



CINERGI: Community Inventory of EarthCube Resources for Geoscience Interoperability

Ilya Zaslavsky (1), Luis Bermudez (2), Jeffrey Grethe (3), Amarnath Gupta (1), Leslie Hsu (4), Kerstin Lehnert (4), Tanu Malik (5), Stephen Richard (6), David Valentine (1), and Thomas Whitenack (1)

(1) UCSD, San Diego Supercomputer Center, La Jolla, United States (zaslavsk@sdsc.edu), (2) Open Geospatial Consortium, Washington, United States, (3) UCSD, La Jolla, United States, (4) Lamont-Doherty Earth Observatory, Columbia University, Palisades, United States, (5) University of Chicago, Chicago, United States, (6) Arizona Geological Survey, Tucson, United States

Organizing geoscience data resources to support cross-disciplinary data discovery, interpretation, analysis and integration is challenging because of different information models, semantic frameworks, metadata profiles, catalogs, and services used in different geoscience domains, not to mention different research paradigms and methodologies. The central goal of CINERGI, a new project supported by the US National Science Foundation through its EarthCube Building Blocks program, is to create a methodology and assemble a large inventory of high-quality information resources capable of supporting data discovery needs of researchers in a wide range of geoscience domains. The key characteristics of the inventory are: 1) collaboration with and integration of metadata resources from a number of large data facilities; 2) reliance on international metadata and catalog service standards; 3) assessment of resource “interoperability-readiness”; 4) ability to cross-link and navigate data resources, projects, models, researcher directories, publications, usage information, etc.; 5) efficient inclusion of “long-tail” data, which are not appearing in existing domain repositories; 6) data registration at feature level where appropriate, in addition to common dataset-level registration, and 7) integration with parallel EarthCube efforts, in particular focused on EarthCube governance, information brokering, service-oriented architecture design and management of semantic information.

We discuss challenges associated with accomplishing CINERGI goals, including defining the inventory scope; managing different granularity levels of resource registration; interaction with search systems of domain repositories; explicating domain semantics; metadata brokering, harvesting and pruning; managing provenance of the harvested metadata; and cross-linking resources based on the linked open data (LOD) approaches. At the higher level of the inventory, we register domain-wide resources such as domain catalogs, vocabularies, information models, data service specifications, identifier systems, and assess their conformance with international standards (such as those adopted by ISO and OGC, and used by INSPIRE) or de facto community standards using, in part, automatic validation techniques. The main level in CINERGI leverages a metadata aggregation platform (currently Geoportal Server) to organize harvested resources from multiple collections and contributed by community members during EarthCube end-user domain workshops or suggested online. The latter mechanism uses the SciCrunch toolkit originally developed within the Neuroscience Information Framework (NIF) project and now being extended to other communities. The inventory is designed to support requests such as “Find resources with theme X in geographic area S”, “Find datasets with subject Y using query concept expansion”, “Find geographic regions having data of type Z”, “Find datasets that contain property P”. With the added LOD support, additional types of requests, such as “Find example implementations of specification X”, “Find researchers who have worked in Domain X, dataset Y, location L”, “Find resources annotated by person X”, will be supported.

Project’s website (<http://workspace.earthcube.org/cinergi>) provides access to the initial resource inventory, a gallery of EarthCube researchers, collections of geoscience models, metadata entry forms, and other software modules and inventories being integrated into the CINERGI system.

Support from the US National Science Foundation under award NSF ICER-1343816 is gratefully acknowledged.