



University of Kentucky
UKnowledge

Library Faculty and Staff Publications

University of Kentucky Libraries

1-2002

Long-term Retention of Electronic Theses and Dissertations

Thomas H. Teper

University of Illinois Urbana-Champaign

Beth Kraemer

University of Kentucky, kraemer@email.uky.edu

[Click here to let us know how access to this document benefits you.](#)

Follow this and additional works at: https://uknowledge.uky.edu/libraries_facpub

 Part of the [Library and Information Science Commons](#)

Repository Citation

Teper, Thomas H. and Kraemer, Beth, "Long-term Retention of Electronic Theses and Dissertations" (2002). *Library Faculty and Staff Publications*. 20.

https://uknowledge.uky.edu/libraries_facpub/20

This Article is brought to you for free and open access by the University of Kentucky Libraries at UKnowledge. It has been accepted for inclusion in Library Faculty and Staff Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

Long-term Retention of Electronic Theses and Dissertations

Notes/Citation Information


Published in *College & Research Libraries*, v. 63, no. 1, p. 61-72.

Copyright © The Author(s)

Long-term Retention of Electronic Theses and Dissertations

Thomas H. Teper and Beth Kraemer

This paper examines the increasing trend of universities to pursue electronic thesis and dissertation (ETD) programs. Although the goal of most programs is similar, procedural variations impact a program's long-term success. As primary research generators, responsibility for providing long-term access to unique materials must be borne by universities. However, this responsibility is in conflict with many ETD program goals, such as increased access and ease of production.

n much the same way that digitization projects came to represent a university library's technical prowess in the 1990s, the growing trend for universities to pursue electronic thesis and dissertation (ETD) projects is something that institutions can no longer ignore. Not only do ETD programs provide universities with the opportunity to promote their programs, they also enable institutions to advertise their technical muscle. Although the goals of most ETD programs are similar, procedural variations among institutions influence the long-term success of these programs. Consequently, ETD programs are projects that institutions should approach with a great deal of measured thought and consideration. Technical variations such as the electronic formats chosen for the submission and retention of these unique documents, combined with an institution's willingness to commit resources for proper long-term migration and storage, will have a significant impact on the long-term reten-

tion of ETDs. If these documents do not survive in the long term, or if the later recovery of stranded data requires significant additional funds, it is the authors' assertion that these programs can hardly be called successful.

The responsibility for providing long-term access to unique materials must be borne by universities. Traditionally, this responsibility has been that of a university's libraries and archives. For those institutions without active preservation, conservation, or records management programs, the principle of benign neglect has occasionally proven a material's greatest ally. However, this only applies to traditional, paper-based materials. History has proven that benign neglect is not an acceptable manner in which to preserve access to electronic and digital information. Moreover, as standards for digital archiving have yet to be established, programs embarking on ETD projects must make decisions that will affect the long-term feasibility of their programs with no specific guidelines be-

Thomas H. Teper is Head of Preservation at the University of Illinois Urbana-Champaign; e-mail: tteper@uiuc.edu. Beth Kraemer is an Electronic Resources Librarian at the University of Kentucky; e-mail: kraemer@email.uky.edu.

yond the pantheon of "best practices" that continue to morph with each new technical iteration.

Unlike the retention of paper documents, the long-term retention of electronic documents is an active, resource-intensive process.

In the authors' investigation of those ETD programs available through the National Digital Library of Theses and Dissertations (NDLTD) Web site, few institutions embarking on ETD programs appear to be actually considering the long-term ramifications of their decisions. If long-term preservation is considered by institutions embarking on ETD programs, the resulting decisions are often based on compromises in which the simplicity of student production and the university's twin desire for immediate publication and an immediate Web presence become the primary considerations while concerns about long-term access are put on the back burner. Even worse, many programs appear to leave long-term preservation issues unspecified, adopting a "we'll deal with that when it comes up" approach. As demonstrated countless times, this cavalier approach could result in information loss. Unlike the retention of paper documents, the long-term retention of electronic documents is an active, resource-intensive process. As a result, universities that intend to maintain their information must undertake long-term preservation planning. Another issue that heightens the need for intensive planning is that, unlike other documents that might be digitized to provide better access, ETDs are inherently "born digital" and do not necessarily have eye-legible backups available. Consequently, a lack of institutional planning for long-term retention may result in the loss of these unique documents.

