

Member Node Description: USGS Science Data Catalog (SDC)

Version 1.0 10/28/15 Mike Frame

General

Name of resource: USGS Science Data Catalog (SDC)
URL(s): <http://data.usgs.gov> ,
Institutional affiliation(s): U.S Geological Survey, Core Science Analytics & Synthesis
Primary geographic location: Denver, CO USA
Project Director & contact info: Vivian Hutchison, vhutchison@usgs.gov
Technical Contact & contact info: Mike Frame, mike_frame@usgs.gov
Age of resource: Since 2013
Funding support: USGS
Proposed Unique Identifier: urn:node:USGS_SDC

Content

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

The Science Data Catalog is an initiative by the USGS Core Science Analytics and Synthesis (CSAS) program to locate, evaluate, and access Earth Science data and information. It contains standardized metadata descriptions of Earth Science data sets and information products -- contributed from USGS research organizations -- to help scientists discover a broad spectrum of data. Records describe geospatial and tabular data, in addition to software tools used in analyzing, integrating, and applying data. Metadata descriptions follow the Federal Geographic Data Committee's (FGDC's) Content Standards for Digital Geospatial Metadata; however, other standards are also accepted. The Clearinghouse is a participating node on the FGDC's National Spatial Data Clearinghouse (NSDI), Data.gov, and the Geoplatform.gov.

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

Science Research Studies, Publications, Geospatial Datasets including shape files, netCDF and ASCII

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

Public - open access, free of charge

Option for embargo (yes/no, duration):

No

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

6,800

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:

~2,000 unique visit 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)

Users contribute data to the USGS Science Data Catalog by creating standardized metadata records that describe the data and contain links providing access to the data. Metadata records can be contributed via various metadata creation tools including the USGS Online Metadata Editor (<https://www1.usgs.gov/csas/ome/>), the USGS ESRI Metadata Wizard (<https://www.sciencebase.gov/catalog/item/50ed7aa4e4b0438b00db080a>), and other tools that support the FGDC Standard and its various profiles such as the Biological Data Profile (BDP). Organizations contributing large metadata holdings are harvested by the Science Data Catalog on a weekly basis or as defined by the data providing organization. All metadata contributors maintain complete control over their metadata records via the USGS Data Provider Dashboard (<http://data.usgs.gov/info/dashboardhelp.html>), which provides various reporting and methods for contributing data to the USGS Science Data Catalog.

How does a user acquire / access data?

Metadata and data are acquired from USGS research organizations who primarily want to make their research results freely available to the broad science community. Users acquire data by searching the Science Data Catalog and downloading data via a “Get Data” capability or by the links to data contained in the metadata records. Various data visualizations services for common file formats (i.e. .shp, .csv, .tiff, etc.) are available for previewing and performing basic data analysis functions.

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

Metadata creation assistance services are available for researchers and others requiring help with using standards to create metadata records. Tools such as the CSAS Online Metadata Editor (OME) are available for metadata creation and record deposit to the Science Data Catalog. Support services are also provided for QA/QC of metadata records and technical assistance with metadata contribution to the Science Data Catalog (some users may provide single documents, other multiple records via harvesting). Additionally, a 4-hour Introduction to Metadata training is periodically provided by CSAS.

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

Owners/Users are responsible for data curation. CSAS provides various tools to assist in aspects of data curation such as the OME, Data Provider Dashboard, taxonomic services (www.itis.gov) and the CSAS Biocomplexity Thesaurus (http://www.usgs.gov/core_science_systems/csas/biocomplexity_thesaurus/) for standardized keyword choices to describe data.

Technical characteristics and policies

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

SOLR/Lucene Search system, various web services including OpenSearch, OAI-PMH and OGC services exists.

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

Hardware Refresh is ever 3 years. Uptime is 99%.

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

Offsite backup of metadata records exists at the USGS Denver Facility and USGS Oak Ridge National Lab facility.

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

None

Data identifier system and data citation policy, if available:

Use of Digital Object Identifiers are being adopted and promoted for data linkages via USGS agreement with DataCite.

Metadata standards (including provenance):

FGDC (including the BDP)

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?

Willing to support hosting other Member nodes data, pending scope and need. Storage exists at multiple facilities (Oak Ridge and Denver) to support this need. Federal security policies will be required to adhere to.

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

Currently, supporting over 150 science projects, over 200 users throughout all USGS science areas related to advanced and computational computing. The USGS Applied Research Computing program (http://www.usgs.gov/core_science_systems/csas/arc/index.html) mission is to provide high end computing capabilities and expertise to USGS scientists for the acceleration and expansion of scientific discovery. The program goals are to provide Research and apply advances in High End Computing (HEC) technologies, enable expertise in HEC, provide access to HEC resources (via interagency agreements with NSF and DOE), and collaborate to create efficiencies in the USGS.

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?

Other Services from this Member Node include:

- Training/Outreach expertise in Data Management
- Metadata Editor/Creation Tool for FGDC BDP Metadata
- Advanced computing capabilities to support research data analysis and visualization (http://www.usgs.gov/core_science_systems/csas/arc/index.html)
- Leading organization in USGS/DOI for Open Data activities including catalogs, standards, governance, training, and policies.