

The NDLTD and Issues of Long Term Preservation and Archiving¹ IT'S ABOUT TIME!

Gail McMillan

Digital Library and Archives, Virginia Tech

gailmac@vt.edu

P.O. Box 90001, University Libraries, Blacksburg, VA, 24062-9001

<http://scholar.lib.vt.edu/>

Keywords: archive, preservation, NDLTD, ETD

Abstract

It's about time the NDLTD did something really significant for the membership. Of course, this is to say it's about time we did something for ourselves such as collaborating to ensure the long-term preservation and archiving of each other's ETDs. This presentation will outline the broad considerations for archiving ETDs as well as the necessary actions that the NDLTD Executive and Steering Committees should initiate on our behalf. Audience participation will also be a component of this presentation so that conference participants can tell the organization what they think it should do for them.

The NDLTD and Issues of Long Term Preservation and Archiving

The Networked Digital Library of Theses and Dissertations is maturing and becoming a more formal organization as evidenced by drafts for a strategic plan (May 2002) and bylaws (March 14, 2003). A theme throughout these and other documents such as the ETDs2003 Berlin conference web site (<http://www.hu-berlin.de/ETD2003/>) is preservation and sustaining access. For example, according to the Strategic Plan, the mission of the NDLTD is to improve graduate education by developing *accessible digital libraries* of theses and dissertations. To fulfill this mission one of the specific objectives is to increase the *availability* of student research for scholars and to *preserve* it electronically. ETDs2003, the conference web site reiterates:

NDLTD's activities are focused on universities, libraries, faculties and graduate students in order to support authoring, indexing, archiving, dissemination and retrieval of electronic theses and dissertations worldwide.

NDLTD's vision is to increase the availability of student research documents for scholars and to preserve it electronically and to empower students to convey a richer message through the use of multimedia and hypermedia technologies.

Because of these pervasive themes and because there has been little activity resulting from these words, I, therefore, assert that it is time for the NDLTD to fully live up to its obligations as well as its name.

Archiving should be a core mission of the NDLTD with a goal to provide a copy of each member institution's ETD collection through collaborative persistent mirroring. From the intellectual property point of view this should not pose a problem, especially if members followed the VT model that has the explicit agreement authors that they grant permission for their institutions to store and provide access. Where the institution stores ETDs is up to it to decide and may include redundant copies stored by the home institution and, by contract, with another entity such as the NDLTD.

One hundred fifty two institutions of higher education have so far banded together to form the NDLTD, 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 it 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 NDLTD 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 may have more expertise, more equipment, more staff, more technology, more money, etc. 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 is obligated to help preserve every member's ETDs.

¹ This paper uses terms such as archiving and long-term preservation interchangeably because it is intended for a broad audience that may not be familiar with the fine distinctions. "Archiving involves the identification, collection and provision of access to digital records in their current format..." [PADI] "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 March 2003 p.2]

The NDLTD strives to be a model digital library with the first concerted initiative providing free access to graduate student research and scholarship. Currently it has a membership list with links to each ETD initiative and/or collection and a rudimentary federated search engine that will undoubtedly improve with time. But, "a web page with a set of links is not a library." [Keller] The dramatic shift that many institutions have made from physical media to 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 a sound preservation strategy.

"The question we all face in digital preservation is addressing the need to adapt the process of creation without sacrificing the beauty of creative endeavor. We will reach this balance only when we have established effective partnerships between creators of documents, and those charged with stewardship of these products." [Spindler, p. 9]

In the limited space allotted for this paper I have not attempted to address all the ramifications of the NDLTD archive. The goal of this paper is to persuade the membership and the conference participants to voice their support of NDLTD archiving to more fully meet the obligations of a research library of academic proportions, including a responsible archiving.

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 dog years that would be more than 21!;>) It is, therefore, time to act responsibly, like mature adults. That is, it is about time to actively foster the long-term preservation - the storage and access of ETDs.

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 propose that the NDLTD take steps immediately to establish archiving and long-term 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.

The mere fact that we are here in Berlin and the NDLTD is holding at its sixth conference demonstrates that we have a stake in the NDLTD and in access to ETDs as do many others who are not here. In the long run, the loss of the source data contained in ETDs may

be as serious a problem as the inability to interpret the data in secondary literature. 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, let us consider 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.

