
Market Analysis

NDIIPP MetaArchive Project



2008-03-08

Summary

The MetaArchive Cooperative is a community-based distributed digital preservation network that provides low-cost, high-impact preservation services to universities, libraries, archives, and other cultural heritage repositories. This market analysis concentrates on the MetaArchive Cooperative's primary markets: University libraries and archives. A separate analysis will be conducted later this year on cultural memory repositories that include museums, state and federal government agencies, and historical centers.

About the Cooperative

The MetaArchive Cooperative is an independent, international membership association that provides collaborative and distributed digital preservation solutions for cultural memory organizations, including university libraries, government and historical research archives, museums, and other similar institutions (<http://www.metaarchive.org>). The central missions of the MetaArchive Cooperative are to support, promote, and extend the practice of distributed digital preservation; to serve as a catalyst and guide for other networks that seek to implement the distributed digital preservation methods it has developed; and to educate cultural memory organizations about distributed digital preservation.

The Market

There are roughly 1,600 universities that contain libraries and archives in the United States alone.¹ The vast majority of these universities, libraries, and archives are currently producing and collecting "born digital" and digitized collections of historical, cultural, and political value. As documented by surveys of these communities, very few have solutions to the pressing problem of preserving these digital collections, but many of them are actively seeking such solutions.² To date, more than four dozen unsolicited inquiries have been made by institutions that are interested in the MetaArchive Cooperative's distributed digital preservation method and services. This constitutes a market segment with an overwhelming need that is currently unmet by other products.

The Proposed Solution

The MetaArchive Cooperative can meet the needs of these communities by 1) hosting content, 2) building networks with member organizations to host collections, and 3) providing the technical and organizational model and tools needed to encourage others to establish their own distributed networks.

¹ See "U.S. Universities" at <http://www.utexas.edu/world/univ/alpha/>; "Museums FAQ" at <http://www.aamus.org/aboutmuseums/abc.cfm>; "National Archives Locations by State" <http://www.archives.gov/locations/states.html> (accessed 1 Dec 2007)

² Pardo, Theresa A., G. Brian Burke, and Hyuckbin Kwon, "Preserving State Government Digital Information: A Baseline Report," (July 2006). Available at http://www.ctg.albany.edu/publications/reports/digital_preservation_baseline/ (accessed 1 Dec 2007); Clareson, Tom. "NEDCC Survey and Colloquium Explore Digitization and Digital Preservation Policies and Practices" RLG DigiNews, 10:1 (February 2006). Available at: http://www.rlg.org/en/page.php?Page_ID=20894#article1 (accessed 3 Dec 2007).

The Data

Data used in this analysis comes from two survey-based sources: 1) two NEDCC/NEH-sponsored “Stewardship of Digital Assets” Workshop surveys (84 participants) and 2) the initial responses to a MetaArchive ETD preservation survey (75 Research Libraries: see Appendix A). Data also comes from unsolicited inquiries into MetaArchive membership that have been made by institutions over the last 18 months

SODA Survey Results

The Stewardship of Digital Assets workshop, hosted by the Northeast Document Conservation Center (NEDCC), teaches attendees how to plan for the preservation of their digital assets. Prior to attending, all registrants are asked to fill out a 66-question survey regarding the current and planned digitization activities of their organizations. To date, two workshops have been held and there have been a total of 84 survey respondents.

Respondents come from a range of institutions including academic, public, and private sector libraries, museums, archives, and government agencies. Respondents held a range of positions—librarians, archivists, and information technologists from university settings comprised the majority of the group. A number of administrators were also present.

The main findings of the survey are as follows. Implications of these findings for the MetaArchive Cooperative and its potential market are discussed below.

1. **There is, at best, a rudimentary understanding of “digital preservation” and of the activities that this term includes.** For example, survey respondents report their very ambitious plans (e.g., 63.2% intend to become Trusted Digital Repositories and 100% intend to retain their digital collections for more than 10 years) which contrast radically with their nearly invisible actions to date (e.g., only 18% have any written digital preservation policies, only 13.6% have a digital preservation plan, and only 12% report that they operate a digital preservation solution. As discovered during the workshop, some of that 12% actually do *not* operate a preservation solution, but rather a backup system—they did not yet understand the definition of “digital preservation” when they responded to the survey.)
2. **There is widespread under funding and understaffing in current digital collections programs, of which digital preservation is cited as a minor component.** (e.g., more than 50% of the respondents say that grant funding is either the most important or second most important source of funding for their digital collections activities, nearly 50% report that they expect to use grants to fund the preservation of their digital collections, and more than 50% of respondents report that *less than six FTE staff and volunteers* work on their digital collections at all.)
3. **Although most institutions are actively creating and acquiring a wide range of digitized and born digital materials, few have reasonable measures in place to protect or preserve these materials.** (Although 94.7% report engaging back up strategies, only 21% report even employing off-site storage of backups. 16.7% report that they are creating no metadata for their digital collections.)
4. **Likewise, few are engaged in creating and/or amending policies and plans for preservation of digital collections.** (Only 33% have amended their deed of gift agreement to include digital content, only 18% have any written digital preservation policies, and only 44.4% of the remainder are currently developing them. Only 13.6% have a digital preservation plan. Only 65.2% of institutions report that the staff members with responsibility for preservation in their organization have any role in digital preservation.

The practice of digital preservation is still in its infancy, at least within the U.S. The institutions that sent attendees to this digital preservation workshop presumably are investing at least limited resources and energy in digital preservation activities, however, those surveyed indicate that most of these institutions are only just beginning to think about the issues and practices this will necessitate.

As a result, many are “dreaming big” and “acting small” (if acting at all). Most striking was the figure on the number of institutions that believe they will become Trustworthy Digital Repositories. The desire here may be something that the MetaArchive Cooperative is currently in a unique position to satisfy. The declaration of this intent by the majority of attendees may indicate that these institutions want to maintain control over their preservation processes rather than outsourcing to other Trustworthy Digital Repositories. Although this is costly to achieve alone (particularly if building your own solution), it is cost-effective to participate in a Cooperative structure—there accomplishing the goal of maintaining an active role instead of a passive, outsourced one, in the digital preservation world. They want to be actors, not purchasers, in other words, and the MetaArchive Cooperative currently offers a unique and rare opportunity for institutions to accomplish this aim.

