

## Member Node Description: Dryad

---

Version 1.2    1/24/14    T. Vision

### General

<b>Name of resource:</b>	Dryad
<b>URL(s):</b>	<a href="http://datadryad.org">http://datadryad.org</a>
<b>Institutional affiliation(s):</b>	Dryad is a nonprofit corporation. The repository has been developed by a collaboration between the National Evolutionary Synthesis Center at Duke University and the Metadata Research Center at UNC Chapel Hill, along with other grant partners.
<b>Primary geographic location:</b>	North Carolina, USA
<b>Project Director &amp; contact info:</b>	Executive director: Laura Wendell < <a href="mailto:lwendell@datadryad.org">lwendell@datadryad.org</a> >
<b>Technical Contact &amp; contact info:</b>	Repository architect: Ryan Scherle, <a href="mailto:ryan@datadryad.org">ryan@datadryad.org</a>
<b>Age of resource:</b>	Launched in 2009
<b>Funding support:</b>	Organizational memberships, data publication charges, and R&D grants from the National Science Foundation and other organizations.
<b>Proposed Unique Identifier:</b>	urn:node:DRYAD

### Content

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

The Dryad Digital Repository hosts files of data, software source code, and other supplementary materials associated with scientific and medical publications (such as serials, proceedings, edited books, monographs, and theses). Files are typically submitted directly by authors in conjunction with submission of their manuscript to the publisher, but submission of material associated with a previously published work is also permitted. Exclusions include content with legal restrictions on reuse or access (*e.g.* human-subject data).

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

Variable; liberal format acceptability policy.

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

Creative Commons Zero (public domain) with an expectation of community norms for attribution.

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

Yes. Optional embargoes of one year from date of publication are allowed where consistent with journal or publisher policy; longer durations may granted at the discretion of the editor.

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

As of January 2014, Dryad contains over 13,000 data files from over 4,600 individual publications. These are associated with 291 different serial titles and a small number of other publication types. The average data package requires approximately 10 MB of storage.

**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:**

Figures are given here for 2013. The Dryad website received approximately 150K visits from individual users. Counting only the most heavily downloaded file for each data package, the median number of downloads per data package was 43 downloads, with a mean of 70. View and download statistics for individual data packages are available from the Dryad interface and from altmetric providers such as ImpactStory and PlumX. Due to the recent nature of most data submissions to Dryad, citations are not yet being comprehensively tracked, although incomplete statistics for individual data packages are available through the Thomson Reuters Data Citation Index.

## **User interactions**

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

Submission is accomplished through a web interface, typically at the prompting of a journal. Dryad provides online guidance on best practices but is liberal in what it accepts and does not set requirements for how files are prepared. Specific instructions regarding appropriate content and format may be provided by the relevant journal. Curators are available to provide individualized assistance to users.

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

Data may be downloaded from the Dryad website, or searched, harvested and retrieved through various machine interfaces (<http://wiki.datadryad.org/API>).

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

Customized support provided through the website and by email.

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

Upon submission, curators check files for integrity, confirm that they are virus-free, flag inappropriate content, and do some metadata quality control (*e.g.* for the final article citation, the article and data DOIs, and some subject metadata).

## **Technical characteristics and policies**

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

Dryad is built on a customization of DSpace. The API interface includes support for the DataONE API, OAI-PMH, SOLR search, and RSS feeds of new content.

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

Dryad employs production and backup servers with a combined uptime of 99.7%. Since the production server is a virtual machine hosted by the North Carolina State University Library, it is regularly migrated to new hardware by the service provider.

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

Repository contents are incrementally backed up every minute. Full backups are made every 24 hours. The repository service is currently replicated between two sites (NCSU and Duke University) and additional service mirrors are to be added. Content is to be replicated as part of the DataONE network and through membership in CLOCKSS. Should Dryad cease operations at some point in the future, existing content would fallback to these content providers.

To guard against digital file format obsolescence, Dryad is developing the capacity to migrate select proprietary or binary file formats to “preservation formats”. Migration is planned to occur both for individual files upon submission and repository-wide at the discretion of curators for targeted file formats. In all cases, the original bitstreams will be retained in recognition of the potential for information loss during of migration. See [http://wiki.datadryad.org/Preservation\\_Policy](http://wiki.datadryad.org/Preservation_Policy).

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

DSpace user authentication is required for data submission, and Dryad aims to provide support for DataONE and ORCID identity management and authentication. Metadata may be modified by the original submitter, and new versions of data files may be submitted, but published data files may not be deleted. No special access privileges are required for viewing or downloading published contents, with the exception of content under review or time-limited embargo.

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

Dryad assigns DataCite DOIs to both data packages and component data files ([http://wiki.datadryad.org/DOI\\_Usage](http://wiki.datadryad.org/DOI_Usage)). When citing data found in Dryad, users are asked to cite both the original article as well as the Dryad data package. (<http://datadryad.org/using#howCite>).

**Metadata standards (including provenance):**

Dryad’s metadata application profile is based on the Dublin Core Metadata Initiative Abstract Model (DCAM). The most recent version is available from [http://wiki.datadryad.org/Metadata\\_Profile](http://wiki.datadryad.org/Metadata_Profile).

**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

We will support Tier 1 at release, aim to become Tier 2 by the end of 2014, and eventually become Tier 4.

**If you can host data from other member nodes, what storage capacity is available?**

Initially, 200 GB, with this allocation expected to double each year as the size of Dryad’s own holdings increases.

**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?**

Dryad technical staff provide ongoing support to the DataONE infrastructure, both through in-kind contributions and funded projects. Dryad maintains an ongoing research program on barriers and incentives to data sharing and reuse, as well as outreach to scientific societies, publishers and funders regarding data policy. Training capacity is provided through DryadLab, which promotes the reuse of research data in undergraduate and graduate classrooms. Dryad also aims to raise awareness of DataONE resources and activities among its large, and rapidly growing, user base of active researchers.