UNAVCO

Semantic software, persistent IDs and controlled vocabularies for geoscience metadata

Gross, M. Benjamin[1], Rowan, Linda R.[1], Mayernik, Matthew[2], Krafft, Dean B.[3], Daniels, Michael D.[2]

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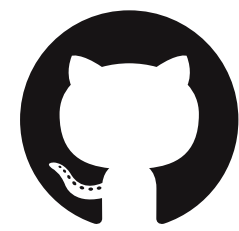
The EarthCollab Project

Enabling Scientific Collaboration and Discovery through Semantic Connections, or EarthCollab, is part of the EarthCube Program at the US National Science Foundation. EarthCollab 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 NCAR's Earth Observing Laboratory (EOL). Cornell, where VIVO was originally developed, is also part of the collaborative project.

The VIVO application was customized to better capture the needs of the geodesy community. The customizations implemented so far include: ontology extensions, mapping capabilities, an expertise and research visualization, and links to UNAVCO's primary data archive interface.

http://earthcube.org/group/earthcollab

EarthCollab: http://git.io/vVErv
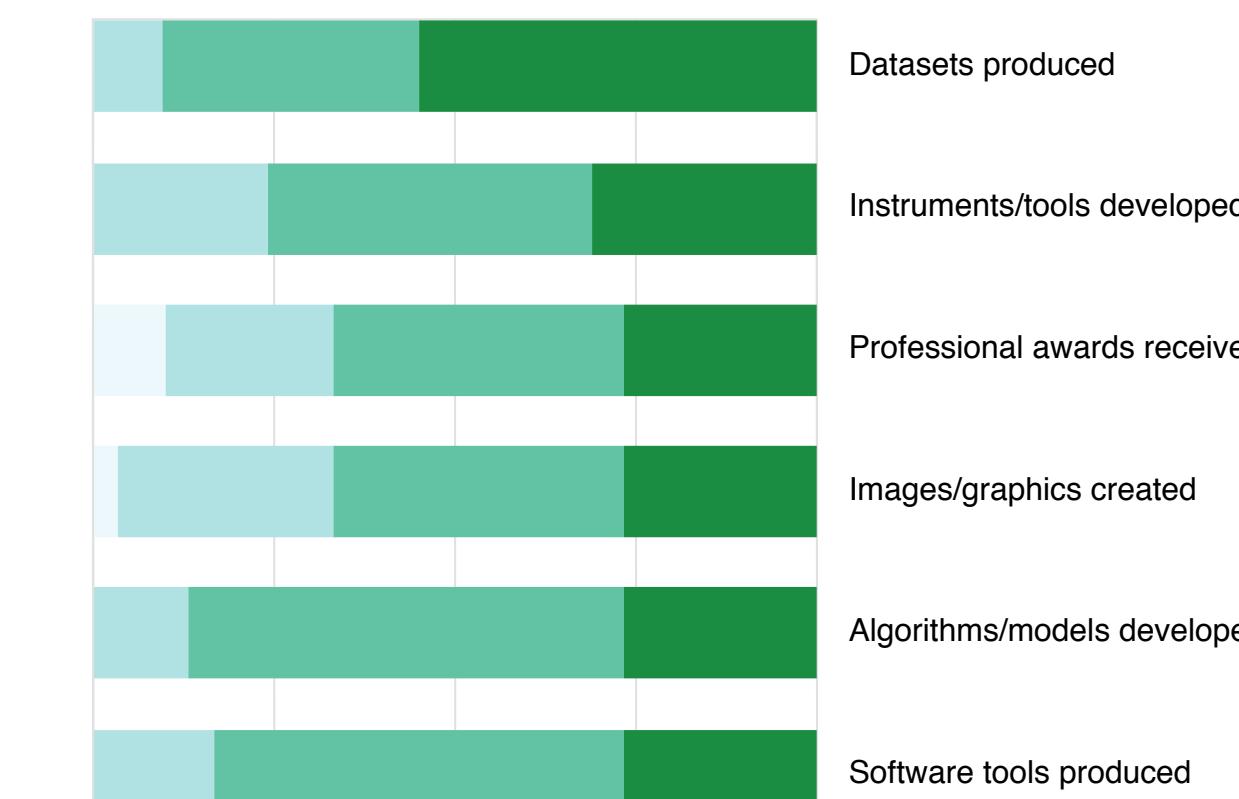UNAVCO VIVO: http://git.io/vG9AJ
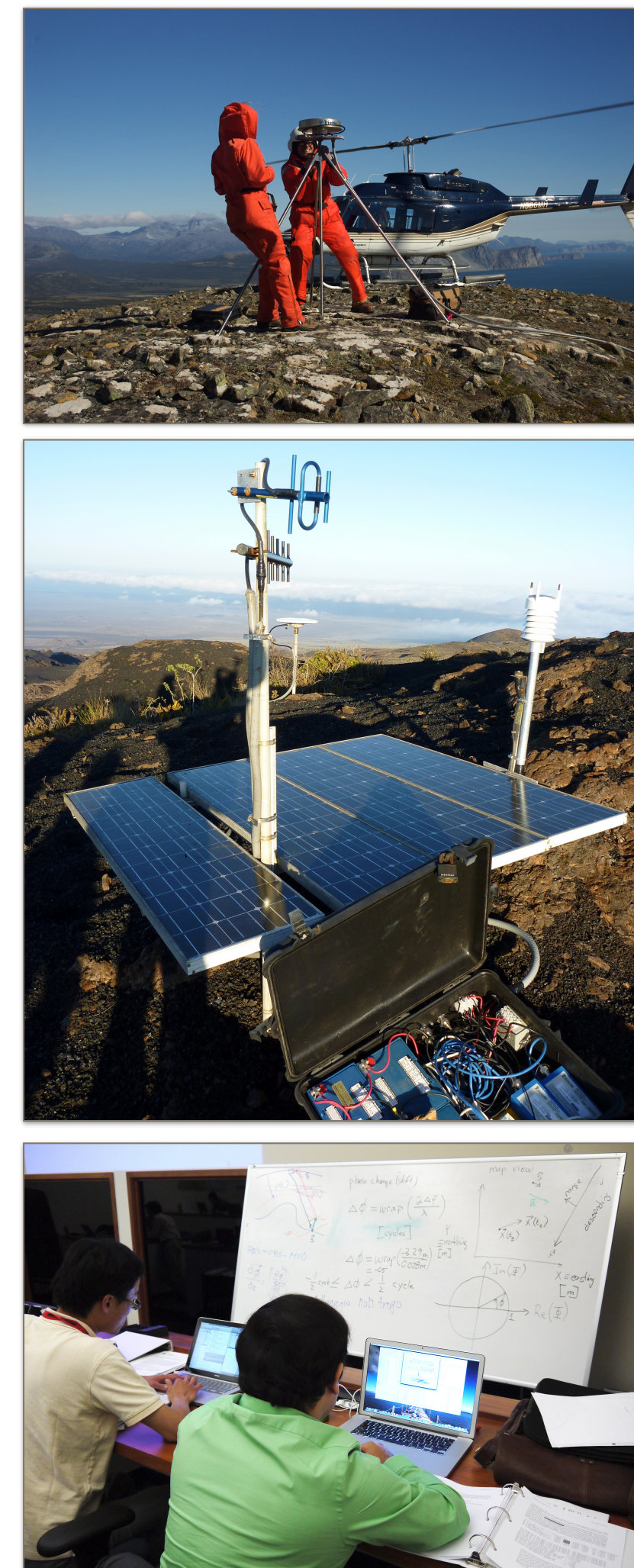
UNAVCO supports a diverse community of researchers with geodetic data and engineering expertise.
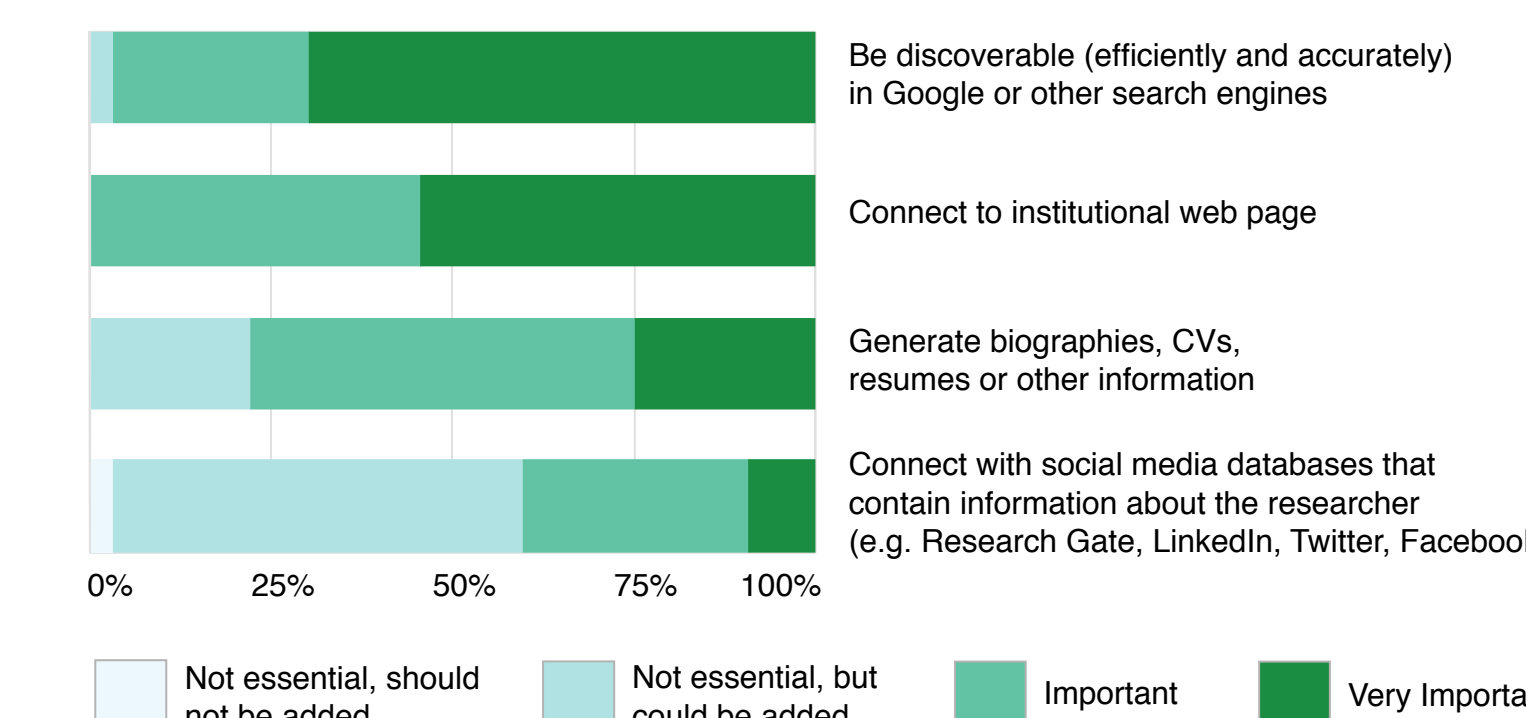
User Engagement