ETD Survey Results

In September 2007, the MetaArchive Cooperative Steering Committee determined that all current members were interested in preserving their Electronic Theses and Dissertations (ETDs) in a dedicated ETD preservation Archive. The Steering Committee also expressed interest in surveying the broader University community to establish what, if any, demand there might be for this service.

The MetaArchive Cooperative’s Community Building Committee designed and disseminated an ETD Digital Preservation Survey in December 2007 - February 2008. This survey was intended to help the Cooperative to gauge the digital library community’s interest in establishing an ETD-specific preservation network. It was distributed to College, Graduate School, and Library administrators through a variety of targeted listserv postings. To date, 90 surveys have been completed and are represented in the findings below.

These findings and the work of the Community Building Committee have already led to the establishment of a partnership between MetaArchive and the Networked Digital Library of Theses and Dissertations to found an ETD archive that NDLTD members may turn to as a preservation solution for their collections.

The major findings of this survey are as follows. Implications of these findings for the MetaArchive Cooperative and its potential market are discussed below.

- 1. Preservation of digital ETDs is a dire need in the academic community.** Of the respondents, 81.2% accept ETDs and 37.68% of those *only* accept electronic versions. Preservation needs cover a wide range of formats, including some that are likely to go obsolete in the next decade. Respondents identified a wide range of formats currently supported, with the largest percent accepting .pdfs but many also accepting .ppt, .qt, .tif, .xml, .wav, .png, .mpg, .mp3, .aif, .avi, .doc, .gif, .html, .jpg, and .mov. They are also using a wide range of repository systems, including ProQuest, DigiTool, CONTENTdm, DSpace, Fedora, ETDdb, Eprints, bepress, and Digital Commons.

2. **Most institutions are not yet engaging a preservation solution for their ETDs.** 74.2% do not have a preservation strategy; the groups that report using a preservation service are largely (and perhaps entirely) working with ProQuest.
3. **There is a high degree of knowledge of LOCKSS among respondents and an even higher degree of interest in using LOCKSS as a preservation strategy.** 67% report experience with or knowledge of LOCKSS; 42.35% report that they would be interested in participating in an ETD-specific LOCKSS-based collaborative and distributed digital archive sponsored by the NDLTD; another 49.4% say they might be interested, leaving only 8.24% respondents explicitly not interested in participating.
4. **Not only are the vast majority of respondents interested in a LOCKSS-based solution, half of those would want to play an active role.** While 45.5% report wanting to have their ETDs preserved without sharing in the preservation responsibilities, the other 54% report wanting to run a secure server for the network. Of that 54%, nearly half would want to participate at the “Sustaining” level, both running a secure server *and* contributing to the growth and maintenance of the network in both technical and organizational ways.
5. **There is a strong desire for the preservation solution to be linked to an open archive with some public access.** 40% of respondents reported that they wanted open archiving with public access; 42.86% report wanting “dim” archiving, and only 17.14% are interested in dark archiving for this purpose.
6. **The two biggest factors respondents cite in determining their interest are cost and staff time.** In written responses, these ranked highest in the concerns cited. Other important concerns included the agreements made for guaranteed recovery of data, long-term goals, recovery information, hardware and software requirements, training opportunities, and sustainability.

The ETD survey strongly suggests that the University community has a strong need for the type of services the MetaArchive Cooperative can offer. Most Universities do not yet have a solution and are recognizing the need to engage one; there are few that satisfy the cost requirements and ideological concerns of the University representatives in this survey.

The MetaArchive Cooperative is poised to engage this community. In order to capitalize on the current interest, the Cooperative should plan to have an ETD archive created and populated (thus ready for new members) no later than May 2008. The Cooperative must also prepare a clear strategy for the quick growth spurt that this community will bring, completing that documentation by no later than May 2008. This should include a business plan and five-year development plan. With these pieces in place, the Cooperative will be ready to work closely with this community to provide the ETD preservation services—and in many cases, the active role in preservation—that they seek.

Competitors for this Market Segment

There are currently few digital preservation solutions available to this market. While this may change as quickly as in the next few years, the only existing solutions that an institution can either purchase or buy into nationwide are: ProQuest (for ETDs), PORTICO (for journal content), Archive It (for web-based content), SRB (largely used for datasets and large files in supercomputer settings), and LOCKSS (for journal content through the public network or for web-based content through a Private LOCKSS Network).

On a more localized level, both California (through the California Digital Library) and Florida (through FCA DAITSS) provide digital preservation solutions for the public institutions in their

respective states. These programs may prove to be models for other states or other consortia, but to date no external adopters have arisen for either system.

Such broader-based vendor solutions as Amazon's S3 service and Yahoo's Flickr service are additional competitors, the latter largely due to continued misinterpretations of what "digital preservation" entails. Because libraries and archives have such specific needs—and hold to particular standards in their work—at least to date, these broader-based services are not serious competitors for this market.

The only service beyond MetaArchive that offers institutions the opportunity to participate in their own preservation solution is LOCKSS (to which MetaArchive is fundamentally connected). Likewise, the only distributed digital preservation solutions are those that are LOCKSS-based, including MetaArchive.

There is no other service to date that offers institutions—nationwide and internationally—the opportunity to preserve their digital collections across genres and formats in a distributed digital archive.

Final Analysis

As evidenced by both surveys and by the unsolicited expressions of interest that the Cooperative has received from the community, the need for a cooperative approach to digital preservation is acute for the university, library, and archive communities. The MetaArchive Cooperative is uniquely positioned to help to meet that need.

There are several distinguishing qualities of the MetaArchive Cooperative that contribute to its current position in this field. These include:

1. It has been created by and for university libraries
2. It is built as an open source solution
3. It enables participants to take an active role in their own preservation processes
4. Its decentralized approach keeps its costs—and with them, its membership fees—very low
5. It accomplishes the two main objectives that preservationists have learned to value from their centuries of experience in the print world: security and distribution

For MetaArchive to have the fullest possible positive impact on the emerging field of digital preservation as a whole, it needs to carefully and systematically disseminate information about the services it can offer to these repositories—in person, in print, on the web, and through a variety of marketing pathways.

