



Extending VIVO, a semantic web app, to share semantic data across institutions

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

(1) UNAVCO, Boulder, CO, United States (rowan@unavco.org), (2) University Corporation for Atmospheric Research, Boulder, United States (mayernik@ucar.edu), (3) Cornell University, Ithaca, United States

The EarthCollab project, a U.S. National Science Foundation EarthCube Building Block, is extending an existing open-source semantic web application, VIVO, to enable the exchange of information about scientific researchers and resources across institutions. VIVO has been implemented by more than 100 universities and research institutions to highlight research and institutional achievements. Most implementations of VIVO, however, gather information about a single organization. The EarthCollab project VIVO extensions enable cross-linking of VIVO instances to reduce duplication of information about the same people and scientific resources, and enable dynamic linking of related information across VIVO installations. EarthCollab is a collaboration between UNAVCO, a geodetic facility and consortium that supports diverse research projects informed by geodesy, The Bering Sea Project, an interdisciplinary field program whose data archive is hosted by NCAR's Earth Observing Laboratory, and Cornell University.

Test cross-linking implementations have been deployed by UNAVCO and Cornell to demonstrate the following core features: 1. Look up people and things at an external VIVO instance, 2. Assert equivalence between URIs at the two institutions (i.e. a person) using the 'owl:sameAs' property, 3. Provide a subset of data as RDF and JSON as a service from one institution, 4. Display the data requested from the service at the other institution's site, and 5. Allow a user to distinguish between data sources when displayed on one page. Currently, a curator makes the 'sameAs' assertions manually, but persistent and unique identifiers such as ORCIDs for people and DOIs for datasets could be used to automate the process.

As development of the cross-linking extension continues, we must address a number of open questions. For example, if cross-linking institutions have duplicate or conflicting information, should one institution be considered the canonical source, or should the information be displayed in parallel? Should cross-linked data be cached locally in case of a service disruption on the external system? For what types of information can we leverage public persistent identifier lookup services to automate 'sameAs' assertions? Solutions to these challenges will be the focus of EarthCollab's continued development and eventual inclusion of cross-linking features in the core VIVO code.