As institutions pursue ETD projects, their practices are going to affect the probability of providing long-term access to the product they are desperately attempting to market. Outside the ETD commu-

nity, institutions have long experimented with additional options for the long-term preservation of, and access to, digital materials. Through a reasoned examination of the strengths and weaknesses of specific formats, format strategies, the regulations governing institutional records, and the purpose of information production, suggestions for ensuring long-term access and the long-term success of ETD projects will be examined.

Theses, Dissertations, and ETDs

The archiving of electronic documents is a hot topic in many institutional communities, including universities, libraries and archives, museums, private businesses, and the records management industry. Although all of these communities differ greatly, all share an interest in what technology can offer. However, three main components make the long-term retention of electronic documents different from that of paper. First, born-digital information has no innate paper backup. Consequently, there is little to fall back on should format changes strand data. Second, electronic documents are different from paper because access and delivery will change in the future. Archiving electronic documents is an active process, and the best format for delivery is not necessarily the best format for retention. Finally, the production and storage of paper documents are relatively straightforward processes that have remained relatively stable over time. The production and retention of electronic documents is not quite so simple. As there is little opportunity for institutions to anticipate format, changes will depend on an institution's ability to maximize its flexibility.

Within the ETD community, long-term retention is an issue because few institutions consider long-term access issues when making format and procedural decisions about their programs. Currently, many rely on proprietary formats for both document delivery and retention. Unfortunately, this practice is not in keeping with current archival and preservation thinking. Many large-scale digitization

projects carried out in the mid-1990s concluded that reliance on proprietary software and hardware was a mistake—and a costly one, at that.

As stated earlier, ETDs differ from traditional theses and dissertations in that they are born-digital documents. This is what makes them simultaneously so tantalizing to some and so feared by others. Within the realm of the traditional thesis or dissertation, there is one format—paper. That paper record is an eye-legible, permanent backup. Moreover, microfilming by UMI provides institutions with an additional backup should something happen to the original.

Few authors expend energy developing a standardized definition of what digital preservation actually means, let alone that various facets of it are more or less applicable depending on the situation.

As technologies developed that permitted students to create multimedia packages, universities began accepting theses and dissertations with a multitude of additional components intended to enhance the research. Reel-to-reel tape, audiocassettes, photographs, videos, floppy disks ranging in diameter from 3.5" to 8", and CD-ROMS all became integral components of a student's research; and many libraries worked to incorporate these within bound volumes. Over time, the functionality of these component pieces declined to a condition of nonfunctionality because institutions could not reasonably ensure continued access. The resulting document could then be considered incomplete. Despite this incomplete state, the functional paper component remains a testimony of the student's original accomplishment.

Present developments are rapidly leading institutions to a point where they envision electronic documents as the normal means by which students submit theses and dissertations. Unfortunately, deficient planning in regard to acceptable nonproprietary format types and the cre-

ation of backup versions of a student's work threaten not just components of the ETD with obsolescence, but also the entire document. No more will the bound volume remain a partial record of the student's work. Long-term access will need to be ensured by administrations that, during times of economic hardship, will be just as likely to fall back on the old stopgap of benign neglect.

Five chief factors affect the longevity of electronic formats. Formats must be well documented, well tested, nonproprietary, widely distributed, and platform independent.¹ Unfortunately, there are no archival standards or accurate gauges for the longevity of electronic formats. This may come in the future, but current projects base their decisions on guidelines and best practices developed by other programs and research projects. Although these do provide incredible assistance, guidelines and best practices inevitably reflect the fluid technological environments of their creation, leaving institutions with outdated projects.

Viewing Preservation within the ETD Program

Within the university library setting, the term *preservation* is an umbrella term that concerns itself with providing access to materials for as long as they are needed by whoever might need them. Preservation involves binding, conservation, deacidification, care and handling, and reformatting programs, and its success depends on cooperation within the institution. Preservation also involves making choices. Unfortunately, the resources available are frequently far below what *could* be spent on preservation programs, and institutions must prioritize how their dollars are spent.