Notes on this document

This document was drafted by Katherine Skinner in 2008-01, was reviewed at the second MetaArchive All-Project Meeting held at University of Louisville on Friday 2008-02-08 by the Steering Committee and additional attending project participants, issued on 2008-03-06, and edited by Matt Schultz in 2009-10.

APPENDIX A: EXPRESSIONS OF INTEREST IN THE METAARCHIVE COOPERATIVE

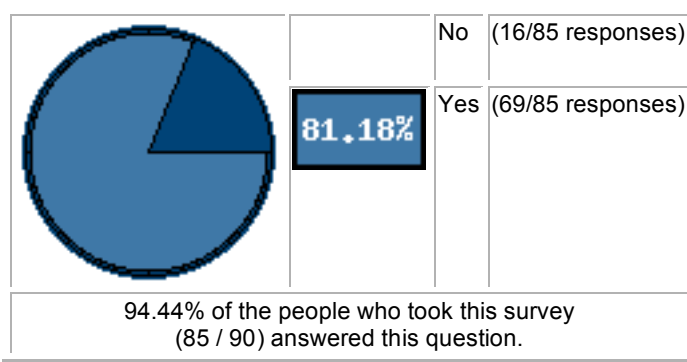
1. University of Mississippi
2. University of Southern Mississippi
3. Mississippi State University
4. Rice University
5. Dartmouth University
6. Purdue University
7. Boston College
8. University of Kentucky
9. Eastern Kentucky University
10. Western Kentucky
11. University of North Carolina at Chapel Hill
12. University of North Carolina at Greensboro
13. State Library of NC
14. University of Tennessee
15. University of South Carolina
16. Michigan State University
17. University of Michigan
18. University of Maryland
19. Arizona State Library and Archives
20. National Transportation Library
21. Alabama Department of Archives and History
22. Spring Hill College
23. Troy University
24. University of Alabama
25. University of Alabama at Birmingham
26. University of North Alabama
27. Utah Statewide
28. San Diego SuperComputer Center
29. University of Calgary
30. Joint Information Systems Committee (JISC)
31. Universidad Federal do Rio De Janeiro
32. University of Hull
33. Victoria University of Wellington
34. York University
35. Cal Tech
36. Texas A&M
37. Brigham Young
38. Tulane
39. Rutgers
40. Sept. 11th Museum
41. Augusta College
42. Kansas State
43. Colorado State University
44. Indiana group (Rachel)
46. University of Rochester
47. Bucknell University
48. University System of Georgia (GA Tech IMLS application)
49. Liberal Arts Scholarly Repository (LASR) (Contact: Abby at Bucknell)

APPENDIX B: ETD DIGITAL PRESERVATION SURVEY RESULTS

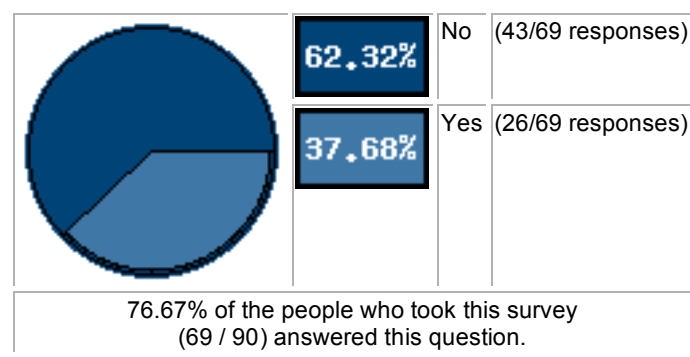


Viewing Results for Digital Preservation of ETDs

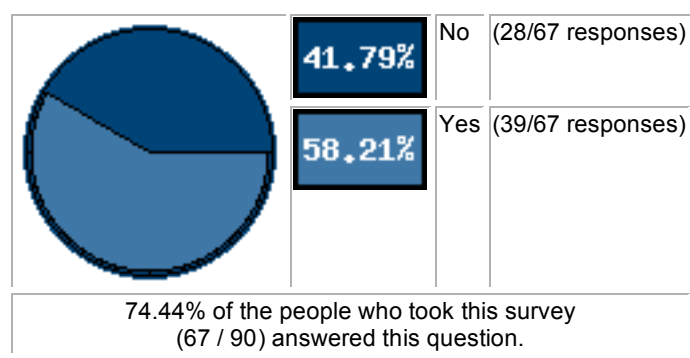
1. Does your institution accept electronic theses and dissertations (ETDs)?
If not, please go to no. 9 below.



1a. If so, does your institution accept only electronic versions?



1b. If so, does your institution also maintain print copies?



2. Please estimate the current size of your ETD collection in kilobytes.

Text Responses (52 for this question)	
".01 k"	
"0 now; begin in 1 month"	
"1,000,000 KB"	
"1,151,243"	
"1.21 GB"	
"10 GB"	
"10,000,000"	
"100,000"	
"1000000"	
"1089382"	
"111gb"	
"12,000,000"	
"12,000,000K (12GB)"	
"12474000"	
"132689920"	

"1708032"	
"2,000,000"	
"201080000"	
"24" - (2 responses)	
"2540416 KB"	
"286,096"	
"3,000,000 KB"	
"3,864,051"	
"3,972,449 KB"	
"3.4 gigabytes"	
"30000"	
"3000000" - (2 responses)	
"319866.53"	
"37,500 MB"	
"375,000,000"	
"4 865 392"	
"4,320,000 KB"	
"40,000,000"	
"500,000"	
"567627 KB"	
"6,782,464 KB"	
"686,080KB"	
"69075838"	
"70,000"	
"8067"	
"825,904,522 bytes"	
"870,000 KB"	

"9,636,213 KB"	
"9,686,808kb"	
"??"	
"???"	
"I don't know"	
"Just started, so none present yet"	
"lots"	
"unknown"	
"very small, just started"	

