

Member Node Description: ONEShare

Version 1.1 4/8/15 Stephen Abrams

General

Name of resource: ONEShare repository
URL(s): <http://oneshare.cdlib.org/>
Institutional affiliation(s): California Digital Library; University of New Mexico
Primary geographic location: Oakland, CA, US
Project Director & contact info: Stephen Abrams, Stephen.Abrams@ucop.edu, +1 510-987-0370
Technical Contact & contact info: Stephen Abrams, Stephen.Abrams@ucop.edu, +1 510-987-0370
Age of resource: Since 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 public open access repository for research data sponsored by the California Digital Library and the University of New Mexico.

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

All types of data are accepted.

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

CC0

Option for embargo (yes/no, duration):

No

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

91 datasets, 125 versions, 1,520 files, 177MB

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:

User interactions

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

Datasets can be contributed to ONEShare via the Dash research data portal at <http://oneshare.cdlib.org/>. Dash supports drag-and-drop file upload and DataCite descriptive metadata entry.

How does a user acquire / access data?

Search and browse for ONEShare datasets via ONEMercury at <https://cn.dataone.org/onemercury/> or directly in the Dash portal at <http://oneshare.cdlib.org/>. Search and browse is also possible via the

ONEShare collection view in the underlying Merritt repository platform at https://merritt.cdlib.org/m/oneshare_dataup. All ONEShare datasets are also indexed in the Data Citation Index (<http://apps.webofknowledge.com/>) and Scopus (<http://www.elsevier.com/online-tools/scopus>).

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

The Dash research data portal mediates dataset deposit to and discovery from the underlying ONEShare repository. Detailed user instructions and answers to other frequently asked questions are available at <https://oneshare.cdlib.org/xtf/search?smode=faqPage>.

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

Upon deposit, a dataset package is processed to extract metadata, generate an OAI-ORE resource map (to represent structural relationships), calculate SHA-256 digests for fixity audit, and assignment of a unique persistent ARK identifier resolvable to the dataset landing page in ONEShare.

Technical characteristics and policies

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

ONEShare is based on CDL's Merritt repository platform with added support for the DataONE MetaCat software implementing a Tier 1 member node. Descriptive information about all ONEShare datasets is automatically synchronized with the DataONE network's coordinating nodes. ONEShare takes advantage of Merritt's distributed micro-services architecture by hosting repository processing at CDL's datacenter while preservation storage is provided by the University of New Mexico. Deposit, management, and discovery of datasets can be done using ONEShare's native Merritt UI or RESTful API, or through the Dash research data portal.

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

24x365. ONEShare is hosted in a managed datacenter on high-availability VM clusters, with a five year hardware refresh cycle. The underlying Merritt repository platform features a granular micro-services architecture and takes advantage of horizontal scaling for load balancing, enhanced performance, and high-availability.

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

Preservation storage is protected by RAID-6 storage arrays. Databases and working file systems are backed up nightly. All datasets in ONEShare are subject to ongoing fixity audit. The underlying Merritt repository platform has not sustained any data loss in over seven years of production operation.

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

OAuth protocol with the Google identity provider.

Data identifier system and data citation policy, if available:

ARK identifiers supplied by CDL's EZID service (<http://ezid.cdlib.org>). Dataset contributors are provided with a formatted citation conforming to DataCite standards at the point of deposit in the Dash portal.

Metadata standards (including provenance):

ONEShare provides an opportunity for data contributors to supply DataCite metadata (<http://schema.datacite.org/meta/kernel-3/>) at the point of deposit. Additional arbitrary metadata is encouraged and can be supplied as file-level components of deposited datasets. The underlying ONEShare repository manages comprehensive provenance and change history tracking.

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?

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

No

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?