



Enhancing Geoscience Research Discovery Through the Semantic Web

Linda R. Rowan (1), M. Benjamin Gross (1), Matthew Mayernik (2), Huda Khan (3), Frances Boler (1), Keith Maull (2), Don Stott (2), Steve Williams (2), Jon Corson-Rikert (3), Erica M. Johns (3), Michael Daniels (2), Dean B. Krafft (3), and Charles Meertens (1)

(1) UNAVCO, Boulder, CO, United States (rowan@unavco.org), (2) UCAR, Boulder, CO, United States (mayernik@ucar.edu), (3) Cornell University, Ithaca, NY United States

UNAVCO, UCAR, and Cornell University are working together to leverage semantic web technologies to enable discovery of people, datasets, publications and other research products, as well as the connections between them. The EarthCollab project, a U.S. National Science Foundation EarthCube Building Block, is enhancing an existing open-source semantic web application, VIVO, to enhance connectivity across distributed networks of researchers and resources related to the following two geoscience-based communities: (1) the Bering Sea Project, an interdisciplinary field program whose data archive is hosted by NCAR's Earth Observing Laboratory (EOL), and (2) UNAVCO, a geodetic facility and consortium that supports diverse research projects informed by geodesy. People, publications, datasets and grant information have been mapped to an extended version of the VIVO-ISF ontology and ingested into VIVO's database. Much of the VIVO ontology was built for the life sciences, so we have added some components of existing geoscience-based ontologies and a few terms from a local ontology that we created. The UNAVCO VIVO instance, connect.unavco.org, utilizes persistent identifiers whenever possible; for example using ORCID IDs for people, publication DOIs, data DOIs and unique NSF grant numbers. Data is ingested using a custom set of scripts that include the ability to perform basic automated and curated disambiguation. VIVO can display a page for every object ingested, including connections to other objects in the VIVO database. A dataset page, for example, includes the dataset type, time interval, DOI, related publications, and authors. The dataset type field provides a connection to all other datasets of the same type. The author's page shows, among other information, related datasets and co-authors. Information previously spread across several unconnected databases is now stored in a single location. In addition to VIVO's default display, the new database can be queried using SPARQL, a query language for semantic data. EarthCollab is extending the VIVO web application. One such extension is the ability to cross-link separate VIVO instances across institutions, allowing local display of externally curated information. For example, Cornell's VIVO faculty pages will display UNAVCO's dataset information and UNAVCO's VIVO will display Cornell faculty member contact and position information. About half of UNAVCO's membership is international and we hope to connect our data to institutions in other countries with a similar approach. Additional extensions, including enhanced geospatial capabilities, will be developed based on task-centered usability testing.