Member Node Description: New Mexico Experimental Program to Stimulate Competitive Research

Version 1.0   3/25/2015   Authors: Karl Benedict and John Savickas

General

Name of resource:	New Mexico Experimental Program to Stimulate Competitive Research (NM EPSCoR)
URL(s):	http://www.nmespcor.org/data_portal/browsedata
Institutional affiliation(s):	University of New Mexico
Primary geographic location:	New Mexico
Project Director & contact info:	Karl Benedict
	Associate Professor / Director, Research Data
	Services College of University Libraries and Learning Sciences
	University of New Mexico
	kbene@unm.edu
	(505) 277-5256
Technical Contact & contact info:	John Savickas
	EDAC IT Manager / Earth Data Analysis Center
	jsavickas@edac.unm.edu
	(505) 277-3622 ext. 242
Age of resource:	Since July 2015 (planned deployment)
Funding support:	This work was partially funded by National Science Foundation (NSF) EPSCoR Program (Track 1 {Awards: 0447691, 0814449, 1301346} and Track 2 awards {0918635, 1329470})
Proposed Unique Identifier:	urn:node:NMEPSCOR

Content

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

The emphasis of the EPSCoR data collection is on providing discovery of and access to the full range of NM EPSCoR program research data products through a variety of data services and search platforms/interfaces. Given that the NM EPSCoR Program encompasses multiple individual projects, each with its own science focus, the diversity of data made available through the NM EPSCoR DataONE member node is high. These data include remote sensing, GIS, and other geospatial data products; experimental data including tabular data; maps and photos; model/simulation data; human subject survey data; analytic and visualization tools; point-time-series environmental observation data; scanned historic documents; and learning modules. While the primary focus of the EPSCoR data collection is on research data products generated by the various EPSCoR Projects, in some instances data obtained from external data sources (such as the NRCS SNOTEL and SCAN weather data collection) in support of EPSCoR researchers is included as well when the access methods and documentation models used by the EPSCoR data portal provide useful added value to those data products.

Types of data (complex objects, text, image, video, audio, other):
Raw, Images, Tabular, PDF, Text, CSV, Vector, NETCDF, Other.

Data and metadata availability (rights, licensing, restrictions):
With the exception of specific instances where there are patent or other specific and more limiting requirements, EPSCoR data and related metadata are made available under Creative Commons CC-By 4.0 license terms (http://creativecommons.org/licenses/by/4.0/). Exceptions to this general model must be approved in writing by the EPSCoR project director.

Option for embargo (yes/no, duration):
With the exception of climate and water data that are collected automatically by State and Federal agencies and that will be made available as soon after collection as the agency allows, other data collected through this project may be embargoed for a period of up to one year to allow time for publication by students and researchers; any exceptions to this embargo period must be approved by the Project Director in writing.

Size of holdings (number and size of datasets, mean and median granules (files) per dataset):
Varies with researcher data submitted.

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:
Tier 4 node not emplemented yet. Expected date July 2015.

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 via web data entry form for NMEPSCoR Track 1 and NMEPSCoR Track 2 via automated process pulling data from Virtual Watershed Platform.

How does a user acquire / access data?
A web portal provides browse, search and retreval systems. Users can either browse data by category or search for specific terms that are in the metadata. The metadata published through the DataONE member node will include data download links and links to related data services. Metadata for parts of the collection may also be registered with other data discovery systems such as data.gov (http://catalog.data.gov) and UNM's institutional repository, LoboVault (http://repository.unm.edu).

What user support services are available (both for depositing and accessing/using data)?
Users can contact the EDAC suport team by email or phone for help with depositing or accessing data.

How does the resource curate data at the time of deposit?
It's submitted to a relational database and assigned a UUID (Universally Unique Identifier) for tracking, managing and retrieval. Depending upon the specific characteristics of the deposited data, some data products will be assimilated into the hybrid data management system developed by EDAC (GSToRE – http://gstore.unm.edu) to enable value added services (such as data extraction and reformatting services; Open Geospatial Consortium Web Map, Web Feature and/or Web Coverage Services) in addition to basic source data discovery and download.

Technical characteristics and policies

Software platform description, incl. data search and access API(s):
The data management tier includes multiple databases: the geospatially enabled PostgreSQL/PostGIS Object Relational database; the JSON document-based MongoDB database; Elasticsearch for metadata search; and file-based storage for data that are appropriately stored outside of the database management systems. The services tier provides a set of RESTful Web services with which client applications (i.e., Web applications, desktop GIS, analytic tools) interact to discover and access the data held in the system.

Service reliability (including recent uptime statistics, frequency of hardware refresh, if known):
99% depending on researcher request and new development features.

Preservation reliability (including replication/backup, integrity checks, format migration, disaster planning):
Nightly offsite backups of the Tier 4 Member Node. as well as all Tier 4 Member Node data. Both data and VM's have been implemented into EDAC's disaster recovery plan.

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

Data identifier system and data citation policy, if available:
UUID and Metadata for every dataset entered.

Metadata standards (including provenance):
FDGC (RS Extension), ISO 19115 and related standards (19110, 19115-2, 19119, 19139)

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?
5TB

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?
Help with other organization setting up new nodes.