

The NDLTD and Issues of Long Term Preservation and Archiving: IT'S ABOUT TIME!

by Gail McMillan

*for Next Steps: Electronic Theses and Dissertations Worldwide
ETD 2003, Humboldt University, Berlin, Germany, May 22, 2003*

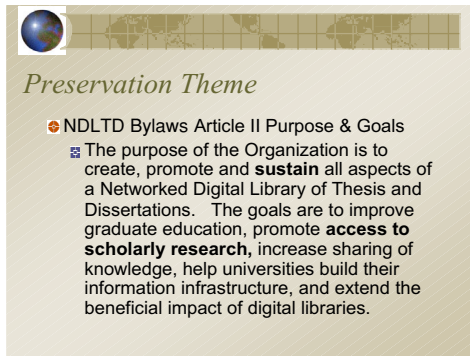
I met a traveller from an antique land
Who said:—Two vast and trunkless legs of stone
Stand in the desert. Near them on the sand,
Half sunk, a shatter'd visage lies, whose frown
And wrinkled lip and sneer of cold command
Tell that its sculptor well those passions read
Which yet survive, stamp'd on these lifeless things,
The hand that mock'd them and the heart that fed.
And on the pedestal these words appear:
 "My name is Ozymandias, king of kings:
 Look on my works, ye mighty, and despair!"
Nothing beside remains: round the decay
Of that colossal wreck, boundless and bare,
The lone and level sands stretch far away.
Percy Bysshe Shelley¹

Will ETDs suffer the fate of Ozymandias? My goal today is to convince you that it is time for the NDLTD to take specific actions to promote the long-term preservation of ETDs.



The Networked Digital Library of Theses and Dissertations is maturing and becoming a more formal organization. Evidence of this maturation includes documentation such as the Strategic Plan, which states that the mission of the NDLTD is to improve graduate education by developing *accessible digital libraries*, to increase the *availability* of research, and to *preserve, preserve!* it electronically.

¹ Thanks to Henry Gladney for using this poem in “Digital Preservation Archiving and Copyright: Problem Description and Legislative Proposal.”



Preservation Theme

- ⊕ NDLTD Bylaws Article II Purpose & Goals
 - ⊞ The purpose of the Organization is to create, promote and **sustain** all aspects of a Networked Digital Library of Thesis and Dissertations. The goals are to improve graduate education, promote **access to scholarly research**, increase sharing of knowledge, help universities build their information infrastructure, and extend the beneficial impact of digital libraries.

Our draft bylaws say that the purpose of the Organization is to create, promote and sustain, **sustain!** all aspects of a Networked Digital Library of Thesis and Dissertations.



Preservation Theme

- ⊕ ETD 2003 Berlin Conference
 - ⊞ NDLTD's activities are focused on universities... to support... archiving...
 - ⊞ NDLTD's vision is to increase the availability of student research documents for scholars and to preserve it electronically.

The 6th conference web site also calls attention to preservation, specifically mentioning focusing on **archiving** and retrieval of electronic theses and dissertations worldwide as well as increasing the availability of student research by **preserving** it electronically.

In each of these and in other forms there is a persistent theme: the importance of preservation and sustaining access to ETDs. But, in spite of this repeated theme and because there has been little activity resulting from mere words, I assert that it is time for the NDLTD to live up to its name, the Networked Digital **LIBRARY** of Theses and Dissertations, as well as its mission.

Archiving and preservation are closely aligned and the finer distinctions don't necessarily need to be dwelt upon.

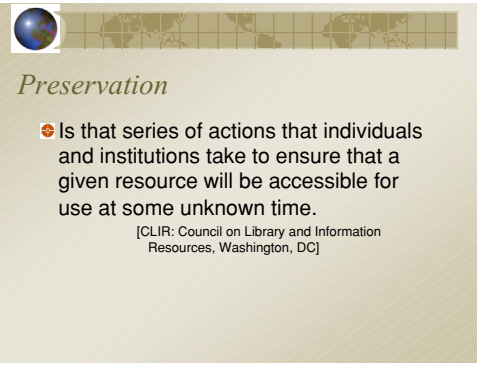
A slide titled "Archiving" with a globe icon and a world map background. The text defines archiving as the identification, collection, and preservation of access to digital records in their current form. It cites PADI: Preserving Access to Digital Information, National Library of Australia.

Archiving

✚ Involves the identification, collection, and preservation of access to digital records in their current form.