Another way to view preservation within the library context is as asset management.² Asset management is the business of providing access and protecting the institution's investment. Although this may be an uncomfortable truth, universities are businesses. They have investment portfolios, assets, insurance, and

asset managers. They are in the business of creating knowledge, educating students, and perpetuating themselves. Within this context, preservation projects are asset management programs that focus on library and archival collections that often include university records and students' theses and dissertations.

Although preservation administrators frequently pursue the same objectives, differences are emerging with the ready adoption of digital media. These differences are causing individuals within the library community to rethink traditional notions of preservation and to formulate methods of handling emerging technologies. One example of this comes from Maggie Jones, the National Library of Australia's director of collection management and retrieval service. In a paper entitled "Preservation Roles and Responsibilities of Collecting Institutions in the Digital Age," she highlighted this transformation in thought:

In the digital environment the links between selection of materials, provision of access to those materials, and preservation of them over time is so inextricably linked that at the National Library we tend to talk increasingly simply of providing short- and long-term access rather than even making a semantic distinction between preservation and access.³

As members of the ETD community view such assertions, the possibility exists there is confusion about what constitutes preservation. Within the preservation community, administrators frequently consider preservation and access integral components of the same goal. Without preservation, long-term access is impossible; without long-term access, preservation is meaningless. Traditionally, the key concept in preservation is maintaining access to the intellectual content of the item, not necessarily the artifact. This does not mean that preservation programs ignore the artifact; indeed, most are based on maintaining access to

the original for as long as possible before resorting to reformatting options.

The second area of apparent confusion deals with the phrase "digital preservation." Much like the term *preservation, digital preservation* is best viewed as a blanket term that incorporates two concepts. The first concept is preservation of physical objects through digital imagery. This involves providing access only through digital surrogates. The reduced usage of the original decreases the likelihood of use-related damage. However, real success depends on an institution's willingness to create written policies and procedures that restrict access to original materials. The second concept is that of preserving born-digital information. This concept pertains directly to the ETD community, and it is the authors' belief that, as links to our collective intellectual history, they need to be preserved.

In the past decade, a great deal of literature appeared about the electronic environment and its impact on education, scholarship, and librarianship. One of the areas most written about and debated within the library field is that of digital preservation. Books, articles, and research publications range from a desire to throw caution to the wind to those that seek to proceed cautiously. These two approaches, opposingly labeled "futurist" and neo-Luddite, continue to produce the bulk of this material.

However, the very definition of digital preservation is incredibly vague and tends to vary from author to author. Few authors expend energy developing a standardized definition of what digital preservation actually means, let alone that various facets of it are more or less applicable depending on the situation. With that in mind, this article maintains that there is a distinction between preserving digital information and preserving artifacts through digital imagery. *Digital preservation*, therefore, is an umbrella term that encompasses a number of different practices.

The third definition, and the one that pertains to the ETD community, is the

preservation of digital information. ETDs represent the further development of our collective intellectual heritage. They are records of a student's creativity. In addition, in some states, theses and dissertations stand as permanent records of a student's academic accomplishment. Consequently, they are permanent records according to the state's records retention schedules.

Understanding that the institution is, in one sense or another, both morally and ethically bound to preserve the materials submitted by students, materials with which institutions have traditionally been entrusted, the problem now becomes one of how an institution goes about preserving something as complex and multidimensional as an electronic thesis and dissertation.

In *Preservation in the Digital World*, Paul Conway concluded that there were three requirements for digital preservation—making use possible, protecting the original item, and protecting the surrogate.⁴ These conditions are valid in the ETD community. Moreover, the need for institutions to preserve digitized information is more important to the ETD community than it is to the community of digital library projects that gave rise to Conway's report.

The reason centers on the simple fact that the programs generally consider ETDs to be electronic entities—created, accessed, and stored in an electronic environment. Although this facilitates short-term access, it should raise serious questions about the potential for long-term sustainability because, unlike scanning projects, there is not necessarily a hard copy to fall back on. The variable created by the ETD is that of an electronic original. This means that institutions face a situation in which the electronic surrogates created by digital preservation projects are now effectively equal to the original items. Failure to protect the original is now equal to a failure to protect the surrogate and therefore negates the possibility of future use. Moreover, as is demonstrated in later sections, the costs of

recovery far outweigh those of proper planning.