3. Please estimate the number of ETDs currently in your collection.

Text Responses (64 for this question)	
"0 now; begin in 1 month"	
"1,970"	
"1000"	
"10474"	
"1055"	
"1127"	
"1200" - (3 responses)	
"125"	
"13,200"	
"13930"	
"140"	
"1450"	
"146"	
"150" - (2 responses)	

"152"	
"1600" - (2 responses)	
"175"	
"1782"	
"1982"	
"2" - (2 responses)	
"2,500"	
"200"	
"200 volumes"	
"2000+"	
"20180"	
"240"	
"3,269"	
"3,600"	
"3,700"	
"300"	
"300 open access; 800 altogether"	
"301"	
"35" - (2 responses)	
"4,200"	
"42" - (2 responses)	
"431"	
"434"	
"45" - (2 responses)	
"466"	
"500"	
"60,000"	

"63" - (2 responses)	
"67"	
"687"	
"697"	
"700" - (2 responses)	
"75"	
"78"	
"8" - (2 responses)	
"855"	
"87"	
"900" - (2 responses)	
"903"	
"<10"	
"don't know - small, just started 2 months ago"	
"I don't know"	
"Only test theses so far"	

4. Please estimate the annual growth of your ETD collection in kilobytes.

Text Responses (54 for this question)	
"1 GB"	
"1 gigabyte per year"	
"1,000,000"	
"1,000,000 KB"	
"1,787,201 KB"	
"1.6 GB"	
"10%"	

"10,000"	
"10,000,000"	
"100 GB"	
"100,000"	
"1000000"	
"1200000"	
"150000KB"	
"170,000"	
"175,000 KB"	
"175,000-200,00 KB"	
"1gb" - (2 responses)	
"2,253,300 KB"	
"200,000 KB"	
"2000000"	
"2135"	
"23,407,000"	
"250,000"	
"263,206"	
"3,300,000"	
"300000"	
"3000000"	
"315,000,000 kilobytes or 315 GB"	
"3344"	
"363520"	
"382688 KB"	
"50,000,000"	
"500,000 kb"	

"500000"	
"5276000"	
"600 MB"	
"600,000"	
"657408"	
"6GB"	
"7000"	
"875000"	
"??"	
"???"	
"????"	
"approx. 267,402 KB/year"	
"don't know"	
"First year; NA"	
"hard to determine, submission becomes mandatory in 2009, perhaps one gigabyte per year"	
"not yet known"	
"unknown" - (3 responses)	

5. Please estimate the number of ETDs added to your collection annually.

Text Responses (65 for this question)	
"1 or 2"	
"1,000"	
"10" - (2 responses)	
"10,000"	
"100" - (5 responses)	
"1000" - (3 responses)	

"120"	
"120/year"	
"140"	
"150" - (4 responses)	
"150 volumes"	
"175"	
"180"	
"190"	
"200" - (3 responses)	
"25" - (2 responses)	
"25-30"	
"250"	
"270"	
"30-35"	
"300" - (6 responses)	
"35-40"	
"350" - (2 responses)	
"40" - (2 responses)	
"400-500"	
"45" - (2 responses)	
"5" - (2 responses)	
"50" - (2 responses)	
"500" - (2 responses)	
"550"	
"600" - (2 responses)	
"67 in 2007, 200+ in 2008, 300+ in 2009 when mandatory"	
"750" - (3 responses)	

"80-100"	
"800" - (2 responses)	
"900"	
"900 to 1,000"	
"approx. 550/year"	
"Approximately 350"	
"First year; NA"	
"too soon to say"	
"~100"	

6. What file formats do you support for your ETDs?
Select all that apply, and/or provide additional formats.

2.27%	.ppt	(6/264 responses)
1.52%	.qt	(4/264 responses)
4.92%	.tif	(13/264 responses)
4.17%	.xml	(11/264 responses)
6.44%	.wav	(17/264 responses)
5.68%	Any format	(15/264 responses)
5.68%	Other formats	(15/264 responses)
3.03%	.png	(8/264 responses)
20.83%	.pdf	(55/264 responses)
	.mpg	(12/264 responses)

4.92%	.mp3	(13/264 responses)
	.aif	(6/264 responses)
5.68%	.avi	(15/264 responses)
	.doc	(11/264 responses)
	.gif	(15/264 responses)
5.3%	.html	(14/264 responses)
7.58%	.jpg	(20/264 responses)
5.3%	.mov	(14/264 responses)
74.44% of the people who took this survey (67 / 90) answered this question.		

Other Responses (15 for this question)

".dwt, .xls, .mix, .csv, .zip"

".snd"

".tex"

".txt"

".txt, .midi, on special request: .exe, .jar"

"Graduate Office prefers PDFs"

"JP2, MIDI, TXT, CSV"

"No policy yet"

"other"(6) responses

"primary file must be either .pdf, ps or html"

7. How do you currently structure your ETD collection (i.e., do you separate them into collections based on discipline, year, or other criteria)?

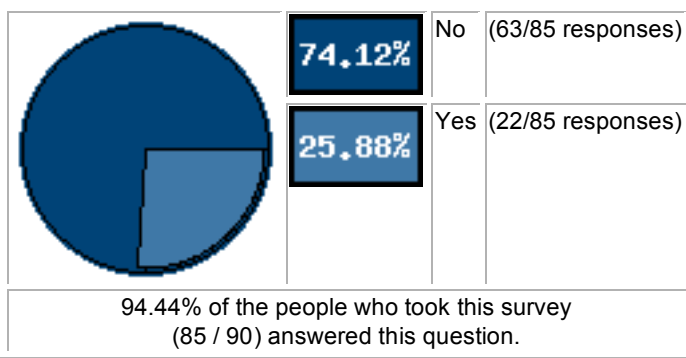
Text Responses (64 for this question)

