

Member Node Description: *IEDA - Interdisciplinary Earth Data Alliance*

Version 1.1

Date:2018-03-12

Authors: Kerstin Lehnert, Suzanne Carbotte, Vicki Ferrini,
Stephen Richard

General

Name of resource:	IEDA - Interdisciplinary Earth Data Alliance
URL(s)-MN:	http://www.iedadata.org/
URL-Data collection:	n.a. There are various repositories. Integrated catalog is at http://app.iedadata.org/catalog
Institutional affiliation(s):	Lamont-Doherty Earth Observatory, Columbia University
Primary geographic location:	Palisades, New York
Project Director & contact info:	Kerstin Lehnert, lehnert@ldeo.columbia.edu
Technical Contact & contact info:	Vicki Ferrini, ferrini@ldeo.columbia.edu ; Neville Shane, nshane@ldeo.columbia.edu
Age of resource:	Since 2010
Funding support:	National Science Foundation
Proposed Unique Identifier:	urn:node:IEDA

Content

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

IEDA is an alliance of science community focused repository systems. The current component systems are: 1. Earth Chem Library (ECL): a repository for geochemical, petrological, and mineralogical data (new analytical and experimental data, compilations) from the Solid Earth; spatial coverage global, and the age of materials analyzed is not constrained; content is commonly, but not necessarily, data supporting published scientific articles, and might appear in publications. 2. U.S. Antarctic Program Data Center (USAP-DC): a repository for data products from projects funded by the NSF Antarctic Polar Research Program; the repository accepts research products that are not deposited in other science domain specific repositories, and collects metadata for all program research products; spatial coverage is restricted to Antarctica and the Southern Ocean, and temporal coverage is not restricted; datasets might include content that is published in scientific journals. 3. Marine-Geo Digital Library (MGDL): a repository for marine geoscience research data acquired throughout the global oceans and adjoining continental margins; temporal coverage is not restricted; content includes research results that might appear in published scientific journals, as well as sensor field data from research vessels that is unique. 4. Academic Seismic Portal (ASP): a repository for academic active-source seismic data (SCS, MCS, CHIRP, OBS, OOS, Sonobuoy and 3.5 kHz) acquired throughout the global oceans and adjoining continental margins; temporal coverage is not restricted; raw data in the repository is unique, processed data and interpretations might be published in scientific journals.

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

Repository items can consist of one or more files in various formats. IEDA does not restrict file formats, but requests use of standard archival file formats. Content includes text (ASCII, CSV),

spreadsheets (Microsoft Excel), ESRI Shape files, images in various formats, grid formats (NetCDF, GeoTIFF), SEGY, SEGD, as well as custom file formats.

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

Metadata are freely accessible via web-accessible folders at <http://get.iedadata.org/metadata>. Datasets are licensed under CC BY-NC-SA 3.0. Some commercially-acquired seismic data in the MGDL are subject to different licensing conditions.

Option for embargo (yes/no, duration):

Yes. For ECL, up to two years. For MGDL, up to two years after data acquisition. Option for other contributor-defined embargo period.

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

About 870,000 files, ranging from ~100KB up to several TB; Mean file size is about 18MB.

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:

File downloads (2017): 210,000; Publications citing IEDA Datasets (2017): 349; EarthChem Library gets hits from 30 to 100 unique IP addresses per month; MGDL gets hits from 300 to 600 unique IP addresses per month.

User interactions

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

Workflows for data submission start with filling out an online form to provide basic metadata. For EarthChem Library and USAP-DC, one or more files are uploaded to an IEDA server, and the files and metadata are reviewed by a data curator. For MGDL, the user provides metadata and data files, and they are reviewed and ingested into the system by MGDL data curators. E-mail contact with curators is available for data submitters to ask questions and obtain help.

How does a user acquire / access data?

Users access repository content at the file-level through one of several web interfaces that allow selection based on domain-appropriate criteria. Actual data file access is generally through download links on dataset landing pages. Some EarthChem Library data is marshalled into the EarthChem Database, allowing more granular sample and analysis-level access. .

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

Helpdesk (e-mail) and telephone contact with curators is available for assistance with data submission and access.

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

ECL, USAP-DC, MGDL, and ASP submissions are reviewed for consistency and completeness of supplied metadata and for completeness of contributed datasets. Geospatial data of all data sets are validated and data content of files is validated (files are introspected) for relevant metadata.

Technical characteristics and policies

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

Partner systems operate their own data submission and data access applications, tailored for specific communities and data types. These include [EarthChem Library search](#), [MGDL search](#), [USAP-DC search](#), which use various software platforms including Java, PHP, and JavaScript. This heterogeneous deployment reflects the independent development of the partner systems that have been brought together in the alliance.