The Problem of Preserving ETDs

The digital environment's flexibility is an incredible benefit to the methods by which users may access materials. However, the instability that accompanies an industry in which the developmental year is measured in six-week intervals means that long-term preservation of digital information is difficult. As a result, it is the authors' assertion that digital preservation, as defined by many individuals, is a misnomer. The process for creating permanent digital surrogates akin to preservation microfilm is not yet a reality.

The research of Jeff Rothenburg, a computer scientist with the RAND Corporation, concluded that born-digital information such as that recorded in ETDs required four things for preservation. First, preserving the item required its ability to be copied perfectly. Second, preservation required that individuals had the ability to access the information without geographic restraint. Third, the preservation of digital information required that the item be machine-readable. Finally, the preservation of born-digital information required that an institution preserve the unique functionality of the original item.⁵ Within the ETD community, preserving functionality is, perhaps, the most important aspect of the equation. The potentially dynamic nature of ETDs makes them so desirable to institutions and students. Without their dynamic functionality, the ETD is little more than a paper document—static.

However, projects that hope to be successful in both mounting their ETDs online and maintaining their long-term access and functionality must weigh the short-term benefits of instant access and an immediate Web presence against other considerations, such as the very real need to maintain long-term access. Traditionally, preservation focused on activities that increase the period in which access to original materials is possible before reformatting them. Because electronic in-

formation is increasingly becoming the norm, a number of models are being developed within the preservation community.

The first is predominantly technological. Through the planned process of migration, some maintain that institutions may preserve the functionality of their original projects. An example of this strategy is the LOCKSS model. LOCKSS, which stands for Lots of Copies Keep Stuff Safe, maintains that institutions must preserve the bits themselves, access to the bits, and the ability to translate the bits. It also maintains that the presence of many distributed, open-source electronic archives is key to maintaining long-term access. However, as characterized by Rothenberg, the research of many computer scientists primarily encourages the development of software emulators.

The CEDARS project, a collaborative project coordinated in the United Kingdom, concluded that both migration and emulation have merits, depending on the situation.⁶ However, one of the project's more controversial conclusions in the digital preservation realm was that preservation and access are not necessarily the same thing. Consequently, preservation administrators within the ETD community must understand that although preservation and access are both integral components of one another, they might be the end goals of two separate, but necessary, processes.

More traditional preservation models rely on analog backups of electronic materials. The greatest problem with these, however, is that the electronic environment does not easily transfer to the analog world. The functionality of dynamic Web pages, databases, and their ilk cannot be replicated in eye-legible media.

The final preservation model takes its cues from early library preservation. In the electronic realm, it has arguably done more to harm our cultural resources than any other preservation activity and can be characterized by the phrase, "put it on the shelf and hope for the best." Although this model has succeeded with some tra-

ditional materials over the span of a few hundred years, Seamus Ross related the chance recovery of digital data to the recovery of archaeological materials:

Information stored in digital form is as delicate as archaeological remains of flora and fauna—it is rare to discover them, the environmental conditions under which they were deposited influences their survival, their recovery and study depends upon substantial investment of labour, and their interpretation requires a vast array of scientific technique.⁷

Within the ETD arena, benign neglect is a guaranteed model for failure.

Format Choices, or PDF and Long-term Permanence

Whereas preservation professionals and working groups are seeking to find a happy medium in which the concerns of long-term access and preservation are both considered and met, many members of the ETD community are taking advantage of a false middle ground between the technological and traditional views of preservation. This middle ground is quickly making itself appear to be some sort of standard, despite the fact that it is not. It is Adobe's Portable Document Format (PDF). Increasingly attached with large-scale efforts at distributing textual information, PDF's faithful replication of printed formats is making it one of the prime means of communicating textual data online. Indeed, it has become the de facto standard for the Government Printing Office's (GPO) publication of government documents as well as the publication of a great deal of the Web's white and gray literature.