"All ETDs are together in one database/collection."	
"All in one collection. Metadata allows searching by year, department, program, etc."	
"All items of the digital collection are together on the digital library system. Programs that perform specific queries on the database allow ETDs to be retrieved by graduate program, author, supervisor, research area, institutions, etc."	
"All Ph.D. and masters theses grouped together."	
"Author, Area, Year of Publication and Academic Grade (Under and Post graduate)"	
"By 'Closed' or 'Open'. Authors opt into 'Closed'. Closed ETDs revert to 'Open' after five years."	
"By academic department"	
"By author and department"	
"by college"	
"By discipline and degree"	
"by name and by department"	
"by year"	
"By year and semester"	
"Collection based on discipline and year"	
"Collection can be browsed by year or discipline. Currently we deposit only the open access ETD's, but this decision is being revised. We allow an embargo period of up to 5 years."	
"Collections structured by faculty and department, degree, subject (Proquest UMI subject categories)"	
"Currently inserted into the library catalog. Also, there is a provider for the Open Archives Initiative. In addition we provide a site where on campus folks can browse and sort by student last name or by department."	
"currently no structure"	
"Department, College, Advisor, Year"	
"Discipline and keywords"	
"discipline and year" - (2 responses)	
"Dissertations"	
"Documents are cataloged by standard meta data captured by libraries. The electronic collection is in a single database capturing all collection data."	
"don't know"	

"Folders stored by semester, and then author. A database indexes a meta information sheet, and works are then searchable by author, title, discipline, degree, etc."	
"Index on author, department, advisor, last name, date, abstract keyword, title"	
"level masters, PhD"	
"No"	
"nope, one big free for all"	
"One collection for theses and one collection for dissertations"	
"One single collection"	
"One single collection, currently plan to separate by year, school, etc. as collection grows"	
"part of the searchable database in the library by usual criteria"	
"Probably will be in a full-text searchable collection by year."	
"Quarter and Year"	
"Searching by author, title, university department, and year"	
"Separate based on year and classification (i.e. 2003-Present and undergraduate honors, respectively)"	
"Separate collection for ETDs with sub-collections grouped according to University hierarchy."	
"separate them into collections at a departmental level."	
"Separate them into discipline (browse). Search can be done by year, by full-text"	
"single collection, divided into theses and dissertations"	
"Storage of bit streams is by year. Browsing is by department name or author name."	
"The are searchable by all data elements in bibliographic record including LC subjects, keywords from title, abstracts. Also browsable by year and author."	
"The collection is organized by access type (world-wide/campus only, restricted), but a user may search the database by keyword."	
"Theses and dissertations are in separate collections divided by year. ETDs are also divided by discipline within our IR."	
"They are initially stored in 2 directories: available and withheld, with all eventually moving into available--in no particular order."	
"They are not separated."	
"they are treated the same as our print - categorized by discipline"	

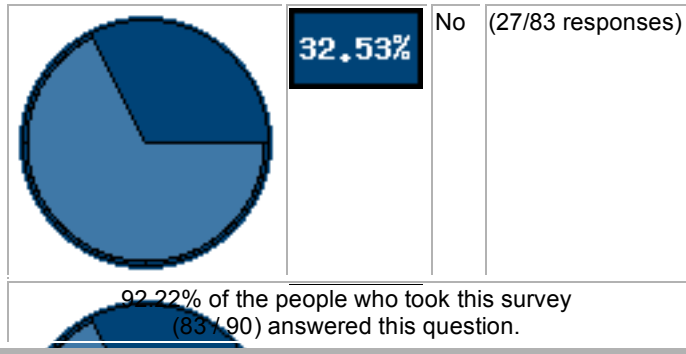
"we are just beginning, so this has not been discussed yet"
"we can browse by degree, dept/program. We can search, and limit by year, degree, dept/program & language"
"We catalog them like other monographs, and identify them as such in the appropriate MARC field."
"We do not separate our ETD collection; however, the user may browse by department or author and search by many different search criteria."
"We have not a structure specified for the collection but rather set up a special ETD portal for multiple search functions."
"We just add them as we get them and find them by searching."
"We keep a dark archive copy of files that are submitted to ProQuest. Access is through ProQuest Digital Dissertations until institutional repository is operational."
"We use a modified VA Tech ETD db system, so the primary file structure division is by access restriction ('available/unrestricted', 'available/restricted' and 'withheld'). Although ETD's for all years and disciplines are stored in the same file structure, the search interface allows filtering by Graduate School, Program, Degree, etc."
"will be tagged with year & discipline; in online catalog primary arrangement is chronological"
"Year" - (3 responses)
"Year and Degree Type"
"Year and ETD type (i.e. BS thesis, MS thesis, Ph.D. thesis)"
"You can browse by department or author."

8. Does your institution have a formalized preservation plan for its ETDs?
 If so, and you are willing to share, please send it to gallmac@vt.edu.

