the health effects from exposure to the excessive pollutants in water and support reasoning over these effects. Then, SWQP can provide queries customized to health concerns. If the user inputs that he/she is concerned with water pollutants that negatively impact kidneys, SWQP can highlight water sources with high levels of cadmium given the rule that long-term exposure to excessive cadmium may cause kidney damage. 3) The architecture of SWQP can be used for other environment topics. We can build semantic web portals for investigating air quality, soil quality, etc. using the same architecture and workflow used in SWQP. For example, the TWC Clean Air Status and Trends demo [12] has gone through an update to include provenance and could be expanded to include the regulation views.

As the portal is expanded for greater usage, its credibility becomes more important. To increase the credibility of the portal, we plan to augment its provenance support by building, linking and displaying proof traces that track how the answers are derived from source data. Our PML and Inference Web provenance infrastructure [16] makes it easy to encode all the data manipulations and use that information for presenting either a complete trace or abstracted trace for user inspection.

[12] http://logd.tw.rpi.edu/demo/clean_air_status_and_trends_-_ozone

We also would like to support provenance granularity options so that users can choose the granularity of the provenance they prefer in certain contexts.

Several e-Science systems have incorporated similar types of provenance support. myGrid [17] proposes the COHSE open hypermedia system, which generates, annotates and links provenance data to build a web of provenance documents, data, services and workflows for biological experiments. The Multi-Scale Chemical Science (CMCS) [18] project developed a general-purpose infrastructure for collaboration across many disciplines. It also contains a provenance subsystem for tracking, viewing and using data provenance. In future work, we intend to leverage the best of these approaches along with domain-specific provenance needs and our Inference Web (IW) [16] provenance infrastructure to provide a more water quality-oriented provenance-aware application. We believe that the IW focus on supporting extraction, maintenance and usage of provenance of answers given by web application and services along with the workflow focus of the other systems will provide a nice complement to this work.

Our SWQP evaluation is currently experiential. We demonstrate capabilities that have previously not been possible or not done as efficiently in other architectures. Additionally, we are not aware of a best practice evaluation benchmark for interdisciplinary environmental informatics portals such as this. It is also difficult to evaluate the ontology against existing related ontologies because the driving use case is different and thus the ontologies have significant differences. However, through the design and implementation of SWQP, we have demonstrated the value and potential of applying semantic technologies to facilitate environmental research and community awareness. In the future, we would like to engage both researchers from the hydrology community and interested citizens to evaluate the portal. Feedback from the two user groups can lead to improvements of the portal.

V. Conclusion

We presented the TWC Semantic Water Quality Portal, which allows user to discover polluted water sources and polluting facilities. We have illustrated benefits of applying semantic web technologies to water quality research. These benefits include support for provenance-aware pollution retrievals, automatic support for identifying overall pollution or pollutant-specific pollutions, semantically-informed data visualizations showing pollution events, and trends over time. We also discussed the extensibility of the portal and the potential for using it for topics beyond water quality. We believe this semantic approach will make it easier to build and maintain environmental informatics portals and empower local communities to track environmental concerns supported by transparent and accessible environmental data.

Acknowledgment

The authors would like to thank the students in the Semantic e-Science 2010 course and Advanced Semantic Technologies 2011 course for their feedback and discussion.

References

[1] "Boil Water Advisory Issued for Bristol County Water Authority" Rhode Island Department of Health, September 8, 2009. http://www.ri.gov/press/view/9685

[2] Morgan, T. J. 2009. "Bristol, Warren, Barrington residents told to boil water" Providence Journal, September 8, 2009. http://newsblog.projo.com/2009/09/residents-of-3.html

[3] Raskin, R. G. and Pan, M. J., 2005. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). Computers & Geosciences 31(9): 1119-1125.

[4] Hobbs, J. R. and Pan, F., 2006. Time Ontology in OWL, W3C Working Draft 27 September 2006. <http://www.w3.org/TR/owl-time/>

[5] Manola, F., Miller, E., McBride, B., 2004. RDF Primer. W3C Recommendation. http://www.w3.org/TR/rdf-syntax/

[6] Lebo, T., Williams, G.T., 2010. Converting governmental datasets into linked data. Proceedings of the 6th International Conference on Semantic Systems, I-SEMANTICS ’10, pp. 38:1–38:3. doi:http://doi.acm.org/10.1145/1839707.1839755.

[7] Hitzler, P., Krotzsch, M., Parsia, B., Patel-Schneider, and P., Rudolph, S., 2009. OWL 2 Web Ontology Language Primer. W3C recommendation. <http://www.w3.org/TR/owl2-primer/>

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

[9] Sirin, E., Parsia, B., Cuenca-Grau, B., Kalyanpur, A., and Katz, Y. 2007. Pellet: A practical OWL-DL reasoner. Journal of Web Semantics 5(2): 51-53. doi:10.1016/j.websem.2007.03.004

[10] Carroll, J.J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., and Wilkinson, K. 2004. Jena: Implementing the semantic web recommendations. Proceedings of the 13th International World Wide Web Conference, pp. 74-83. doi: 10.1145/1013367.1013381

[11] McGuinness, D.L., Ding, L., Pinheiro da Silva, P., and Chang, C. 2007. PML 2: A Modular Explanation Interlingua. Workshop on Explanation-aware Computing, July 22-23, 2007.

[12] Pinheiro da Silva, P., McGuinness, D.L., and McCool, R. 2003. Knowledge Provenance Infrastructure. IEEE Data Engineering Bulletin, vol. 26.

[13] Chen, Zhiyuan, Gangopadhyay, Arrya, Holden, Stephen H., Karabatis, George, McGuire, Michael P., 2007. Semantic integration of government data for water quality management. Government Information Quarterly 24(4): 716–735. doi: 10.1016/j.giq.2007.04.004.

[14] Chau, K.W., 2007. An Ontology-based knowledge manage ment system for flow and water quality modeling. Advances in Engineering Software 38(3): 172-181.

[15] Ceccaroni, L., Cortes, U. and Sanchez-Marre, M., 2004.OntoWEDSS: augmenting environmental decision-support systems with ontologies, Environmental Modelling & Software 19(9): 785-797.

[16] McGuinness, D. L., and Pinheiro da Silva, P., 2004. Explaining answers from the semantic web: The inference web approach. Journal of Web Semantics 1(4):397-413.

[17] Zhao, J., Goble, C. A., Stevens, R. and Bechhofer S., 2004. Semantically linking and browsing provenance logs for e-science. Semantics of a Networked World 3226: 158-176. doi: 10.1007/978-3-540-30145-5_10

[18] Myers, J., Pancerella, C., Lansing, C., Schuchardt, K., and Didier, B., 2003. Multi-scale science: Supporting emerging practice with semantically derived provenance. ISWC workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data.