Electronic theses and dissertation (ETD) repositories: What are they? Where do they come from? How do they work?
Kristin Yiotis

Article information:
To cite this document:
Permanent link to this document:
http://dx.doi.org/10.1108/10650750810875458

Downloaded on: 17 February 2016, At: 23:09 (PT)
References: this document contains references to 46 other documents.
To copy this document: permissions@emeraldinsight.com
The fulltext of this document has been downloaded 2084 times since 2008*

Users who downloaded this article also downloaded:

Access to this document was granted through an Emerald subscription provided by emerald-srm:515654 []

For Authors
If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com
Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.
Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.
Electronic theses and dissertation (ETD) repositories
What are they? Where do they come from? How do they work?

Kristin Yiotis
Sutter Health, Sunnyvale, California, USA

Abstract
Purpose – The purpose of this paper is to introduce the electronic theses and dissertation (ETD) repository as a subset of local institutional digital repositories. The paper discusses the originating institutions and organizations including Virginia Tech Initiative, the Networked Digital Library of Theses and Dissertations, the United Nations Educational, Scientific, and Cultural Organization and the United States Department of Education.

Design/methodology/approach – This paper is informational in nature and explores the topic of ETD repositories. It provides information relevant to academic and digital librarians interested in including an ETD repository in their institution’s digital library. The paper discusses interoperability among repositories and the Open Archives Initiative Protocol for Metadata Harvesting. The paper discusses issues related to ETD repositories including intellectual property rights, publishers’ views of ETDs as prior publications, plagiarism issues, development costs, and long-term preservation issues.

Findings – It was found that library administrators who implemented ETD repositories at various universities adapted their models to the needs of their institutions and their graduate students. ETD administrators made decisions about implementation models and software and hardware infrastructure in terms of human and technical resource allocation.

Practical implications – The paper argues that ETD repositories benefit students and universities by enhancing graduate education, expanding graduate research, increasing a university’s visibility, and instructing students, faculty, administration, and librarians about digital technology.

Originality/value – The value of this paper for digital and academic librarians concerned with EDT repositories is in providing a historical overview, a discussion of the benefits, and a review of the issues involved with implementing an ETD repository at their institution.

Keywords Database management, Digital storage, Theses, Academic libraries

Paper type Research paper

Theses, including doctoral dissertations, are a “cherished academic genre” of literature that American academic libraries have held in their collections for over a century dating back to the first dissertation submitted, “a six-page, hand written thesis at Yale University in 1860” (Moxley, 2001, Tradition and ETDs section, para. 2). “The quality of a university is reflected by the quality of its students’ intellectual products. Theses and dissertations reflect an institution’s ability to lead students and support original work” (UNESCO, 2001, 2.1 section).

Theses and dissertations from all academic departments become part of the university’s library collection. Once approved, students submit to the university library several print copies that the library binds, catalogs, and shelves. In general, circulation is limited to the university community, as most libraries do not lend theses...
and dissertations through interlibrary loan. Circulation records from Virginia Polytechnic Institute and State University (Virginia Tech) show that 15,335 theses and dissertations were approved between 1990 and 1994. In 1998, 3,967 were checked out, which is 3.86 percent of the total (Moxley, 2001, para. 2). In other words each one was checked out 0.0386 times.

Technological changes in the 1970s enabled students to produce theses and dissertations on electric typewriters, a vast improvement from the manual typewriters before that. By the 1980s the digital revolution arrived and word processing software programs eliminated the need to create carbon copies. Users could save copies of their work in digital format, edit their work before printing, and print multiple copies that were all originals.

When students want a wider audience for their work, they submit print copies at their own cost to University Microfilms Dissertation Abstracts (UMI). UMI, now owned by ProQuest, is the central repository and disseminator for North American print dissertations for the past 50 years (Moxley, 2001, Lost opportunities section, para. 4). UMI converted them to portable document format (PDF) files and made the copies available through their current research service.

Moving to electronic format
In 1987, UMI initiated a movement toward digital dissertations by calling a meeting to discuss the concept (Crowe, 1998, UMI section). In 1992, the Coalition for Networked Information (CNI), Virginia Polytechnic Institute and State University (Virginia Tech), the Council of Graduate Schools, and UMI started a project called “The capture and storage of electronic theses and dissertations” for the purpose of promoting applications and standards for capturing and storing theses and dissertations electronically (Crowe, 1998, Background section). One outcome was UMI’s ProQuest digital dissertations program that ensured digital formatting for the 1997 submissions. UMI began accepting submissions in electronic format and scanning and digitizing paper and microform submissions (Crowe, UMI section).

In December 2006, UMI current research service migrated to the ProQuest platform and was re-named Dissertations & Theses Full Text (Dissertations & Theses @, 2007). This subscription-based service provides free access to citations and abstracts in the ProQuest Dissertations and Theses (PQDT) database and allows users to view for free the first 24 pages of dissertations and download full text copies when available. Users can also pay per print or digital copy if their institution does not subscribe to the ProQuest service.

Another outcome was the CNI agreement to work with Cornell University, the University of Michigan, Penn State, and Virginia Tech to “develop and disseminate a standard method of using SGML (standard generalized markup language) to make dissertations available online” with initial funding coming from the Southeastern Universities Research Association (SURA) (Crowe, Background section). Virginia Tech submitted a grant proposal to the United States Department of Education to create the national digital library of theses and dissertations. Fox et al. (1996), the proposal writers, defined the problem: students receiving the 400,000 master’s and doctoral degrees awarded each year lacked basic information literacy skills and were unprepared for futures in which electronic publishing and accessing networked information systems would be commonplace (abstract). In addition, human and
material resources invested in generating and preserving theses and dissertations were wasted because access to them was “severely constrained” and “opportunities to unlock valuable university resources” were missed, “greatly limiting possibilities of knowledge transfer and re-use, and causing the whole academic enterprise to suffer” (Fox et al., 1996, Abstract).