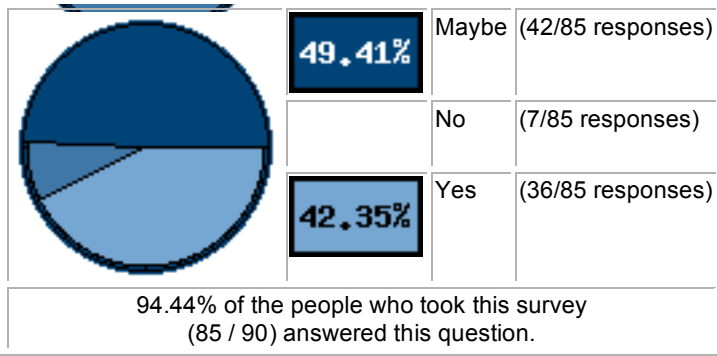


9. Do you have experience with or knowledge of [LOCKSS](#)-based preservation networks?

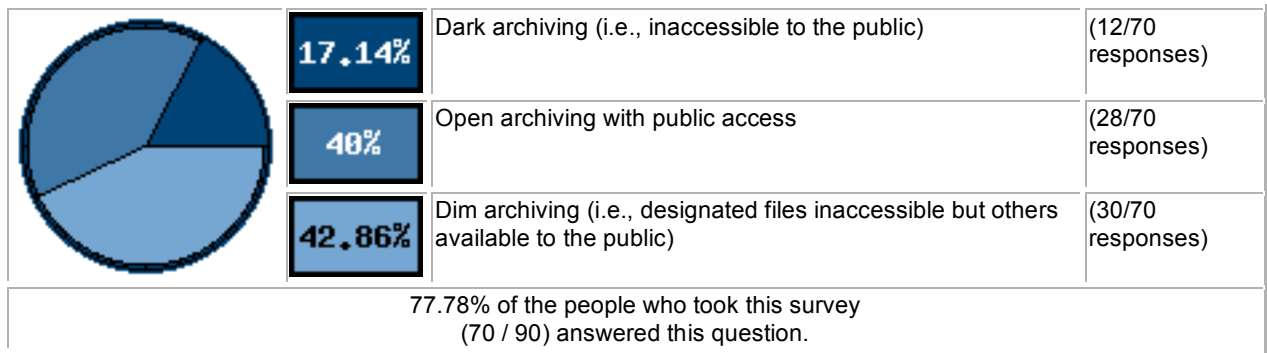




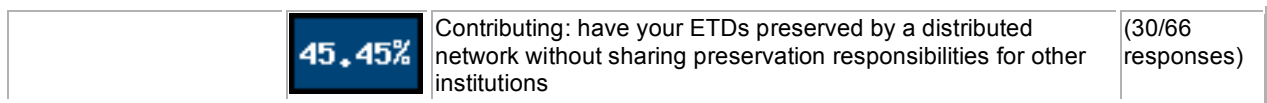
10. Would your institution be interested in participating in an ETD-specific [LOCKSS-based](#) collaborative distributed digital archive sponsored by the [NDLTD](#)?
If not, please go to no. 11 next.

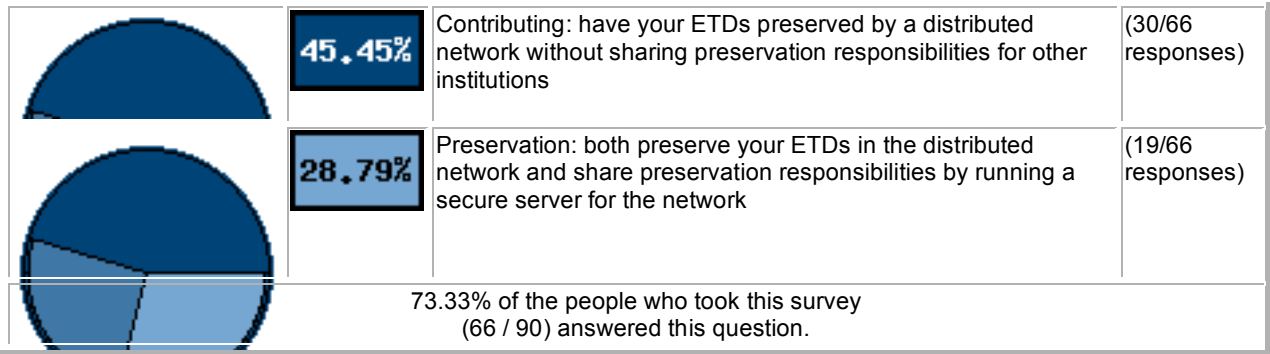


10a. If yes, would there be a preference for
(select only one)

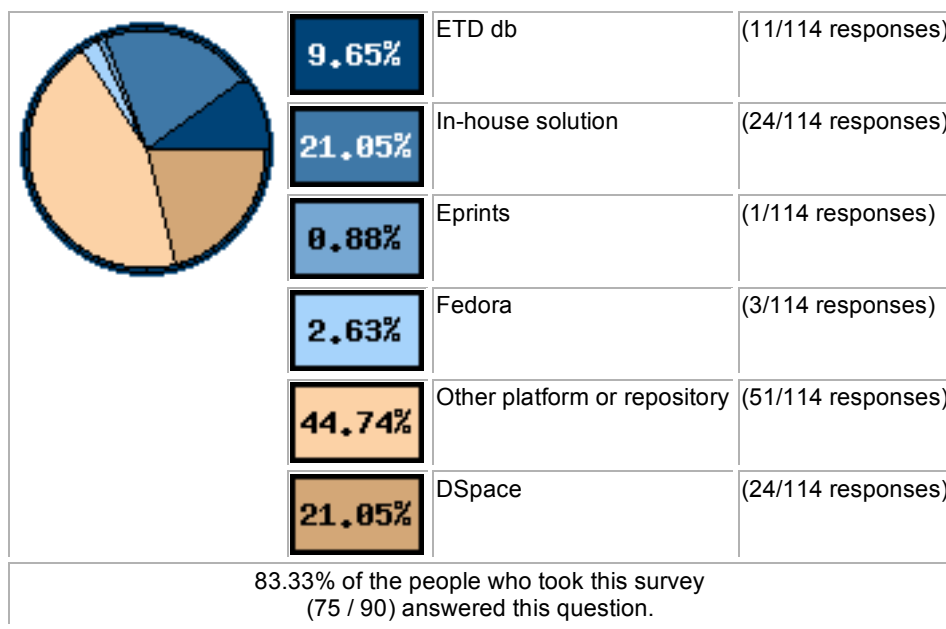


10b. If yes, what level of participation might your institution support?
(select only one)





11. What platform or repository structure are you using to collect, disseminate, and store your ETDs?
Select all that apply.



Other Responses (51 for this question)
"Archimede"
"bepress / digital commons"
"content-DM"
"CONTENTdm"(3) responses
"CYBERTESIS"
"DIAS"
"Digital Commons"

"DigitalCommons"
"DigiTool"(2) responses
"DLXS"
"Documentum"
"Fez"
"Millennium Online Catalog"
"not applicable"
"OhioLINK and locate CDs"
"Oracle"
"other"(24) responses
"ProQuest"(2) responses
"ProQuest - will be a Fedora-based solution in approx 1 year"
"Sesame + in-house Java code"
"UMI ProQuest"
"VTLS Vital"
"We are getting ready to implement a ProQuest hosted solution."
"we're still in the planning stages"

