Developing Preservation Policy for Dryad Digital Repository

Sara Mannheimer, MSIS 2013
Metadata Research Center, Dryad Digital Repository
mannheim@live.unc.edu

Ayoung Yoon, PhD Student
Metadata Research Center, UNC-Chapel Hill
ayyoon@email.unc.edu

Introducing *Dryad Digital Repository*

DataDryad.org is a curated general-purpose repository that makes the data underlying scientific publications discoverable, freely reusable, and citable. Dryad has integrated data submission for a growing list of journals; submission of data from other publications is also welcome.

Dryad is governed by a nonprofit membership organization. Membership is open to any stakeholder organization, including but not limited to journals, scientific societies, publishers, research institutions, libraries, and funding organizations.

PURPOSE OF THE STUDY

The purpose of this research was to develop preservation policy for Dryad Digital Repository.

- The policy aims to guide current and future preservation practice at Dryad and facilitate the long-term preservation of Dryad's digital assets.
- Dryad's mission is to ensure access to its data over the long term in order to facilitate data availability, data sharing, and scholarly communication. A preservation policy and plan are important to this mission.

SIGNIFICANCE

- The aim of the preservation plan is to ensure authenticity, reliability, and integrity of research data over the long term* so that data can be reused for research and education.
- *Our understanding of Long-term (from OAIS reference model): *"a period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community."*

METHOD

1. Review existing preservation practices, models, and policies, then develop an initial draft
2. Consult with a working group of digital preservation experts
3. Revise the draft and structure it to fit Dryad's specific needs

SUMMARY OF DRAFT

Overview of preservation strategy

- Dryad's preservation policy is under the broad guidance of the Open Archival Information System (OAIS) reference model (ISO 14721:2012).
- Dryad will make preservation decisions to balance the constraints of costs, level of authenticity, and legal requirements, taking a three-pronged approach: bit-level preservation, collection of adequate description to facilitate preservation, and file-meaning preservation.

Format support and levels of preservation

- Non-proprietary, openly documented formats are always preferred.
- Based on the assessment of future usability of files, different level of support will be provided (full support, limited support, or as-is Bitstream Access). Format support levels will be continually reevaluated.

Implementation strategy

- Both short-term and long term plans have been identified for ingest, curation, and archival storage.
- Examples from short-term plan: evaluate migration tools for non-preferred formats, test automated file inspection, replicate using CLOCKSS, etc.
- Examples from long-term plan: implement automated file inspection, adopt PREMIS or other standard metadata, establish additional service mirrors, conduct mass transformations of outmoded formats, etc.

Data integrity, security, and versioning

- Fixity checks, MD5 checksum and nightly integrity checks
- Ensure the physical safety and security of data, monitor IT infrastructure
- Display a version history that allows users to navigate to earlier versions

Sustainability plans

- Technical sustainability: DSpace and DuraSpace
- A business model that covers the repository's operating costs

IMPLICATIONS

- Protect the repository's scientific data over the long term
- Serve as a blueprint for other repositories facing similar challenges