The GPO's adoption of PDF has made it one of the most preferred formats because many institutions mistakenly assume that the government's current use will ensure the format's long-term viability. However, it is the authors' assertion that this belief should be adopted only

with extreme caution. Although many individuals are dutifully working at developing long-term preservation and access strategies, the federal government's record for preserving data is far from exemplary. In the 1970s, the National Aeronautical and Space Administration (NASA) transferred to magnetic tape a great deal of information that is no longer readable. The National Archives and Records Administration (NARA) lost the records for many Vietnam veterans after electronically encoded files became corrupted. Making assumptions about the longevity of magnetic tape, NARA destroyed paper records decades before this tragedy. As a result, verifying the service records of tens of thousands of veterans is next to impossible. Similarly, Adobe's "agreement" with the government to ensure backward capability for twenty-five years should not be counted on when it comes to the long-term preservation of ETDs and, as is demonstrated later, should not be considered synonymous with a government-sanctioned preservation plan.

Recently, the GPO's use of PDF has become a major component of the Federal Depository Library Program. The distribution of government documents, traditionally characterized by monthly catalogs, orders, and a multitude of formats ranging from paper and fiche to floppy disks and CD-ROMS, has been a nightmare for many government documents librarians. Although the lack of an overarching government information policy continues to make the lives of government documents librarians difficult, PDF is making their lives easier, just as it is making the lives of records managers and others charged with organizing and disseminating information in the electronic environment simpler. Despite this use, PDF is a distribution format; paper remains the preservation format at the GPO.

What leads the authors to urge caution in accepting PDF as the permanent file format for document imaging stems from three sources—the government, record

managers, and private industry. As noted earlier, the government's record of accomplishment in preserving access to electronic information is not the best, and the advent of PDF is no reason to believe otherwise. At a conference jointly sponsored by the CEDARS project, the Online Computer Library Center, the Research Libraries Group, the U.K.'s Office for Library Networking (OLN), and the Joint Information Systems Committee (JISC), George D. Barnum of the GPO presented a paper entitled "The Federal Depository Library Program: Preserving a Tradition of Access to United States Government Information." In this presentation, Barnum stated that there is no reason to assume that the government will continue to rely on PDF.

Presentation of electronic publications that rely on an open standard ... will presumably remain straightforward as the Web and its successor technologies develop. Publications, however, that rely on a proprietary format or commercial software for their use pose serious challenges, since backward compatibility in newer technology will depend on market forces and demand. GPO cannot consider content separate from access and access mechanisms; thus the greatest challenge over the coming years will be to keep publications captured in 2000 viable despite the advance of technology. Transfer of all publications in the archive to a single, migration-friendly, open standard format has not, in the interest of preserving the official nature of the publications, been pursued thus far. Such transfer may, however, present itself as the best alternative for keeping archived publications alive.⁸

PDF was not mentioned once throughout the entire presentation. Barnum did, however, mention that the GPO's three guiding principles for pursuing electronic access were the trend in government to

adopt electronic media for communicating with the public, the rapid adoption of electronic media in libraries generally, and the clear direction of Congress to implement greater electronic access and to seek reductions in the cost of disseminating information. He then stated that, though the preservation of electronic files was important, the third reason, the reduction in distribution costs, was the most imperative force behind the push for electronic access.⁹

If this were not enough to give pause, the federal government is currently undertaking an interagency project involving twenty-five federal agencies. Its goal is to develop a Portable Document Delivery Format (PDDF) and a Federal Information Processing Standard (FIPS).

FIPS will provide a means for Government agencies to archive final form electronic documents in an open, transportable, format while maintaining document integrity. The importance of this achievement cannot be over-stated, since the availability of public domain non-proprietary software will enable virtually any Internet user to submit complex electronic documents (audio, text, graphics) in a form that can be retrieved in its original form with full retention of document integrity (no loss of format, content, color, etc.).¹⁰

Although the federal government is currently taking advantage of the blessing that PDF provides in widening access and reducing the cost of dissemination, it also is realistic about proprietary software products and the maintenance of long-term access to digital information while still preserving document integrity. The result is the government's effort to seek a permanent, nonproprietary software system that will permit simultaneous delivery and the preservation of document integrity.