EarthCollab has solicited feedback by conducting usability testing, holding a workshop, and asking the community to complete a survey on how they find and share research, which is partially summarized below.

Which of the following "products" are most important to include in an information platform that displays/describes your work (e.g. on a faculty webpage or in a researcher profile)?

- Datasets produced
- Instruments/tools developed
- Professional awards received
- Images/graphics created
- Algorithms/models developed
- Software tools produced

Which other features would you like an information platform that displays/describes your work to have?

- Be discoverable (efficiently and accurately) in Google or other search engines
- Connect to institutional web page
- Generate biographies, CVs, resumes or other information
- Connect with social media databases that contain information about the researcher (e.g. Research Gate, LinkedIn, Twitter, Facebook)

0% 25% 50% 75% 100%

Not essential, should not be added | Not essential, but could be added | Important | Very Important

Unformatted text

VIVO
connect • share • discover

RDF/XML, N3, Turtle, JSON-LD

Persistent and Unique Identifiers

Unique identifiers make it much 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 begun assigning DOIs to GPS/GNSS and select inSAR datasets through the EZID service.

Name ambiguity presents a challenge when populating a semantic database, partly because unique identifiers are not commonly implemented for people. Publication records often only include partial names, making the process of connecting authors with non-unique last names to their unique record in the VIVO database difficult. A publicly available, unique identifier for a person solves this problem.

ORCID is an open-source, non-profit effort to provide persistent identifiers for people. ORCID has been adopted by publishers such as AGU and EGU as an optional part of the submission and publication process and is increasingly becoming an essential part of a researcher's academic identity. We are leveraging ORCID in the disambiguation process and to facilitate updates to staff and member records using the ORCID public API.