Within higher education we are teaching our students to document scholarship through electronic media and we simultaneously acknowledge the challenges of digital 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 kidding ourselves and acting like our most naïve students. 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 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. Each member customizes delivery to best meet their principle users' needs. We should not be prescriptive and avoid attempting to normalize ETDs (e.g., what technology they take advantage of) 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 identify 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, among others, the following:

- Format registry specifying documentation and best practices
- Interoperability for searching, storage, and access
- 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
- Create low-cost, persistent digital caches of authoritative versions of content (e.g., LOCKSS open-sources software)

- OAIS-type archive (Lavoie):
 - Obtain sufficient control of the information to ensure long-term preservation
 - 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
 - Make the information available to the designated user community
- Access warranties without public access
 - The response to a search could be, "Yes, it's in the NDLTD but you have to go to the originating institution first to get it. If they can't find their copy, however, the NDLTD will provide one."
- Accept whatever its members create and call ETDs
 - Some institutions will have data format standards
 - Suggest file formats whose migration and management are the easiest (especially with active commercial partners like Adobe and UMI)
- Initially provide various levels of service (e.g., Harvard):
 1. Render normative formats
 2. Keep bits in order but not necessarily render files (e.g., DSpace)
 3. Other complex file formats
- 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
- Internal-only links will be archived
- Balance the completeness and currency of the archive and the burden on the SYSTEM resources
- Metadata for description, reuse, administration, and preservation
- Actively encourage faculty to become partners in preservation to educate graduate student authors so that they consider preservation as a component of creating ETDs
- Persistent identifiers
- Open-source turnkey SYSTEM for digital asset management (DSpace)

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 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.

What are the good and obvious reasons for the NDLTD to provide long-term preservation and archiving services to its members?

- Collections are worthy of dissemination and continued access in combination with as well as separate from their originating institutions.
- It is part of the organizations' mission.
- It would add to the prestige of the organization.

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. We need to be part of the scalable and sustainable digital library endeavors demonstrating responsible behaviors. In the recently released "Plan for the National Digital Information Infrastructure and Preservation Program," the Library of Congress proposes that sectors such as higher education and academic and research enterprises take primary responsibility for collecting, curating, and ensuring the preservation of their information assets as certified trusted repositories in a networked environment. This report on "Preserving Our Digital Heritage" says higher education and research enterprises should assume preservation responsibilities because protecting its 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] There are several NDLTD members who have noteworthy e-archiving initiatives, including California Institute of Technology and the University of Virginia, as well as Harvard. Our Australian collaborators in the NDLTD are also very active through PADI in defining all aspects of digital archiving and long-term preservation.

Unlike some of these institutions, however, the NDLTD does not have to make the difficult choice of deciding what not to keep because the collection development policies are in place at each member institution when the Graduate School (or its counterpart) approves an ETD. These works have already been determined to be acceptable and of enduring value to the institution and, therefore, to the NDLTD. Similarly, some acceptable formats have already been determined because the vast majority of ETDs are submitted as PDF files. We are fortunate that Adobe is a partner with the NDLTD and is already well known for allowing third-parties to develop PDF products. [Ockerbloom] We should take advantage of this collaborative attitude as we develop the preservation goals of the NDLTD.

The members of the Networked Digital Library of Theses and Dissertations are research institutions of higher learning and all are represented by their libraries active participation in the organization and/or in their institutional ETD initiatives.

"...preservation as one of the fundamental responsibilities of the research library community... A principal goal of research libraries is to preserve the record of knowledge from the past and present ... and to carry out this basic responsibility into the future. To invest in protection of this

asset is a wise and responsible act to insure against loss." [ARL]

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. "Archiving is not a disconnected activity... If the user community does not trust the competence, values, and viability of the archive, the necessary social support for the archiving activity may be missing." [CLIR March 2003, p. 33] Is this the case with the NDLTD?

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. "The NDLTD leadership was interested in positioning the initiative in the burgeoning context of digital library and institutional repository development." [NDLTD Strategic Plan]

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.

If the NDLTD decides not collaborate to assume its inherent library responsibility for archiving its collections for long-term preservation and access, then it should develop a philosophy for library collections and subsequently align itself with an organization (currently referred to as "trusted third-party repositories") that will do this job for us. But we cannot ignore this responsibility; we commissioned a strategic plan that advises the organization to clearly address archiving concerns and propose strategies for preservation of ETDs. We must act now on our own advice; the NDLTD must accept its curatorial and preservation responsibilities or suffer the fate of Ozymandias.

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 Shelley

Bibliography

- 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.dpconine.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/ttdo.pdf> [May 9, 2003]
- Gladney, H. M. "Digital Preservation Archiving and Copyright: Problem Description and Legislative Proposal" (ver. 2, 15th August 2000)
- 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/re-pubs/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]
- 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 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]