Both inside and outside the government, record managers have long been

experimenting with the management of electronic files. A tour of the Association of Records Managers (ARMA) product floor at an annual conference five years ago would have led visitors to believe that very few companies were producing traditional micrographic equipment anymore—the products from Kodak, Canon, and other imaging companies were dominated by purely digital systems. Attendees at current ARMA conferences will see that the tables have actually turned back about ninety degrees. The imaging and records management industry now is dominated by the hybrid imaging system, dual output—electronic files such as PDF and other digital raster images for access and microfilm records for long-term preservation.

In a highly technical paper entitled "Permanent Digital Records and the PDF Format: Defining a Permanent Transform Record Management System, a Hierarchy of Record Storage Formats, Five PDF Formats, and Document Copying/Migration," Stephen J. Gilheany, a certified records manager and certified document imaging archivist, addressed the challenges of creating a long-term records management system.¹¹ Though speaking favorably of raster formats and, in particular, Adobe's PDF, Gilheany's model for successful management and preservation carried an underlying hint of caution. This vague caution centered on the necessity of preserving electronic documents in a multitude of formats, including native formats such as Microsoft Word for contextual information, raster format such as PDF or TIFF for access, the OCR output so that documents will be searchable, a structured format such as SGML or XML for preservation purposes, and an ASCII document to serve as a last-ditch method of recovering lost data. The paper also noted the existence of five PDF formats, some of which are more migration- and preservation-friendly than others.

The final reason for not accepting PDF, or any other single format, as a preservation tool, emerges from private industry. Pro Quest (formerly Bell & Howell Info-

Learning, formerly UMI) is primarily known to the academic community as a microfilm producer. Pro Quest films theses, dissertations, newspapers, books, serials, and other materials too numerous to name. Currently, it plans to receive the *New York Times* and a number of other major newspapers electronically. Using I-Beam Readers, the raster images they receive of these newspapers will be output to microfilm for sale and permanent storage. As Bell & Howell Info-Learning, the company's net earnings in 2000 exceeded \$275.2 million. The company maintains digital files of the materials it receives electronically; however, it does not trust the long-term accessibility of these electronic files.

Impact on the ETD Community

When specifying document format requirements for ETD programs, institutions have to balance ease of production for the student with ease of migration/retention for the institution. PDF documents are very easy for students to create. The appeal of a PDF-based ETD program is that students do not need to be highly computer literate to create this document format. The typical word processor a student uses to do writing can export the final document as a PDF with the push of a button. (Of course, the usability of this final document will depend on enhancements such as internal linking, proper embedding of unusual fonts, and other "details" that institutions specify in the instructions that students receive. Apparently, some PDFs are more equal than others.) PDF also has the advantage of being easily deliverable via the Web. The Acrobat Reader program is available as a free download and permits seamless viewing of PDF documents using any Web browser. PDF documents are among the most common document formats available via the Web.

Simple creation and delivery are very tempting features to an administration looking to start an ETD program from scratch. However, the authors believe that this choice is shortsighted. ETD programs

are new, and none have been tested in format migration, which is inevitable. Migration is likely to be an unfortunate, yet inevitable, reality. At present, a file format's ability to migrate *can be assessed* and should be a primary consideration when administrators make decisions.

The question that then arises is, Can an ETD program have it all? Several programs are investigating XML as the format of the future for ETDs. XML (eXtensible Markup Language) is a tagged ASCII text. Interpreted by a non-proprietary "browser" for maximum readability, plain-text readers such as Notepad read XML. XML can be difficult for the average student to produce and will initially require more hand-holding by institutions. However, for those universities interested in fostering true information literacy among their students, the hand-holding will have more lasting results than the push-button methods described above.

XML is very promising for ETDs, but other alternatives might be easier to implement in the short term. One possibility that would improve an institution's chances for migration in the future would be to require submission by the student of both the PDF (as a handy delivery format) and the "native" format used to create the original document (e.g., Word). The word processor file is still proprietary and thus is not the best candidate for long-term storage. Nevertheless, it can be converted to other formats much more easily than PDF can. If the campus has a standard program used for word processing, the documents could be converted in batch mode to each new version and, ultimately, to some other software in the future. At present, one format does not have to serve all purposes. PDF is a fine delivery format, but institutions need something else for long-term retention.