12. What information would your institution need to make a decision about participating in an ETD preservation network?

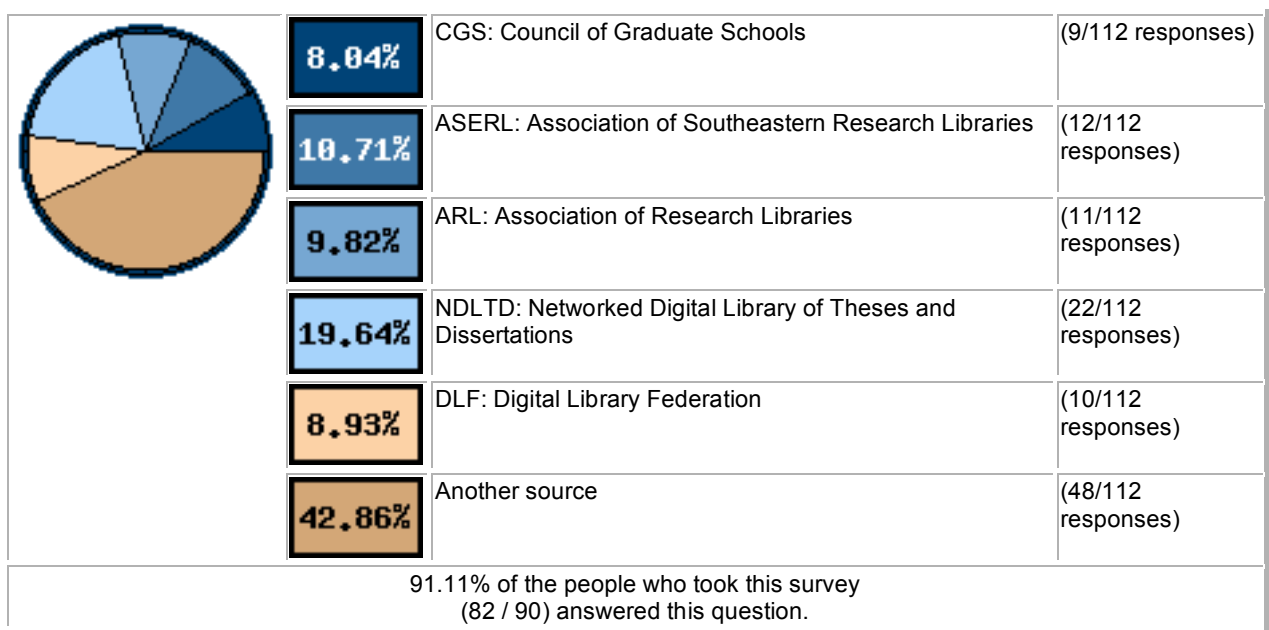
Text Responses (59 for this question)	
"(1) Our Director have reservations about making our full text available for harvesting: will each institution preserve just their own content, or will all participating institutions preserve all available collections? (2) we are not sure about 10b. possible involvement will depend on availability of disk space and other hardware requirement; we are already running a LOCKSS server for journals"	
"--Costs --Service level agreements for guaranteed recovery of data"	
"Administrative, policies & procedures, costs, and long-term goals."	
"Cost (human resources as well as other), benefits (including accessibility), any migration strategies for materials, information about metadata required"	

"cost and how we would access the data when needed"	
"Cost and staff resources."	
"Cost and staff time."	
"Cost and staffing expectations"	
"Cost in terms of fees and staff time required."	
"Cost, access methods, system requirements"	
"Cost, operating agreement terms, expectations for participants."	
"cost, organization, staffing implications"	
"Cost, responsibilities, access to content, submission process, long-range preservation plan"	
"cost, scope"	
"Cost, server space required, other technical issues"	
"Cost, software and hardware needs, expertise of staff required"	
"cost, technical requirements & level of technical difficulty involved"	
"cost, work in preparing archives, security of information."	
"Cost. Timing. Liability."	
"Costs associated"	
"costs, other commitments"	
"costs; terms"	
"Defined responsibilities; frequency of harvesting; costs; targets; hardware specifications"	
"Especially information on the legal aspects of the preservation network, access scenarios and the like."	
"Hardware, software and staffing requirements for our institution."	
"Hardware/software requirements"	
"How it would work intra country, NZ participates in the ADT Australasian Digital Theses, the joint group CONZUL has not renewed membership of NDLTD because of the cost."	
"In principal we would be interested in exploring this possibility. Obviously we would need a clear understanding of the resources - both monetary and technical - required to participate in such a program."	
"information on cost - initial and on-going, other details of requirements, etc."	
"Letter of invitation"	

"Level of work involved, relationship to other activity (via British Library) and relationship between LOCKSS and Fedora"	
"LOCKSS and NDLTD importance documentation"	
"Much more detail on the implementation strategies for this project."	
"Need preservation and access policy from you."	
"Need to know the data storage requirements for the LOCKSS box. Need a LOCKSS plug-in to work with ETD db software. Will one be developed?"	
"None."	
"Not applicable we are a community college and don't get ETDs."	
"Not sure - someone from UBC library will probably answer this survey; they'd know"	
"Not sure at this stage"	
"requirements for submission, sustainable preservation, annual cost"	
"resource commitment required (staffing, system, monetary)"	
"resource commitments, business plan, intellectual property issues"	
"resource requirements"	
"Scope, membership, technical requirements, monetary investment, lots more."	
"Server, server platform, and price."	
"specific responsibilities and technical details"	
"Staff, technical, support, financial etc. resource demands for the institution"	
"Sustainability."	
"Technical details of how it would work. Our actual content is not in a dspace or eprints like repository. We would probably have to contribute content or setup another server potentially."	
"Technical requirements, choice of platforms, data transfer methods, etc"	
"The benefits, the cost, staffing requirements, commitments"	
"Time and cost and training involved."	
"We have heard that Library and Archives Canada is planning to become a Trusted Digital Repository. So, we would first discuss such an initiative with Theses Canada Portal."	
"We would need hardware specifications as well as the programming needed to integrate with our system."	
"We would need to know what the cost (one time, on-going, etc.) would be, the procedures for participation,	

institutional requirements, etc."	
"What kind of resources (people, equipment & money) commitment would this involve?"	
"What skills would we need to participate? Amount of time? Buy-in from administration. Cost? Staffing? Access to other institutions' theses and dissertations."	
"What the preservation strategy will be for e-theses, who will do the work, and what the business model will be for the archive. Also what sort of access control is available."	
"Would there be additional responsibilities beyond digital storage?"	