[PADI: Preserving Access to Digital Information, National Library of Australia]

Today, NDLTD member institutions identify, collect, and provide access to their ETDs. But what if one or more ETD servers go down, are compromised, or simply cannot handle the access load of query and retrieval from worldwide sources? Guaranteeing access means making servers more robust and providing alternate access—possibly through distributed, collaborative, persistent mirroring. More copies available means more patrons can be timely served.

A slide titled "Preservation" with a globe icon and a world map background. The text defines preservation as a series of actions that individuals and institutions take to ensure that a given resource will be accessible for use at some unknown time. It cites CLIR: Council on Library and Information Resources, Washington, DC.

Preservation

✚ Is that series of actions that individuals and institutions take to ensure that a given resource will be accessible for use at some unknown time.

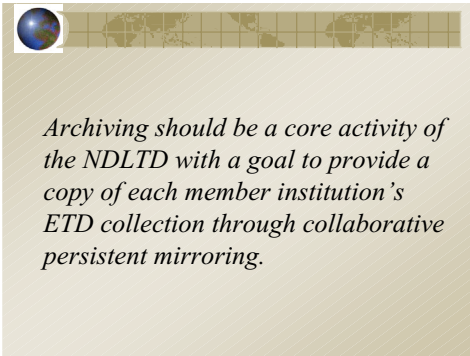
[CLIR: Council on Library and Information Resources, Washington, DC]

Face facts: software rots. While the data of PDF, Word, or XML can withstand repeated, perfect copying to new media, the applications that render these documents are continuously improved, as are the operating systems on which these applications run. PDF has gone from version 1.1 to 1.4 since the NDLTD began; Acrobat Reader is now version 6.0; and XML version 1.1 is now a W3C candidate recommendation. The NDLTD must continually refresh its archiving, indexing, and rendering software in addition to moving its collection of media. Maintaining server hardware, operating systems, and application software licenses is an ongoing process.

In my limited time today I am not attempting to address all the ramifications of the NDLTD archive. My goal is to convince you and the membership at large to voice their support for NDLTD archiving to more fully meet the obligations of a research library of academic proportions, including a responsible preservation.

It is time to examine what we're doing, what in good conscience we are recommending that others do considering what will be the past as well as the future. The collective NDLTD should no longer perpetuate just short-sited goals and foster the notion that 'how to do ETDs' is enough.

Our organization is more than five years old. (In Web years, like dog years that is quite old.) It is, therefore, time to act responsibly, like mature adults. That is, it is about time to actively foster the preservation—the storage and access of ETDs.



Archiving should be a core activity of the NDLTD with a goal to provide a copy of each member institution's ETD collection through collaborative persistent mirroring. Some may think of archiving as a passive activity, developing the way dust bunnies accumulate while housework is avoided as one drafts a dissertation. But, archiving is not a passive activity; it takes conscious and continuing effort. I suggest that the NDLTD take steps immediately to establish archiving and preservation goals that will (1) strategically serve every member of the organization and (2) place it among the international leaders in this area. Let's truly make the NDLTD a model digital library by not only moving graduate student scholarship onto the easily accessible Internet. Let's keep it there by tackling the really tough issues like archiving and preservation and by being here it demonstrates that we all feel that we have a stake in the NDLTD and in access to ETDs as do many others who are not here. According to the recent Council on Library and Information Resources publication, *New-Model Scholarship; How Will It Survive?* long-term preservation of digital library objects favors disaggregating functions so that each institution may, but does not have to, have its own repository. [Smith] Libraries do not operate alone and monolithic. Therefore, the Networked Digital **LIBRARY** of Theses and Dissertations should consider its, our, responsibilities for ensuring the longevity of these important information resources not just within our individual institutions but also across institutions through the NDLTD. I call for the NDLTD to accept the preservation and access responsibilities collaboratively, not just individually.

Over 150 institutions of higher education have so far come together to form this organization, looking to each other to learn about careful and responsible management of electronic theses and dissertations. In order for our organization to be beneficial to the full compliment of member institutions (including over 20 supporting organizations), the NDLTD must become a more vigorous organization that appeals equally to those new to as well as those with well established ETD initiatives.

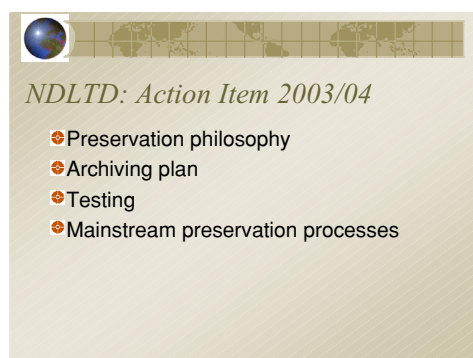
The NDLTD must also continue to have a membership base that includes all institutions of higher education who have as a goal ETDs whether institutional financial support of the NDTLD is available or not. We also do not want to disenfranchise any university community knowing

that digital divide issues such as type of connection can be a more important factor than other digital divide demographics such as education, race, or gender. [Davison] The NDLTD will become a stronger and more vibrant organization if the haves help those who have less. The haves with more expertise, more equipment, more staff, more technology, more money; whatever. Those who have less of any of these are still at institutions where graduate students are producing ETDs that contain information that is desired by someone outside their institution and the NDLTD should feel obligated to help preserve every member's ETDs.



