

## Member Node Description: Merritt Repository

---

Version 1.1      4/21/2015      Stephen Abrams

### General

<b>Name of resource:</b>	Merritt Repository
<b>URL(s):</b>	<a href="http://merritt.cdlib.org/">http://merritt.cdlib.org/</a>
<b>Institutional affiliation(s):</b>	California Digital Library
<b>Primary geographic location:</b>	Oakland, CA, US
<b>Project Director &amp; contact info:</b>	Stephen Abrams, <a href="mailto:Stephen.Abrams@ucop.edu">Stephen.Abrams@ucop.edu</a> , +1 510-987-0370
<b>Technical Contact &amp; contact info:</b>	Stephen Abrams, <a href="mailto:Stephen.Abrams@ucop.edu">Stephen.Abrams@ucop.edu</a> , +1 510-987-0370
<b>Age of resource:</b>	Since March 2010
<b>Funding support:</b>	University of California
<b>Proposed Unique Identifier:</b>	urn:node:CDL

### Content

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

A cross-domain repository for data and metadata provided by the California Digital Library at the University of California.

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

All types of data are accepted.

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

Public or private.

**Option for embargo (yes/no, duration):**

No

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

16,804 datasets, 50,965 versions, 728,518 files, 2.3 GB

**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)**

The Merritt data model permits deposit of any dataset that can be represented as one or more versions of a file hierarchy. A serialized form of this representation is submitted for deposit, at which time it is assigned a persistent identifier (unless it already has one). Minimal metadata is strongly recommended, but not required. Deposit can occur via the Merritt user interface or a RESTful API. Each Merritt collection carries a curatorially-assigned designation controlling whether or not

submitted datasets are included in the Merritt member node and registered with the DataONE network for aggregated discovery.

**How does a user acquire / access data?**

Any data object, or versions, or files thereof, can be retrieved by users at any time, subject to curatorially-defined access control rules. Datasets assigned DOIs (dependent upon a curatorially-assigned designation at the collection-level) are indexed for aggregated discovery in Scopus and Data Citation index.

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

Generic tools are provided to deposit and access the data. Metadata is indexed for local search.

**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 or DOI identifier resolvable to the dataset landing page.

**Technical characteristics and policies**

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

UC3-authored, portable open source software based on the CDL micro-services architecture, with additional support for the DataONE MetaCat software implementing a Tier 1 member node. Descriptive information about datasets is automatically synchronized with the DataONE network's coordinating nodes. RESTful APIs are available for automated submission, search, and retrieval.

**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 replication. Five copies at three geographic locations are made of each dataset. Data replicas are validated by fixity audit. Databases and working file systems are backed up nightly.

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

LDAP-based credentials and IdM management.

**Data identifier system and data citation policy, if available:**

ARK and DOI 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):**

DataCite, Dublin Core, Dublin Kernel, EML, FGDC.

## 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?