Dialog with the student should be a primary component of any ETD program and is necessary to discuss the rationale behind format restrictions. Requiring students to do extra work (such as submitting two versions of the ETD) might dis-

courage participation in programs. Providing options and explaining ramifications will shift some of the responsibility to the student author. A student might want to include a file type that might not migrate because the increased value for the document *now* is worth the risk. A program also might want to restrict these variations to appendices, asking that the student craft the document so that it can stand alone, should the appendix not migrate.

The final reason for planning is that many believe failure may not occur until after decision makers are out of the picture.

Such "optional obsolescence" might be OK theoretically, but it raises questions about the documents that become part of the permanent record. The document that is available in fifty years with no appendix is not the same document created by the author. In Kentucky, state law requires that the University of Kentucky preserve university records, including dissertations, in perpetuity. Consequently, this project presents administrators with the need to make format decisions with an awareness of the long-term implications. The institutional library or archives, as the unit frequently responsible for long-term retention of university documents, will be well in tune with the long-term preservation and access issues affecting ETDs, but other participants in the process can tend to focus on the other function of these documents, as a step completed in partial fulfillment of the requirements for a degree. In this view, the document has served its purpose and does not need to be retained forever. Institutions in Kentucky cannot afford to think that way.

Asset Management, or Short-term Benefits versus Long-term Costs

In the course of touring just about any preservation lab at a large institution, visitors see signs that say, "Think Twice; Cut Once." By urging caution and restraint in cutting boards and using supplies, pres-

ervation departments have generally managed to keep collections conservation, the preservation of library materials, a lower-cost alternative to the wholesale replacement of damaged and destroyed materials.

In the realm of electronic records and information, a similar adage applies. Alter it slightly, and the reader might imagine a sign that states, "Plan Properly; Plan Once." The cost of planning properly is time and, perhaps, a little bit of the prestige that accompanies being the first institution out there with the sexiest Web page for accessing ETDs.

The cost of failing to plan is far greater. In a publication recently released by the British Library's National Preservation Office, Seamus Ross of Glasgow University's Humanities Advanced Technology and Information Institute noted: "The short-term economic and productivity advantages offered by digital storage, manipulation, and communication encourages us to depend on them more and more. Although some are aware of the preservation risks, society in general is ignorant of them."¹² Ross then proceeded to outline the two digital preservation strategies that he believes most often characterize projects—a proactive approach and an approach represented by accident and rescue.

Regardless of what preservation strategy an institution chooses, preservation—whether characterized by a proactive approach or by stumbling into it—essentially centers on dealing with the results of institutional choice. Consequently, the costs must be considered.

In very black-and-white terms, the cost of planning preservation activities is great, but it consists primarily of time and resources devoted to developing an infrastructure capable of dealing with preservation activities. The other cost is, in the case of many ETD programs, the greater motivator—institutional recognition. In an academic world increasingly driven by the speed and competition that have long characterized the for-profit sector, being the second institution to develop a viable program is not good enough.

As institutions plan for ETD programs, they must remember three things. The first impetus for preservation planning is that the economic costs of not planning are greater than the initial outlay. At the 1995 meeting of the ISO Archiving Standards working group, participants reported that it cost between \$2.65 and \$3.75 per megabyte per year to retain electronic records created in the engineering sector, but about \$662.50 to reconstruct them if they were lost or destroyed. Oil survey records are even more costly to recreate. The National Archives of Australia (NAA) holds 600,000 computer tapes of oil survey data. In the early 1990s, the NAA estimated recreation of the offshore data at \$5,300 per meter, or \$5.3 billion in total.¹³ Of course, all the data cited above pertain to materials that could be salvaged from analog sources. As increasing volumes of materials exist in solely electronic forms, the potential cost rises.

The second reason for planning preservation activities is the danger that institutions face from a lack of memory. Print materials have served as the primary tools for historians and researchers for years. However, the advent of audio and video technologies has shown scholars that technologies are frequently fleeting. There is no reason to believe that nontextual digital formats will be any different. What they are is more complex. Frequently, hardware and software must interpret the data in question before ma-

nipulation or display. In their raw form, they are often meaningless.