The NDLTD strives to be a model digital library with the first concerted initiative providing free access to graduate student research and scholarship. The dramatic shift that many institutions have made, replacing physical media with digital sources that are sometimes the only forms of theses and dissertations, has not been accompanied by sufficient action to ensure that these works will survive transitory interest. To steward ETDs, the NDLTD must also provide access through long-term preservation. The NDLTD cannot be a model digital library if it does not, among other things, develop and enact a sound preservation strategy.

Within higher education we are teaching our students to document scholarship through digital media while at the same time we acknowledge the challenges of electronic preservation, even if only from one version to the next. We are aware of the potentially short life of any file format, but we passively await the answers to preservation of digital resources. Let's stop pretending to be naïve.



Let's begin by acknowledging and accepting the challenges to properly handle the works we are causing to be created and actively work toward finding the path to cost-effective use of

appropriate technologies to create a responsible NDLTD that has active digital archiving and historic preservation. Let's out digital preservation, moving concertedly from a philosophy to a plan to a project to a mainstream preservation activity.

ETDs are models of digital scholarship and the NDLTD should be a model organization. We should develop the system-wide infrastructure to support common digital preservation needs, from metadata standards to linking services to persistent identifiers, and beyond. Each member customizes delivery to best meet their principle users' needs. We should not be prescriptive and avoid attempting to normalize ETDs as we have not been prescriptive about their content and scholarship. These are decisions most appropriately made by each individual member institution.

As it formalizes, the NDLTD needs to specify the scope of its activities, both functions and services, so that it can begin to adopt enabling agreements about roles and responsibilities pertinent to preservation. Along these lines, the NDLTD needs to establish a digital preservation infrastructure and could begin by considering these categories and policy issues, among others.

Technology Policy – Single Item	Technology Policy – Archive	Organizational Policy
Format registry specifying documentation and best practices	Persistent identifiers	Independent legal entity empowered to manage the archive if something happens
Learn from working with other formats to preserve the content, not necessarily the software application	Interoperability for searching, storage, and access	Access warranties without public access
Obtain sufficient control of the information to ensure long-term preservation (OAIS-type archive Lavoie)	Create low-cost, persistent digital caches of authoritative versions of content (e.g., LOCKSS open-sources software)	Institutional deposits vs. web-crawling capture of ETDs to address the preservation of restricted and embargoed works along with being password protected and other secured access mechanisms
Accept whatever its members create and call ETDs. <ul style="list-style-type: none"> Some institutions will have data format standards Suggest file formats for easier migration and management (e.g., Adobe) 	Initially provide various levels of service (e.g., Harvard): <ul style="list-style-type: none"> Render normative formats Keep bits in order but not necessarily render files (e.g., DSpace) Other complex file formats 	Make the information available to the designated user community (OAIS; Lavoie)
Metadata for description, reuse, administration, and preservation	Open-source turnkey system for digital asset management (DSpace)	Document policies and procedures to ensure the information is preserved against reasonable contingencies, and to enable the information to be disseminated as authenticated copies of the original or as traceable to the original (OAIS-type archive Lavoie)
	Balance the completeness and currency of the archive and the burden on the system resources	Actively encourage faculty to become partners in preservation to educate student authors to consider preservation a component of creating ETDs
	Internal-only links will be archived	

It is time that the NDLTD make the commitment to active, not passive, preservation. It is necessary for the organization to make this decision before beginning the discussions about cost and technology that are better left to those topic experts. As the organization becomes more formal, specific services should develop through institutional support and these may even include revenue-generating services, but always with the goal to include among its membership as many institutions as want to be part of the NDLTD, financial support aside.



