Ontology-Enabled Virtual Observatories: Semantic Integration in Practice

Deborah L. McGuinness[12], Peter Fox[3], Luca Cinquini[3], Tony Darnell[3], Patrick West[3], James L. Benedict[2], Jose Garcia[3], and Don Middleton[1]

[1] McGuinness Associate, Stanford, CA 94305 {dlm | jbenedict}@mcguinnessassociates.com
[2] Stanford University, Stanford, CA 94305 dlm@cs.stanford.edu
[3]National Center for Atmospheric Research, Boulder, Co
{pfox | luca | tdarnell | jgarcia | pwest | don}@ucar.edu

Abstract. Data integration across multiple disciplines is a key requirement for virtual observatories. We have designed and implemented a virtual observatory for the areas of solar, solar-terrestrial, and space physics that leverages semantic technologies. Our work provides an implemented demonstration of ontology-enabled semantic interoperability. We have encoded an ontology in OWL covering topics of importance to atmospheric and solar researchers. This ontology is used to inform our virtual observatory services. Our goal was to provide semantically-enabled search and analysis functionalities that access a virtual repository of scientific data. The target user community includes a broad range of scientific researchers over a wide array of data sources. We have implemented a number of use cases that demonstrate semantic integration across various solar atmospheric data sets and the impacts of solar activity on the atmosphere. The implementation is deployed and two scientific user communities are migrating over to full usage in October, 2006. This abstract and accompanying demonstration augments a paper in the Semantic Web in Use track of the main ISWC conference.

Keywords: semantic integration, scientific data, data integration, virtual observatories, semantic web in action

Our work focuses on enabling sharing of scientific data across diverse sub-disciplines. We observed that many times successful data sharing still requires personal interactions among lead researchers in each of the efforts contributing data sets. Our goal was to enable data sharing through online tools and services that have access to multiple data sets, thereby minimizing and/or eliminating the need for human intervention in the data sharing process. Our team includes experts in research (and related data sets) in multiple science disciplines in addition to experts in semantic technologies. Our long range goal is to design, develop, and implement a virtual observatory service for a wide range of scientific data. Our current work [2, 3] across multiple projects focuses on solar terrestrial data and thus we will report on our efforts on the Virtual Solar Terrestrial Observatory.

A Virtual Observatory [1] can be viewed as an online environment for searching, integrating, and analyzing distributed scientific databases. In order to provide results that leverage multiple (overlapping) datasets effectively, data integration is required. We use background domain ontologies to enable query, analysis, and plotting tools that understand enough about the data sets and individual data elements to effectively retrieve and manipulate the data in a meaningful and explainable manner.

The Virtual Solar Terrestrial Observatory has two driving data services aimed at atmospheric and solar researchers. The CEDARweb – Coupling, Energetics, and Dynamics of Atmospheric Regions [4] – focuses on the Earth's upper atmosphere. It contains data holdings from numerous instruments and previously embodied a number of domain-specific controlled vocabularies and associated data such as SPDML: Space Physics Data Markup Language [7] and CISM: Center for Integrated Space-Weather Modeling [5]. MLSO: Mauna Loa Solar Observatory [6] – focuses on the sun. Our background ontology [3] contains information about observatories, instruments (along with their operating modes, measurement parameters and capabilities), data products, etc. The implemented system was built using semantic web tools: Protégé and Swoop for building the ontologies, Pellet for reasoning, and Java-based services for access to the data repositories. The ontology drives the services available.

Our demonstration shows how prototypical users access scientific data across our deployed services in solar, solar-terrestrial, and atmospheric research data. We highlight how the background ontologies support the user interaction – in particular in forming semantically and syntactically appropriate queries, supporting retrieval across multiple data collections, and providing semantically-informed data plots including related information and appropriate axis/labeling choices. Semantically-enabled workflows help users access data from multiple observatories containing a variety of instruments that may be operated in various operating modes. Users may want to obtain data in time periods associated with particular atmospheric events. Background knowledge is used to determine possible/appropriate operating modes, associated parameters, classes of instruments, and appropriate plotting options for the data. The services have been deployed beginning in September, 2006 and we expect full community migration in October in CEDAR and MSLO. The services and related materials can be accessed from http:// www.vsto.org.

References

1. Robert Brunner, S. George Djorgovski, and Alex Szalay. The US NVO White Paper: Toward a National Virtual Observatory: Science Goals, Technical Challenges, and Implementation Plan, Virtual Observatories of the Future, ASP Conference Proceedings Astronomical Society of the Pacific, p.353, 2001.
2. Peter Fox and Deborah L. McGuinness. Semantically-Enabled Large-Scale Science Data Repositories, ISWC Semantic Web in Use Track, Athens, Ga, Nov., 2006.
3. Deborah L. McGuinness, Peter Fox, and Don Middleton. Solar-Terrestrial Ontologies. Proceedings of the Geoinformatics Conference, Reston, Virginia, May 10-12, 2006.
4. CEDARweb: Coupling, Energetics, and Dynamics of Atmospheric Regions. http://cedarweb.hao.ucar.edu/
5. CISM: Center for Integrated Space-Weather Modeling http://www.bu.edu/cism/
6. MLSO: Mauna Loa Solar Observatory. http://www.bu.edu/cism/
7. SPDML: Space Physics Data Markup Language. http://sd-www.jhuapl.edu/SPDML