The final reason for planning is that many believe failure may not occur until after decision makers are out of the picture. However, the probability that failure may occur after administrators leave is not an excuse. "Preserving digital assets cannot happen as an after-thought, it must be planned: media degrade, technical developments make systems obsolete, or information is rendered inaccessible by changes in encoding formats."¹⁴ Preservation requires active intervention. Unsecured, it is susceptible to loss through the physical breakdown of the media, rendered inaccessible by technological advances, or left meaningless through a lack of or insufficient contextual evidence.

Currently, a multitude of choices is available. The ETD community is faced with choosing between their obligations as educators and scholars and their obligations as members of the ETD community. One of those dictates that they seek to preserve and enrich human understanding of the world through traditional scholarship; the other dictates that they seek to preserve and enrich human understanding of the world through more contemporary scholarship. The authors do not believe that these are mutually exclusive concerns. Rather, the authors believe that administrators must simultaneously accept that ETDs are here and that the reality is that ETDs are far from perfect.

Notes

1. Richard Ficher and Charles Dollar, "File Formats to Support Long-term Access to Electronic Records," in *2000 Managing Electronic Records Conference Proceedings* (Chicago: Cohasset Associates, Inc., 2000).

2. Sally A. Buchanan, "Too Big, Too Expensive, Too Time-consuming," *Wilson Library Bulletin* 67 (Oct. 1993): 64.

3. Maggie Jones, "Preservation Roles and Responsibilities of Collecting Institutions in the Digital Age," in *1995 National Preservation Office (NPO) Conference Multimedia Preservation: Chasing the Rainbow* (Brisbane, Aus: National Library of Australia, 1995). Available online from <http://www.nla.gov.au/nla/staffpaper/npomj.html>. A print version is also available through the National Library of Australia. Ordering information is available with the online proceedings at <http://www.nla.gov.au/niac/meetings/npo95.html>.

4. Paul Conway, *Preservation in the Digital World* (Washington, D.C.: Council on Library and Information Resources, 1996). Available online from <http://www.clir.org/pubs/reports/conway2/index.html>.

5. Jeff Rothenberg, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation: A Report to the Council on Library and Information Resources* (Washington, D.C.: Council on Library and Information Resources, 1999), 3. Available online from <http://www.clir.org/pubs/reports/rothenberg/introduction.html>.

6. CEDARS (CURL Exemplars in Digital Archives) is an international, higher education effort charged with promoting awareness of the need for digital preservation, developing collection management strategies for digital materials, and investigating methods of digital preservation. Information on the CEDARS project is available online from <http://www.leeds.ac.uk/cedars/>.

7. Seamus Ross, "Changing Trains at Wigan: Digital Preservation and the Future of Scholarship," National Preservation Office Preservation Guidance Occasional Papers (London: The British Library, 2000): 3.

8. George D. Barnum and Steven Kerchoff, "The Federal Depository Library Program: Preserving a Tradition of Access to United States Government Information" (paper presented at Preservation 2000: An International Conference on the Preservation of Long-term Accessibility to Digital Material). Available online from <http://www.rlg.org/events/pres-2000/barnum.html>. Full conference proceedings are available online from <http://www.rlg.org/events/pres-2000/prespapers.html>.

9. *Ibid.*

10. Library of Congress, "Services of the Preservation Research and Testing Division" (Last updated 28 June 1999). Available online from <http://lcweb.loc.gov/preserv/resear.html>.

11. Stephen J. Gilheany, "Permanent Digital Records and the PDF Format: Defining a Permanent TransFormat Records Management System, a Hierarchy of Record Storage Formats, Five PDF Formats, and Document Copying/Migration." Available online from www.ArchiveBuilders.com.

12. Ross, "Changing Trains at Wigan," 6.

13. *Ibid.*, 4. Ross's paper originally cited the figures in Australian dollars. The authors converted the numbers to U.S. dollars. Figures are current as of March 2001.

14. *Ibid.*, 6.