It's about time that the NDLTD

- ✦ Learn from its members' archiving activities
 - ▣ Cal Tech, MIT, Adobe
- ✦ Participate in digital preservation initiatives
 - ▣ Harvard, LOCKSS, OCLC

We need to be part of the scalable and sustainable digital library endeavors demonstrating responsible behaviors. The United States Library of Congress recently released its report on “Preserving Our Digital Heritage.” It says higher education and research enterprises should assume preservation responsibilities because protecting institutional assets for the future is central to furthering the creation of knowledge, a the core value of the academic enterprise in higher education. [CLIR March 2003 p.20] If the membership decides that the NDLTD should formally and actively pursue archiving strategies, it should cooperate with national and international organizations and existing e-archiving and preservation initiatives many of which are among the NDLTD members.



It's about time that the NDLTD

- ✦ accept its mission and establish a preservation plan
- ✦ develop archiving that crosses the digital divide
- ✦ recognize responsible archiving is not a passive activity
- ✦ members collaborate to provide archiving services for each other

In the NDLTD, in this library, we have developed a trusted public institution and by preserving library resources, we are protecting free access to information now and in the future. We have all joined the NDLTD to be part of an organization because of our shared goals. Some members may feel that it just happens that this organization is called a library, but it was a conscious decision by the director, and the founding members accepted this moniker and what it means.

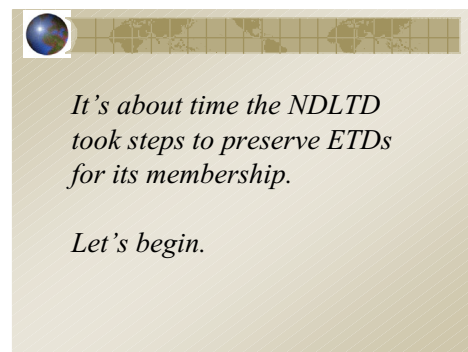
We have created a public academic library that includes access but without preservation there will be nothing to access. We cannot pretend that it is not our responsibility, the NDLTD's responsibility, to provide continuing access to ETDs when we have for many years encouraged the world's institutions of higher learning to allow or to require their graduate students to create them. I believe that the NDLTD has a moral imperative to determine and instigate the preservation strategies for archiving data and primary sources that underlie these scientific, technological, and scholarly advances.



It's about time that the NDLTD

- investigate the path to effective digital archiving
- develop the system-wide infrastructure to support common digital preservation needs for its members
- address the full range of archiving issues, including

If the NDLTD membership decides not to collaborate to assume its intrinsic library responsibility for archiving collections for preservation and access, then it should at least develop a philosophy for library collections and subsequently align itself a trusted repository that will do this job for us. But we cannot ignore this responsibility. We must act now on our own advice; the NDLTD must accept its curatorial and preservation responsibilities or suffer the fate of Ozymandias.



*It's about time the NDLTD
took steps to preserve ETDs
for its membership.*

Let's begin.

REFERENCES

- American Library Association Preservation Policy, rev. 2001. ALCTS Publishing.
http://www.ala.org/Content/ContentGro7ups/ACLTS1/Publications10/WWW_Publications/Preservation3/ALA_Preservation_Policy/ALA_Preservation_Policy.htm [May 9, 2003]
- [ARL] Association of Research Libraries. "The Responsibility of Research Libraries for Preservation." May 22, 2002 <http://www.arl.org/preserv/responsibility.html> [May 9, 2003]
- Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources*. Mountain View, CA: RLG, Aug. 2001
- Berthon, Hilary, Susan Thomas, Colin Webb. "Safekeeping: A Cooperative Approach to Building a Digital Preservation Resource." *D-Lib Magazine*, vol. 8, no. 1, Jan. 2002.
<http://www.dlib.org/dlib/january02/berthon/o1berthon.html>
- Building a national Strategy for Digital Preservation: Issues in Digital Media Archiving*. Washington, DC: Council on Library and Information Resources and the Library of Congress, April 2002.
- Davison, Elizabeth, and Shelia R. Cotton. "Connection Discrepancies: Unmasking further Layers of the Digital Divide." *First Monday* vol. 8, no. 3, March 3, 2003.
http://www.firstmonday.org/issues/issue8_3/davison/ [March 7, 2003]
- Digital Preservation Coalition. 2. Digital Preservation 2.2 Preservation Issues 2.2.2 Organisational Issues.
<http://www.dpconline.org/graphics/digpres/presissues.html> [May 9, 2003]
- Flecker, Dale. "Digital Archiving What Is Involved." *EDUCAUSE Review* _____: EDUCAUSE, Jan./Feb. 2003, 10-11. <http://www.educause.edu/ir/library/pdf/ERM0316.pdf>
- Gladney, H. M. "Trustworthy and Durable Digital Documents: Overview."
<http://home.pacbell.net/hgladney/tdo.pdf> [May 9, 2003]
- Gladney, H. M. "Digital Preservation Archiving and Copyright: Problem Description and Legislative Proposal" (ver. 2, 15th August 2000) <http://www.ndltd.org/ndltd-sc/20000915/ArchCopy.pdf> [May 14, 2003]
- Hodge, Gail. "Best Practices for Digital Archiving: An Information Life Cycle Approach," *D-Lib Magazine*, vol. 6, no. 1, Jan. 2000 <http://www.dlib.org/dlib/january00/01hodge.html>
- JSTOR News, March 2003, no. 7, issue 1.
- Keller, Michael A., Victoria A. Reich, and Andrew C. Herkovic. "What is a library anymore, anyway?" *First Monday* http://www.firstmonday.org/issues/issue8_5/keller/ [May 9, 2003]
- Kenney, Anne R., and Oya Y. Rieger. "Preserving Digital Assets: Cornell's Digital Image Collection Project" *First Monday* http://www.firstmonday.org/issues/issue5_6/kenney/ [May 9, 2003]
- Lavoie, Brian. "OAIS: Open Archival Information System." *OCLC Newsletter*, Jan./Feb. 2003, p. 26-30.
<http://www.oclc.org/research/publications/newsletter/repubs/lavoie243/> [May 13, 2002]
- LOCKSS <http://lockss.standord.edu/> [May 9, 2003]

