Researchers from institutions around borehole seismometers, for airborne laser swath mapping, geophysical methods including GPS, Terrestrial Laser Scanning (TLS), lidar for extreme laser swath mapping, borehole seismometers, strainmeters, and seismometers, and interferometric synthetic aperture radar (InSAR).

Researches from institutions around the world use UNAVCO instruments, data, and expertise to study the earth and how it is changing.

UNAVCO provides services that support research informed by geodesy (the study of Earth’s shape, gravity field, and rotation)

Data Structures
VIVO has been primarily implemented by research universities, especially in the life sciences. This is reflected in the VIVO-ISF ontology. To capture concepts relevant to UNAVCO, we must extend the ontology. We will use concepts from established ontologies wherever possible.

• VIVO-ISF: people, organizations, publications. VIVO-ISF is already a compilation of ontology terms, including concepts from the OBO Foundry (Open Biological and Biomedical Ontology), CITO (Citation Typing Ontology), and BIBO (Bibliographic Ontology).

• DCAT: datasets and catalogs. DCAT (the Data Catalog Ontology) was designed to facilitate interoperability between data catalogs published on the web. DCAT contains concepts not captured in VIVO’s treatment of datasets as a type of publication, such as landing pages and distribution information.

• GCIS: scientific instruments, platforms, and projects. GCIS (Global Change Information System) includes concepts such as sensors, scientific models, and platforms.

• EC local EarthCollab ontology. UNAVCO operates hundreds of continuously operating GPS stations. Station is a concept not defined in any of the above ontologies and will be defined in EarthCollab’s local ontology.

UNAVCO Data Ingest Process

Employees
MS Dynamics

Datasets
UNAVCO API (in development)

Abstracts
web scraping > plain

Publications
Google Scholar > asix

UNAVCO VIVO ingest (python scripts)

RDF

VIVO Triple Store

Get metadata from CrossRef using CrossRef API

Check for existing objects in VIVO using VIVO Query API

UNAVCO Data Ingest Process

Data Ingest Details
UNAVCO’s current data ingest process is illustrated by the flowchart on the left (Figure 15). A major challenge has been collecting data for VIVO from a number of separate databases maintained by separate people.

Ingest process highlights:
• Python code uses rdf8 library to allow output of multiple RDF file types.
• Using the VIVO Query API, checks VIVO database for duplicate objects. Unique identifiers, such as DOIs, are used when possible and string matching used otherwise. This module is also used to check for collisions when generating random URIs.
• Includes a rough implementation of a disambiguation process using fuzzy string matching and manual curation. Name variations that are confirmed to be stored in a Match List file, which is checked by the script before requesting curator input.

• Currently, triples are loaded into VIVO using the Add/Remove RDF tool. As development continues and we begin to automate ingest, triples will be loaded using the VIVO Update API.

Progress
Figure 17 (right): UNAVCO has made significant progress identifying data sources and ingesting the data into VIVO. Additionally, we have begun requesting employees and select community members sign up for ORCID, a service that provides a unique identifier, publications search capabilities, and a public API.

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.
• Add GPS station data and link to datasets already in VIVO.
• Conduct task-centered usability testing to determine how the VIVO application can be further tailored to a geoscience-centered use case.
• Explore integration with other EarthCube web projects.