Abstract

Following work presented at the 2010 AGU Fall Meeting, we present a number of real-world collections of semantically-enabled scientific metadata ingested into the Tetherless World RDF2HTML system as structured data and presented and edited using that system. Two separate datasets from two different domains (oceanography and solar sciences) are deployed for use in three different web environments, Drupal, MediaWiki, and a custom web portal written in Java, to highlight the cross-platform nature of the data presentation. In addition, a single domain dataset is shared between two separate portal instances to demonstrate the ability of this system to offer distributed access and modification of content across the Internet. Lastly, we will present how future improvements in this and other systems will evolve distributed, decentralized collaborations for scientific data sharing across multiple research groups.

Visualization, creation, and editing of Semantic Content within a CMS

Content can be generated using queries from multiple locations. For example, the VSTO 3.0 data portal, currently being developed, is using query and view files from its own repository, the BCO-DMO SPARQL endpoint, and the Tetherless World Constellation endpoint. Content can be generated using queries from multiple locations.

Challenges and Experiences

What are we sharing here? Query files, XSL files, semantic information accessible through SPARQL endpoints and varying storage mechanisms across multiple physical locations. The visualization of and storage of semantic information is distributed. RDF2HTML is linking together scientific communities with research organizations, academic institutes, government agencies, and industry.

Why not Drupal 7 or Semantic MediaWiki? The information is localized, the views to the one site, and the semantic capabilities are built off of services within that were originally written for other features. What about when the next CMS comes along, or a system that does not use a CMS? RDF2HTML is built so that new CMS and new web sites can display and manage semantic content using the same infrastructure.

Acknowledgments:

Professors Jim Hendler, Deborah McGuinness, and Peter Fox – Tetherless World Constellation Chair. Stephon Zednik, Eric Rozell, Tim Lebo, Greg Williams, Dominic DiFonzo – staff and students at TWC who have contributed to the design and development of this project.

Sponsors:

Tetherless World Constellation, Rensselaer Polytechnic Institute

Glossary:

BCO-DMO/WHOI – Biological and Chemical Oceanography Data Management Office at Woods Hole Oceanographic Institution
CEDAR/HAD/NCAR – Coupling, Energetics and Dynamics of Atmospheric Regions/High Altitude Observatory, National Center for Atmospheric Research
CMS – Content Management System
OWL – Web Ontology Language
RDFa – Resource Description Framework Schema
RDF – Resource Description Frameworks
RDF/OWL – The Virtual Solar Terrestrial Observatory / Tetherless World Constellation
RDF2HTML – Extensible Stylesheet Language
RPI/TWC – Virtual Solar Terrestrial Observatory
Semantic MediaWiki – semantic extensions to MediaWiki
VSTO 3.0 – Virtual Solar Terrestrial Observatory

The Tetherless World Constellation web site, http://www.twc.rpi.edu, is created within a Drupal environment, as are some of our research projects. By writing three separate Drupal modules, one for creating and editing semantic content, one for displaying the semantic content, and one for managing our semantic information for documents (Publications, presentations, papers, book chapters, journal entries, and more) TWC collaborates with many different scientific communities, such as the Coupling, Energetics and Dynamics of Atmospheric Regions (CEDAR – http://cedarweb.hao.ucar.edu), CEDAR utilizes a MediaWiki to organized and provide services for their Solar Terrestrial Community. Included in this are extensions that provide for the display of semantic content. This content is also available for the data access portal VSTO (Virtual Solar Terrestrial Observatory – http://www.vsto.org).

The Biological and Chemical Oceanography Data Management Office (BCO-DMO) MapServer (right) uses a SPARQL endpoint to a semantic triple-store to provide custom and mapping capabilities for their data portal. This same endpoint is queried to provide semantic content for their web site and also project information at the Tetherless World Constellation. Both the BCO-DMO web site and TWC project pages use RDF2HTML to view content.