13. How did you learn about this survey?
Select all that apply, or enter another source.



Other Responses (48 for this question)
"colleague"
"ETD list"
"ETD listserv"(2) responses
"ETD-L"(4) responses
"ETD-L (ETD-L@LISTSERV.VT.EDU)"
"ETD-L discussion list"
"ETD-L mailing list"

"ETD-L@LISTSERV.VT.EDU"
"Forwarded from CGS by Grad Office Dean"
"Gail's email to me"
"Graduate School"
"Listserv"
"listserv.vt.etc"
"MetaArchive"
"MetaArchive list serve and FSU Library director via forwarded ASERL email"
"NDIIPP MetaArchive"
"other"(24) responses
"Virginia Tech NDLTD project"
"VIVA"(2) responses
"VIVA colleague"

14. We'd like to receive your comments and concerns about digital preservation of ETDs, particularly the distributed (i.e., [LOCKSS](#)) model that the [MetaArchive Cooperative](#) is considering.

Text Responses (33 for this question)	
"A welcome opportunity for academia to regain control of its intellectual properties, and cost-saving through cooperation. Our institution is committed to LOCKSS."	
"All that LOCKSS does is to help with saving bit-strings. This is far from sufficient for digital preservation! What about ensuring that what is retrieved is authentic? How about ensuring that is intelligible?"	
"Currently using LOCKSS on local servers for some eResources (journals)."	
"Currently, we are doing our own preservation by burning a CDs at the end of each quarter of a Word and PDF file for each student. When we were smaller (2000-2005), they required the student to place the final documents on a diskette, CD, or Zip Disk. We are not considering moving these to CD media, because as the media becomes outdated, so does our ability to retrieve the documents. I am interested in your survey results. Thanks, Angie McCutcheon Director of Electronic Thesis and Dissertation Ohio University 44 University Terrace Athens, OH 45701"	
"don't really know very much about it"	
"generally, I think distributed models for preservation are the best way to go"	
"have no opinion as have no knowledge"	

"How would confidential ETDs be handled, such as those with patents pending?"	
"I am the Dean and only able to answer some of these questions. I am sending the survey to the person in our library who handles the ETDs"	
"I would like to know more about your logic preservation strategies (renderable, readable) for multiple types of file formats."	
"I'm very concerned about how to migrate materials to archival quality formats while enabling students to work with a variety of formats. Are you extracting text alone into XML? Creating tiffs? Our oldest ETDs are PDF version 1.3 and already display badly. What good is maintenance of the bit stream without migration to supportable formats?"	
"Issues of viability of LOCKSS as a true preservation model: bit streams are maintained and "single point of failure" concerns addressed; provision to fully understand and migrate file formats [etc.] not currently in place. Would want a clear plan or model for improving LOCKSS functionality before relying on it for long-term digital preservation."	
"It was somewhat difficult for us to separate out LOCKSS preservation from the larger question of a TDR... is that even part of this model? Question 10a. seems to be about both embargoed files and the branding of institutional content.... In Question 10b. It is our experience that the technology currently recommended for LOCKSS boxes is insufficient for actually serving up the content."	
"It would be great to be able to access them electronically."	
"Keep our institution informed of any developments."	
"LOCKSS would provide for the replication of e-theses (an aspect of a preservation strategy) but much more is needed to preserve them for the long-term, so that strategy will need to be public to make this service a viable alternative to local solutions based directly on our IR."	
"LSU is currently considering various alternatives for off-site digital repositories. A repository specifically run for ETD's would be most appropriate."	
"Note from question 1 -- if student submits electronically, no paper version is submitted. When ETD is mandatory in 2009, no paper copy will be submitted. For preservation, we store a dark copy on a secure, backed up server and students submit a copy to ProQuest."	
"The library doesn't have electronic copies, only paper copies. However, individual departments do have electronic copies - kept on cds for the most part. My guess is that the Graduate Studies office would be in charge of preservation."	
"The national ETD project for South Africa is in the beginning phases. We don't know if the LOCKSS model can be applied to national projects like ours. In other words, do we participate as individual institutions if our country has a national project? (Hussein Suleman is the expert for the national project); Another concern is that we are not yet convinced that the LOCKSS model is working well for archiving journals; based on our experience with participating in the LOCKSS-SA project."	
"This looks interesting, but I'm not sure we need it as we have a preservation repository."	
"We believe that it is a very useful solution, especially for institutions that do not have specific and formalized preservation plan."	
"We don't feel like it's necessary because Bepress is providing the back up for our ETDs and they are submitted to UMI as well."	

"We have a LOCKSS box running currently. Re: formats, we are asking for PDF, but are open to any format if requested. We are also working on an ETD preservation plan in coordination with the British Library EThOS service development."

"We have been talking with MetaArchive about LOCKSS networks and are interested in finding out more."

"We have limited Systems Staff so would want to minimize demands on their time."

"We may be interested later but are trying to avoid scattering our efforts currently. The other issue is that we're not a LOCKSS institution. We're currently involved with Portico."

"We need to do this for our own institution but I'd also like to help institution in less developed countries."

"We'd be interested in knowing how the project would handle embargoed ETDs. Also, does this address any kind of file migration, or are you thinking about the static preservation of the original submitted formats? We anticipate having to move files to different formats as standards change."

"We're in the process of negotiating with our Graduate School about how best to transition to a digital-only program managed collaboratively by the libraries and the graduate school. We have some limited experience with LOCKSS but I think we're still a ways from implementation."

"We're interested in exploring a LOCKSS-based preservation network for ETDs. But what we could really use help with is choosing a repository platform."

"We've already begun using VTLS' VITAL for our digital photographic archive and are planning on expanding its use into other areas of our collections such as our theses."

"We've been part of the ASERL pilot project to preserve ETDs with the LOCKSS model and are very encouraged that it will work."

http://lumiere.lib.vt.edu/surveys/results/view_results.php3

Generated at 03:33 pm on Monday, February 4th, 2008