Alliance Testbed Project: Data Submission Hub

Richard, S.M., Carbotte, S., Ferrini, V., Lehnert, K.A., Shane, N., Song, L., Zaslavsky, I.

Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY

The Alliance Testbed Project (ATP) is an EarthCube Integrative Activity developing the architecture and services of a data repository infrastructure that is shared by Alliance partners to advance usability, interoperability, sustainability, and efficient operation of domain-specific data providers and integration with an EarthCube architecture. ATP has focused on developing a data submission hub (DASH) platform that serves as an entry point for scientists to register, publish, and preserve new resources. The Interdisciplinary Earth Data Alliance (IEDA) is a microcosm of the EarthCube community, managing a wide variety of solid Earth data types. DASH balances general and repository-specific metadata acquisition and curation, a key challenge in managing long-tail geoscience data. Its goal is to provide a platform to acquire standardized high-level metadata, and provide configurable workflows that guide data submitters through resource-specific data validation, quality evaluation, and documentation according to requirements of target partner repositories and the submitted resource type. The project has developed a high level metadata profile based on the DataCite requirements, and tested basic submission workflows for two resource types (tabular geochemistry data and gridded bathymetry data) as a starting point for more generalized resources. The prototype framework includes OpenID-based login, a dashboard for tracking submission status, invoking file introspection processes to extract metadata, form-based metadata review and editing, messaging to reviewers, and accession processing to publish metadata via the IEDA catalog, put files at distribution end points, and produce a bagit package for long-term archiving. Metadata for IEDA resources are made available using the Geoportal catalog application, with ISO 19139 XML as the metadata interchange format. Next steps are to integrate the submission hub workflow into a configurable, pluggable metadata enhancement workflow (‘pipeline’) in collaboration with the Data Discovery Hub EC BB project, to provide an EarthCube workbench resource for registering and releasing research information products.