Guiding development of a semantic web app: End-user engagement in the EarthCollab project

Gross, M. Benjamin¹, Johns, Erica M.², Rowan, Linda R.¹, Mayernik, Matthew², Khan, Huda², Daniels, Michael D.¹, Krafft, Dean B.²

¹) UNAVCO, Boulder, CO 2) Cornell University, Ithaca, NY 3) National Center for Atmospheric Research, Boulder, CO

**Shaping the project**

Enabling Scientific Collaboration and Discovery through Semantic Connections, or EarthCollab, is part of the EarthCube Program at the National Science Foundation: EarthCube has proposed extending an existing open-source semantic web application, VIVO, to highlight connections between people, datasets, grants, and research output.

The project includes two use cases: a geodesy-focused implementation at UNAVCO and another at NASA’s Earth Looming Library (ELL). Convey, where VIVO was originally developed, is also part of this collaborative project. EarthCollab held a workshop at the American Geophysical Union Fall Meeting in December 2014 to identify community needs and how EarthCollab might address them. Workshop participants completed a survey in how they find and share research; the survey, which was completed by 34 researchers including the workshop participants, is summarized below (Figure 1).

**Why semantic?**

Semantic technologies use controlled vocabularies and common formats to store machine-readable data that can be easily moved across applications, semantic applications store information in triple-store databases. A traditional relational database holds information in columns and rows; a triple store can think of as a web, where each piece of information is connected to another according to controlled vocabularies.

**Defining Semantic Connections**

Triples are the basic unit of a semantic database. A triple consists of a subject, predicate (verb), and object, each of which are usually described by a uniform resource identifier (URI). For example:

```
```

In plain English, this translates to: 

```
This publication is an important work product.
```

**Using Unique Identifiers**

Unique identifiers make it easier to connect semantic data. Most peer-reviewed science publications now have Digital Object Identifiers (DOIs) to track and connect publications. Similarly, UNAVCO has recently been assigning DOIs to datasets through the EZID service. GIS/GRDC and select NSF4 datasets are currently being assigned DOIs.

**Building out the application**

The VIVO application was customized to better capture the needs of the geodetic community. The customizations included to include ontologies extensions and mapping capabilities.

For the initial release, the database was populated from a diverse array of sources, summarized below (Figure 5). As the application matures, relevant data will be added automatically using APIs from the NSF, ORCID, and CrossRef.

**Future Work**

- Begin cross-linking VIVO instances across institutions.
- Enhance geospatial capabilities of VIVO by extending ontology and application.
- Automate ingest process, including ingest from ORCID.
- Explore integration with other EarthCube web projects.
- Continue customizations based on feedback from usability testing.

**Check it out at** http://connect.unavco.org