

Preservation Guidelines *DRAFT*

Purpose: To outline best practices that will ensure long-term access to the content of the University of Notre Dame's Institutional Digital Repository (IDR).

Scope: Digital objects collected and preserved in the IDR. At present, these include approved Notre Dame Electronic Theses and Dissertations in PDF format, PDF files of student and faculty papers, and digital objects created from the Library's collections, i.e. the Art Image Library.

Recommendations:

External partners:

- Monitor DSpace's preservation practices. DSpace currently provides bit level preservation for supported, known, and unsupported formats. Functional preservation is provided for supported formats which include TIFF, SGML, XML, AIFF, and PDF. See <http://www.dspace.org/faqs/index.html#preserve> for preservation level definitions.
- Monitor ProQuest's viability. The University contracts with ProQuest to manage its Electronic Theses and Dissertations (ETDs). ProQuest has a contract with the Library of Congress that states that LC will assume responsibility for the ETDs in the event that ProQuest can no longer provide access.

Documentation:

- Maintain access to the complete documentation used to create the digital content in the IDR. It is not feasible to sustain the content over the long term without information about how the files are encoded.
- Advocate the use of non-proprietary and open standards which make it easier to access and maintain full documentation. For proprietary tools, advocate for the maintenance of full documentation in an independent, trusted archive to facilitate the preservation of digital information in proprietary formats.

Formats

- Use widely adopted formats which are less likely to become obsolete and for which emulation and migration tools are more likely to emerge from industry.
- Use formats in which the underlying information is simply and directly represented. Eliminate use of encryption and compression for preservation copies. Their use for transmission is expected to be routine.
- Consider that the use of formats that are dependent on particular hardware, operating system, or software will increase preservation costs. Sustaining dynamic content with such dependencies is more difficult than sustaining static content.
- Use formats that are not bound to a particular physical carrier or to the use of a particular device. Such formats are not suitable for preservation purposes.

Metadata

- Use metadata to identify and describe the content, to document the creation of the digital file, and to provide technical details to support the transition to future technical environments. Embedded metadata that does not impact the content is preferred for preservation purposes.
- Consider the use of preservation metadata at the beginning of the object's life cycle. See <http://www.oclc.org/research/projects/pmwg/> for detailed information on preservation metadata.

The entries listed under *Documentation*, *Formats*, and *Metadata* are adapted, with permission, from the National Digital Information Infrastructure and Preservation Program's (NDIIPP) recommendations for preserving digital objects. (<http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>). NDIIPP provides excellent preservation guidelines which are suitable for the IDR. Following such recognized standards and guidelines will enable us to provide long-term access to the digital objects collected in the IDR.

Note: The guidelines should be viewed as a "work in progress" since they may require modification as the technology changes. And, at some future point, we will need to revisit the feasibility of a dark archive.

Selected Further Reading:

DSpace. MIT Libraries. <http://libraries.mit.edu/dspace-mit/>

Digital Preservation: The National Digital Information Infrastructure and Preservation Program. The Library of Congress. <http://www.digitalpreservation.gov/index.html>

Digital Preservation Policy, State Library of Victoria. Last updated January 31, 2006. www.slv.vic.gov.au/about/information/policies/digitalpreservation.html

Digital Preservation Policy Tool. Electronic Resource Preservation and Access Network, September 2003. www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf

Final Report of the PREMIS Working Group, May 2005. Preservation Metadata Implementation Strategies (PREMIS). A working group jointly sponsored by OCLC and RLG. www.oclc.org/research/projects/pmwg

Gladney, H.M., *Perspectives on Trustworthy Information*. Digital Document Quarterly, v. 5, no. 2. home.pacbell.net/hgladney/ddq_5_2.htm#repositories

Kansas State University, *Policy on Electronic Theses, Dissertations, and Reports, Chapter 3465*, issued April 25, 2006. www.k-state.edu/policies/ppm/3465.html

LeFurgy, William G., *PDF/A: Developing a File Format for Long-Term Preservation*. RLG DigiNews: v.7, no.6. www.rlg.org/preserv/diginews/v7_n6_feature1.html

METS: Metadata Encoding & Transmission Standard. www.loc.gov/standards/mets

Why LaTeX? University of Cambridge, Department of Engineering.
www-h.eng.cam.ac.uk/help/tpl/textprocessing/latex_advocacy.html