

Member Node Description: EDI (Environmental Data Initiative)

Version 1.0 01/23/17 Mark Servilla

General

Name of resource: Environmental Data Initiative data repository
URL(s): portal.edirepository.org
Institutional affiliation(s): University of New Mexico, University of Wisconsin
Primary geographic location: Albuquerque, New Mexico USA
Project Director & contact info: Corinna Gries; cgries@wisc.edu
Technical Contact & contact info: Mark Servilla; mark.servilla@gmail.com
Age of resource: Since July 2016
Funding support: National Science Foundation
Proposed Unique Identifier: urn:node:EDI

Content

Content description/collection policy (1 paragraph, domain and spatial/temporal coverage, uniqueness of content, exclusions, as applicable):

Content includes data collected by environmental and ecological communities, with specific content from the Long-term Research in Environmental Biology, Organization of Biological Field Stations, and Macrosystems Biology programs of the National Science Foundation. All content will be public access with default licensing for any non-licensed data package to be Creative Commons "Public Domain" CC0.

Types of data (complex objects, text, image, video, audio, other):

Data types may include tabular, gis, imagery, audio, and model types.

Data and metadata availability (rights, licensing, restrictions):

All data will be public access, but may contain various license restrictions. Contributed data without an explicit license will default to Creative Commons "Public Domain" CC0.

Option for embargo (yes/no, duration):

No

Size of holdings (number and size of datasets, mean and median granules (files) per dataset):

The data holdings are anticipated to be less than 1,000 individual data packages and less than 10Mb per data package. A data package will consist of one science metadata object, one data package quality report, and 1 to many data objects.

Please describe recent usage statistics, if known, including information on annual data product downloads, annual number of users, annual number of data products used in publications:

Unknown

User interactions

How does a user contribute data? (what can be deposited, how are data prepared, are specific software required, documentation/support available)

Data are contributed as complete packages. Science metadata are uploaded into the repository with direct access to data objects. Data objects are subsequently uploaded by the system and stored on-site/in-situ.

How does a user acquire / access data?

User access is through either a user driven web-based interface or REST-based API.

What user support services are available (both for depositing and accessing/using data)?

The Environmental Data Initiative provides direct user support and consultation.

How does the resource curate data at the time of deposit?

All data packages are quality assured and checked at time of upload; data packages are backed-up with ongoing integrity analysis.

Technical characteristics and policies

Software platform description, incl. data search and access API(s):

All data are stored in the Provenance Aware Synthesis Tracking Architecture (PASTA) data repository. This repository is developed primarily in the Java software language and provides a REST-based API for direct access. User portals are secondary to the repository, also developed primarily in Java and Python, and may be customized for use by multiple communities.

Service reliability (including recent uptime statistics, frequency of hardware refresh, if known):

The PASTA data repository has been in continuous operations since January 2013. All services are operated within virtual environments supported by VMware ESXi. Virtual guest hardware images are refreshed approximately every two years. Virtual host hardware has a planned refresh of every 4-5 years.

Preservation reliability (including replication/backup, integrity checks, format migration, disaster planning):

Preservation reliability includes near-line and off-site data package backups, including to Amazon AWS Glacier for long-term durable storage. Integrity analysis is performed at all stages and on all systems.

User authentication technology (incl. level of create/modify/delete access by users):

Current CRUD operations require user authentication through an existing LDAP user directory registry.

Data identifier system and data citation policy, if available:

The data identifier system follows the historical pattern set forth by the LTER Network when the LTER Network utilized Metacat as their data catalog – this pattern utilizes identifiers to be composed of a string scope value, a numeric identifier value, and a numeric revision value (e.g., “scope.id.revision”: edi.12.5); this document identifier pattern is used internal to PASTA. All data packages are assigned a Digital Object Identifier, which is registered with DataCite through CDL/EZID. DOIs are only assigned at the package level and are not assigned to individual components of the data package. Public facing identifiers and those anticipated to be used with DataONE will include a combination of PURL and DOIs. The data citation policy follows the policy recommended by the Earth Science Information Partners organization.

Metadata standards (including provenance):

The Environmental Data Initiative only supports the Ecological Metadata Language (EML) standard and incorporates provenance into the metadata in an idiosyncratic fashion.

Capacity/services to DataONE

At what functional tier will you initially be operating? (see <http://bit.ly/MNFactSheet> for definitions)

- Tier 1: Read only, public content**
- Tier 2: Read only with access control**
- Tier 3: Read/write using client tools**
- Tier 4: Able to operate as a replication target**

If you can host data from other member nodes, what storage capacity is available?

NA

Can you provide computing capacity to the broader network? If so, please describe.

NA

Other Services

What other services or resources (such as expertise, software development capacity, educational/training resources, or software tools) can be provided of benefit to the broader network?

NA