- Martin, Julia, and David Coleman. "Change the Metaphor: The Archive as an Ecosystem." *Journal of Electronic Publishing*. 7:3 (April 2002) <http://www.pres.umich.edu/jep/07-03/martin.html> [May 9, 2003]
- Mellor, Phil. "CAMiLEON: Emulation and BBC Domesday." RLG DigiNews April 15, 2003, vol. 7, no. 2 <http://www.rlg.org/preserv/diginews/diginews7-2.html#feature3> [May 13,2002]
- Ockerbloom, John Mark. "Archiving and Preserving PDF Files." *RLG DigiNews* vol. 5, no. 1, Feb. 15, 2001 <http://www.rlg.org/preserv/diginews/diginews5-1.html> [May 9, 2003]
- PADI: Preserving Access to Digital Information. National Library of Australia. "Digital Preservation Strategies," including "Roles and Responsibilities" and "Archiving." <http://www.nla.gov.au/padi/topics/> [May 9, 2003]
- Palgrave, Francis T., ed. *The Golden Treasury of the Best Songs and Lyrical Poems in the English Language*. London: Macmillan, 1875. <http://www.bartleby.com/106/246.html> [May 12, 2003]
- Shelley, Percy Bysshe. *Ozymandias*. Henry Gladney used this poem in "Digital Preservation Archiving and Copyright: Problem Description and Legislative Proposal." (ver. 2, 15th August 2000) <http://www.ndltd.org/ndltd-sc/20000915/ArchCopy.pdf> [May 14, 2003]
- Smith, Abby. "Cooperating to Build Collections in the Digital Era." *CLIR Issues*, No. 31, Jan./Feb. 2003, [1], 5. Washington, DC: Council on Library and Information Resources. <http://www.clir.org/pubs/issues/issues31.html#build> [verified May 7, 2003]
- Smith, Abby. "Issues in Sustainability: Creating Value for Online Users." *First Monday* vol. 8, no. 5 (May 2003) http://firstmonday.org/issues/issue8_5/smith/ [May 9, 2003]
- Smith, Abby. *New-Model Scholarship: How Will It Survive?* Washington, DC: Council on Library and Information Resources, 2003. www.clir.org/pubs/reports/pub114/forward.html [May 9, 2003]
- Smith, MacKenzie, et al. "An Open Source Dynamic Digital Repository." *D-Lib Magazine*, vol. 9, no. 1, Jan. 2003. <http://www.dlib.org/dlib/january03/smith/01smith.html> [March 21, 2003]
- Spindler, Rob. "Preserving Interactive or Multi-version Web-based Documents. 3rd International Symposium on ETDs. March 2000
- The State of Digital Preservation: An International Perspective*. Washington, DC: Council on Library and Information Resources, July 2002.
- Waters, Donald J., et al. "The Digital Preservation Consortium Mission and Goals." Commission on Preservation and Access. <http://www.clir.org/pubs/reports/dpcmiss/dpcmiss.html> [March 17, 2003]



Joan Lippincot, Washington, DC; Ana Pavani, Brazil; Gail McMillan, Virginia; Shalini Urs, India
at

Next Steps: Electronic Theses and Dissertations Worldwide
ETD 2003, Humboldt University, Berlin, Germany, May 21, 2003

Thanks to Ed Fox for the digital image!