Member Node Description: ONEShare Repository

Version 1.0   2/2/13   John Kunze, Carly Strasser

General

Name of resource:	ONEShare Repository
URL(s):	https://merritt.cdlib.org/m/oneshare_dataup
Institutional affiliation(s):	University of California Curation Center, California Digital Library and University of New Mexico
Primary geographic location:	Oakland, California, USA
Project Director & contact info:	Trisha Cruse, patricia.cruse@ucop.edu, 510-987-9016
	Carly Strasser, carly.strasser@ucop.edu, 510-987-0179
Technical Contact & contact info:	Stephen Abrams, stephen.abrams@ucop.edu, 510-987-0370
	John Kunze, jak@ucop.edu, 510-987-9231
Age of resource:	Since October 2012
Funding support:	Subsidized by CDL and UNM
Unique Identifier:	urn:node:ONEShare

Content

Content description/collection policy (1 paragraph, domain and spatial/temporal coverage, uniqueness of content, exclusions, as applicable):
A repository for tabular data and metadata created using the DataUp tool, sponsored by the California Digital Library and the University of New Mexico.

Types of data (complex objects, text, image, video, audio, other):
Currently only spreadsheet data accepted (XLSX, CSV).

Data and metadata availability (rights, licensing, restrictions):
All public (CC-Zero license).

Option for embargo (yes/no, duration):
No.

Size of holdings (number and size of datasets, mean and median granules (files) per dataset):
5 data objects, 1.4 MB.

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:
Too early to be meaningful.

User interactions

How does a user contribute data? (what can be deposited, how are data prepared, are specific software required, documentation/support available)
ONEShare is an instance of the Merritt repository, with a front door in Oakland (CDL) and a storage node in Albuquerque (UNM). The data model permits deposit of one or more versions of any dataset. Deposit is currently entirely managed by the DataUp too, which supports EML metadata.

How does a user acquire / access data?

All data objects are accessible form the URL above after logging into the Merritt repository as "guest".

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

The DataUp tool manages all deposits. Content is searchable via DataONE interfaces, as well as via the Merritt inventory. Identifiers are obtained from CDL's EZID service.

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

Upon deposit, a dataset ("object") is processed to extract metadata, compute file digests, generate an OAI-ORE resource map (to represent structural relationships), and an object manifest. Objects are replicated, as well as scanned to generate characterization information (e.g. format validation and technical metadata).

Technical characteristics and policies

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

UC3-authored, community-supported, portable open source software based on the CDL micro-services architecture and the MetaCat software stack (supporting DataONE protocols). APIs and command-line interfaces provide a superset of all functionality available in the user interfaces. All APIs are RESTful.

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

24 x 365, 1-4 scheduled downtimes per year.

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

RAID 6, tape backup of all system components, one extra replica of all content, fixity service running every 3 days, high-availability server clusters, standard PCI data center protocols for physical and electronic security.

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

HTTP Basic over SSL, LDAP-based account management, create/modify/delete operations only available to depositor.

Data identifier system and data citation policy, if available:

ARKs using the EZID service (http://n2t.net/ezid).

Metadata standards (including provenance):

EML (Ecological Metadata Language)

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?

N/A

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

In principle, yes, but would have to be subject to review for alignment with CDL priorities and UC computer usage policies.

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?

We are considering expanding the file formats that can be deposited in ONEShare.