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

Defining Semantic Connections

The VIVO application comes pre-packaged with an ontology that doesn't cover many geoscience concepts. The figure below illustrates a relationship not fully supported in the VIVO-ISF ontology. A project will have many stations which may host multiple instruments (e.g. GPS and meteorological equipment), which are attached to one or more platforms. Similarly, not all roles within the UNAVCO community are defined by VIVO-ISF, as seen on the diagram to the right.
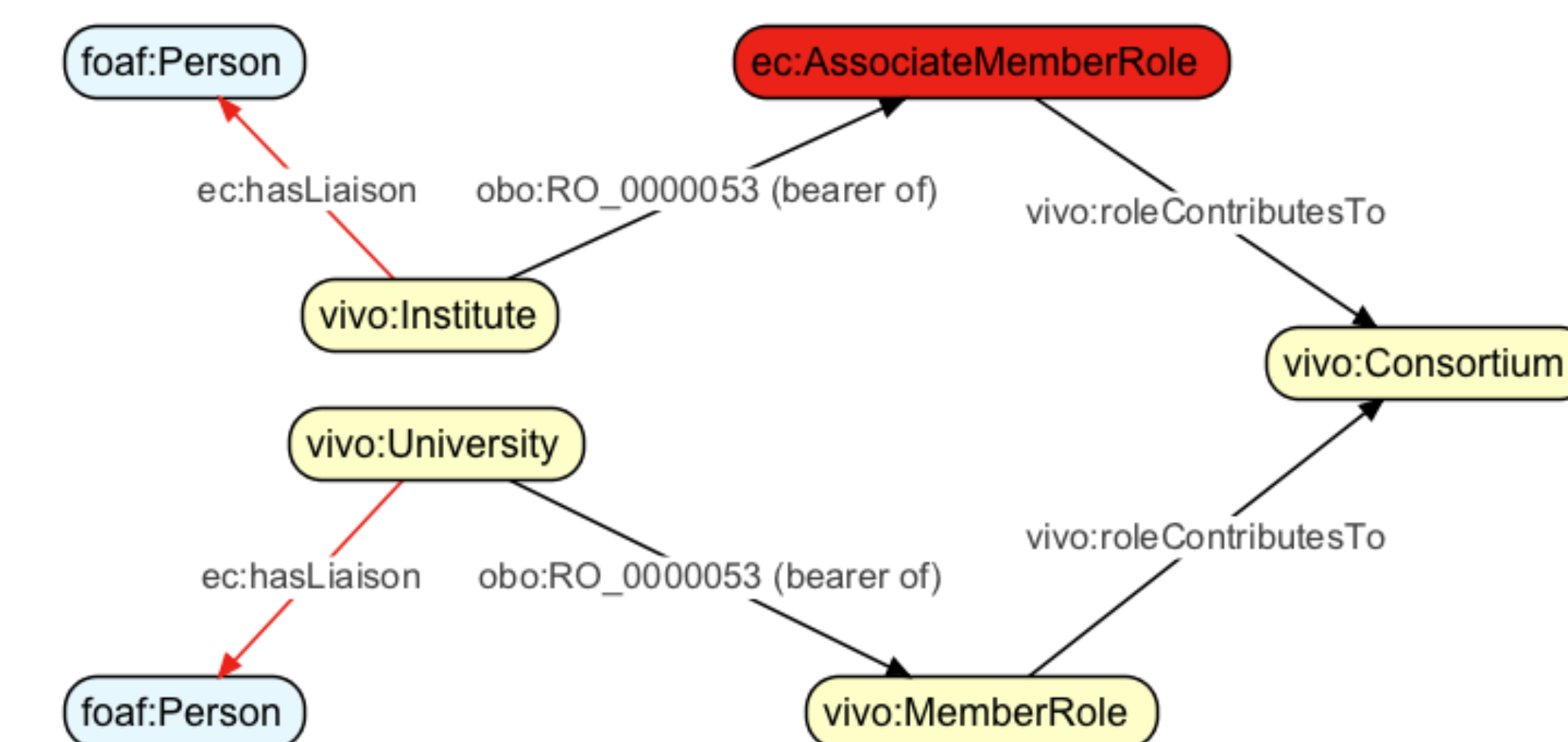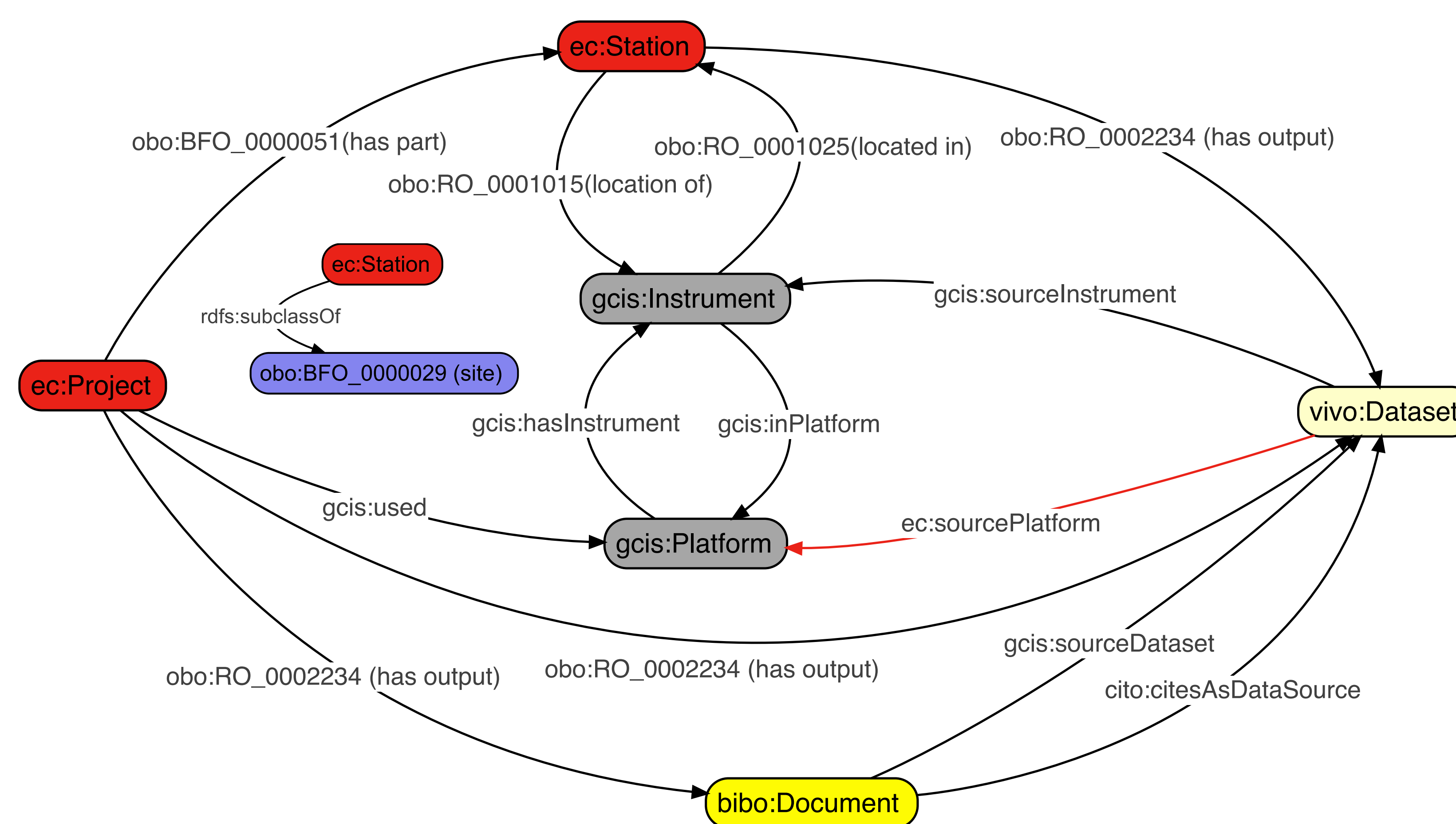
Ontology Prefix Key

