Member Node Description: Earth Data Analysis Center (EDAC)

Version 2.0   1/23/14   Soren Scott

General

Name of resource:	EDAC Gstore Repository
URL(s):	http://gstore.unm.edu/dataone/v1
Institutional affiliation(s):	Earth Data Analysis Center, University of New Mexico
Primary geographic location:	Albuquerque, New Mexico, United States
Project Director & contact info:	Karl Benedict, kbene@edac.unm.edu
Technical Contact & contact info:	Soren Scott, sscott@edac.unm.edu
Age of resource:	Since 1964
Funding support:	Various
Proposed Unique Identifier:	urn:node:EDACGSTORE

Content

Content description/collection policy (1 paragraph, domain and spatial/temporal coverage, uniqueness of content, exclusions, as applicable):
Earth Data Analysis Center's (EDAC) Geographical Storage, Transformation and Retrieval Engine (Gstore) platform archives data produced by various NM organizations, including NM EPSCoR and RGIS. Gstore primarily houses GIS and other digital documents relevant to state agencies, local government, and scientific researchers. See RGIS and NM EPSCoR for more information on the scope of those collections. Gstore currently uses the FGDC metadata standard to describe its holdings.

Spatial coverage: state of New Mexico; temporal coverage: at least 1935 to present;

Types of data (complex objects, text, image, video, audio, other):
Complex data (rasters, vectors, non-spatial tabular data), documents (PDF, Word, Excel)

Data and metadata availability (rights, licensing, restrictions):
Public, currently no access restrictions

Option for embargo (yes/no, duration):
Yes; duration depends on the specific project (up to one year for NM EPSCoR, for example). Embargoed data do not appear in any searches against the system.

Size of holdings (number and size of datasets, mean and median granules (files) per dataset):
Roughly 13TB for ~300,000 datasets. Number of files depends on data type and supported representations (files and rasters may just have one file, vectors up to 7 if shapefiles).

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:
Roughly 750,000 requests per year through RGIS (90,000 pages) with an average 1.5GB transferred per day.

User interactions

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

Users provide data to EDAC Clearinghouse staff or through specific projects. In most cases, the users provide basic information about the files – attribute definitions, location information, abstracts, etc, that is used by EDAC staff to generate valid FGDC documentation. In some cases, data files are reformatted to provide additional services (vector conversions, WxS, etc).

How does a user acquire / access data?

Data can be accessed through the RGIS Clearinghouse, the NM EPSCoR data portal or directly through the GSTORE API.

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

Help for data access/data discovery is available through the RGIS Clearinghouse. Help for depositing data is available through EDAC staff through RGIS or NM EPSCoR.

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

When provided data by a third party, such as an NM EPSCoR researcher, the data files are examined to determine if the data can be made available through the expanded services (Is it a multitabbed spreadsheet? Can it be disaggregated meaningfully as separate data files? Is it well-structured or can it be safely reformatted? Is the spatial information reliable?). Before posting, all datasets are assigned to an appropriate category (GSTORE-specific) and ISO Topic Category (external repository registration).

Technical characteristics and policies

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

Pyramid (Python 2.7.2) API with dataset metadata stored in PostgreSQL database and vector attribute data stored in MongoDB 2.2. Metadata/datasets search handled by ElasticSearch 0.9.

Search is available through the GSTORE API by publication date, dataset dates, spatial extent, ontology, data type and geometry type, and keyword.

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

New hardware for the app/databases every couple years as resources are available. Uptime is tricky – the app cluster is generally stable; individual app uptime is unknown.

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

Regular backups of data and applications. The three database systems are replicated clusters.

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

Create/modify/delete by EDAC staff only. Read access through the API is public.

Data identifier system and data citation policy, if available:

All datasets are assigned a UUID(4). Data citation policy forthcoming.

Metadata standards (including provenance):

FGDC, FGDC-RSE, ISO-19115 (139), ISO-19110, ISO-19119, ISO-19115 DS, W3C PROV support planned.

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?
Undetermined at this time.

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

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?
Undetermined at this time.