The IEDA DataBrowser (<http://app.iedadata.org/databrowser/>) provides a map-based search for holdings of partner repositories and related data syntheses based on sample locations, seismic lines, and cruise tracks. The application accesses partner data via OGC Web Feature Services.

The IEDA integrated catalog (<http://catalog.iedadata.org>) is implemented using Geoportal OpenSource software, forked from <https://github.com/Esri/geoportal-server-catalog> at <http://github.com/iedadata/geoportal-server-catalog>; the application is implemented in Java on the server side and JavaScript on the client side. The catalog harvests ISO metadata for all partner resources from web accessible folders that are updated as part of data accession workflows. Web services to access IEDA resources are described at <https://www.iedadata.org/help/web-services/>.

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

Service uptime for 2017 was 99.4%. Hardware refreshes occur as funding allows.

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

All file systems and repository databases are backed up locally within central data center and replicated to a secondary site within the Lamont campus. The MGD sensor field data and ASP seismic data from the ships are mirrored to a secondary local Lamont site and submitted to NCEI in Boulder for Long Term Archiving. Copies of data archives are also sent to the Amazon Glacier for disaster recovery.

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

Some IEDA systems use a home-built authentication tool, GeoPass, based on JOSSO technology. Other software is using OAuth to enable users to login using Google or ORCID credentials. External login accounts are only used to identify data submitters for processing incoming data. Currently, create/modify/delete access is only available to users inside the IEDA local-area network.

Data identifier system and data citation policy, if available:

All repository items have a unique Persistent Identifier (PID), and DOIs are also assigned to most. See <https://www.iedadata.org/help/data-publication/>, for citation policies and recommendations.

Metadata standards (including provenance):

Metadata are accessible in DataCite XML format, and ISO19139 XML. Provenance metadata follow domain-specific best practices for method and data quality documentation.

Capacity/services to DataONE

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

N.A.

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

N.A.

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?

With appropriate resources, IEDA curators can share their expertise in user interaction, data and metadata curation, and linking data, publications, and samples. Software that might have utility for other data facilities is available on GitHub at <http://github.com/iedadata>. IEDA also provides software and services for assigning Persistent IDs to physical samples and linking these to datasets in the metadata.

Harvest endpoints

US Antarctic Program Data Center (USAP): <http://get.iedadata.org/metadata/iso/usap/>

Marine-Geo Digital Library (MGDL) including ASP: <http://get.iedadata.org/metadata/iso/mgdl/>

EarthChem Library: <http://get.iedadata.org/metadata/iso/ecl/>

Metadata Formats:

MGDL: root element gmi:MD_Metadata, namespace <http://www.isotc211.org/2005/gmi>

```
<gmd:metadataStandardName>
```

```
<gco:CharacterString>
```

```
ISO 19115-2 Geographic Information - Metadata - Part 2: Extensions for Imagery and Gridded Data
```

```
</gco:CharacterString>
```

```
</gmd:metadataStandardName>
```

```
<gmd:metadataStandardVersion>
```

```
<gco:CharacterString>ISO 19115-2:2009 (E)</gco:CharacterString>
```

```
</gmd:metadataStandardVersion>
```

ECL, USAP

root element gmd:MD_Metadata, namespace <http://www.isotc211.org/2005/gmd>

```
<gmd:metadataStandardName>
```

```
<gco:CharacterString>
```

```
ISO 19139 Geographic Information - Metadata - Implementation Specification
```

```
</gco:CharacterString>
```

```
</gmd:metadataStandardName>
```

```
<gmd:metadataStandardVersion>
```

```
<gco:CharacterString>2007</gco:CharacterString>
```

```
</gmd:metadataStandardVersion>
```

1. Identifiers

Identifies the **Metadata Record**

```
<gmd:fileIdentifier>  
<gco:CharacterString>  
urn:ieda:metadata:Hydrologic-Margins-Research-Project--2004-2008--McMurdo-Dry-Vall  
eys  
</gco:CharacterString>  
</gmd:fileIdentifier>
```

Identifies the **resource** described by the metadata record. The content item that is in the repository.

```
<gmd:identificationInfo>  
<gmd:MD_DataIdentification>  
<gmd:citation>  
<gmd:CI_Citation>  
.....  
<gmd:identifier>  
<gmd:MD_Identifier>  
<gmd:code>  
<gco:CharacterString>doi:10.15784/600082</gco:CharacterString>  
</gmd:code>  
</gmd:MD_Identifier>  
</gmd:identifier>  
....
```