Bundled with VIVO software ⇒   Yellow = VIVO-ISF ontology

⇩ EarthCollab additions   contains part of

- Dark Gray = Global Change Information System Ontology (GCIS)
- Green = Virtual Solar Terrestrial Observatory (VSTO)
- Pink = Data Catalog Vocabulary (DCAT)
- Red = EarthCollab Schema (EC)
- Blue = Event ontology
- Light blue = Friend of a Friend (FOAF)
- Gold = Bibliographic Ontology (BiBO)
- Dark purple = Basic Formal Ontology (BFO)
- Light purple = Eagle-I Resource Ontology (ERO)
- Light Green = Simple Knowledge Organization System (SKOS)

No Goldilocks Geoscience Vocabulary
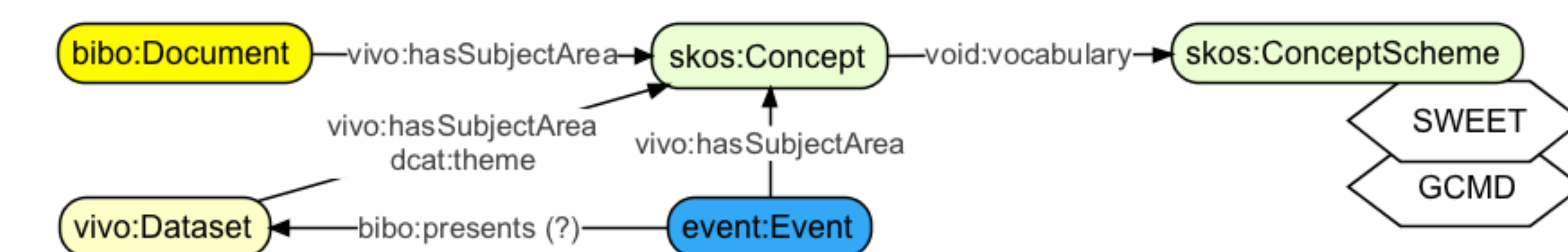
Local term	GCMD[1]	SWEET[2]	FAST[3]	AGU[4]	GSA[5]
Geodesy	geodetics	geodesy	geodesy	geodesy	geodesy
Geomorphology	geomorphic landforms/ processes	-	geomorphology	geomorphology and weathering	geomorphology
Data visualization	data analysis and visualization	visualization	information visualization	data presentation and visualization	-
Archaeology	-	-	archaeological geology	-	archaeological geology
Atmospheric science	-	-	-	atmospheric monitoring with geodetic techniques	-
Unmanned aerial vehicles (UAVs)	-	-	drone aircraft	-	-

1: Global Change Master Directory, 2: Semantic Web for Earth and Environment Technology, 3: Faceted Application of Subject Terminology, 4: America Geophysical Union Index Terms, 5: Geological Society of America Manuscript Classifications