The proposal called for the creation of the Networked Digital Library of Theses and Dissertations (NDLTD, 2006), a federation of member institutions and organization that would “support and encourage the production and archiving of electronic theses and dissertations” (Suleman and Fox, 2003, Abstract). They argued that repositories for electronic theses and dissertations (ETDs) would benefit the entire university and further its academic enterprise by enhancing graduate education, preserving and disseminating knowledge, and advancing current technology to support multimedia publishing in digital libraries (Suleman and Fox, 2003, Introduction section). Funding of more than $1,200,000 from the Department of Education and Corporations, including Adobe, IBM, and Microsoft, enabled Virginia Tech to become the lead institution in promoting ETDs as a new genre of literature that, according to the developers, would have far reaching effects on graduate education (Crowe, 1998, Virginia Tech Initiative section).

**NDLTD and UNESCO support ETDs**

NDLTD focused on three main issues: improving graduate education by making students information literate, “developing and testing models . . . to arrive at standards for document formats and interoperability,” and on encouraging institutions to join NDLTD in free membership (Crowe, 1998, Virginia Tech initiative section). In 1999, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) became interested in ETDs as resources that promote access to information in the public domain through the use of the internet. UNESCO’s mandate, “to ensure the ‘free exchange of ideas and knowledge’ was consistent with NDLTD’s goals for ETDs” (Smith, 2002, ETDs are here section). UNESCO:

> [...] hosted an ETD workshop to discuss the strategy for an international ETD initiative [...] in the interest of developing an international project that facilitated transfer of expertise from developed countries to developing countries in scientific areas [...] (Smith, 2002, ETDs are here section).

In 2000, UNESCO supported Virginia Tech in initiating an online guide of best practices for developing ETD programs. This 420-page document, The Guide for Electronic Thesis and Dissertations, (http://etdguide.org) collects articles by experts from all over the world, many previously published, on practical information and best practices for all areas of ETD development and implementation. This manual serves to document the ETD initiative and to address the concerns of universities hosting ETD repositories and students writing ETDs (UNESCO, 2001, Contents section).

**Benefits of ETDs**

The primary mission of NDLTD is to help graduate students become information literate by making ETD an institutional requirement. According to NDLTD, the underlying purpose of ETD activity is “to prepare the next generation of scholars to function effectively as knowledge workers in the Information Age” (Fox, 2001,
Purpose, goals, objectives section). NDLTD’s founders envisioned a worldwide program of ETDs that “enhance graduate education, promote sharing of research, and support university collaboration” (Fox, 2001, Purpose, goals, objectives section). These goals would be achieved through specific objectives:

- students knowing how to contribute to and use digital libraries;
- universities developing digital library service;
- worldwide sharing of university research; and
- a higher quality and greater expressiveness of graduate theses and dissertations (Fox, 2001, Purpose, goals, objectives section).

Students, then, are the primary beneficiaries of ETDs. Fox (2001, Students section) writes, “Students are the most important participants in ETD activities”. By learning to use and promoting ETDs, students will benefit in two ways: “The first benefit is that new, better types of [theses and dissertations] (TDs) may emerge as ETDs develop as a genre” (Fox, 2001, Students section). Moxley (2001) sees a future where “creative researchers will challenge our conception of academic writing. Linear text with one-inch margins will give way to hypertextual writing, streaming multimedia, interactive chat spaces, three-dimensional modeling…” (Providing the tools section).


The second benefit to students is the improved visibility and increased exposure their work undergoes when placed in an electronic repository. ETDs showcase the intellectual achievements of a university by making them available to a worldwide audience. A university’s ETD, according to Moxley (2001, para. 2), “raises significant interest in the work of its graduate students.” “Graduate students and their sponsoring faculty could benefit from increased exposure of their work, both in job and other financial opportunities and in profession reputation” (Moxley, 2001, para. 1). ETDs are more accessible than traditional theses and circulate more. In 2000-2001, when Virginia Tech had 3,393 ETDs in its collection, 1,565,151 PDFs were downloaded by users, meaning that on average each ETD was downloaded 461 times (Moxley, 2001, para. 2).

Universities also potentially benefit from ETD repositories. Moxley (2001), writing in the UNESCO Guide, reason:

Theses and dissertations reflect an institution’s ability to lead students and support original work […] As digital libraries of ETDs become more commonplace, students and faculty will make judgments regarding the quality of a university by reviewing its digital library (2.1. Why ETD’s? section).

Moxley (2001, para. 1) states, “In future a university’s quality will be linked to its digital library of theses and dissertations…”

UNESCO (2001) envisions that the benefits ETDs have on students will reflect on universities and on entire societies, countries, and regions (2.1.1., Reasons and strategies section). Benefits overlap from student to university to society in the following ways (UNESCO, 2001, 2.1.1., Reasons and strategies section):
• Scholarship builds on scholarship such that increased access to information and research enhances the quality of theses and dissertations and knowledge in general.

• ETDs are a way of sharing intellectual production because they make the results of graduate programs widely known.

• Theses and dissertations (TDs) present the methods used during research, thus allowing these methods to be used by others.

• To electronically publish TDs makes the results known nationally and internationally, and ETDs can identify and connect national and international research groups.

• Wide knowledge of good quality TDs strengthens the faculty, the graduate programs, and the university such that graduate programs may be evaluated by the number of theses and dissertations that are accessible electronically.

• TDs are part of the assets and of the history of the universities. Since they are published on paper, why not publish them electronically where they require less storage space?

• In countries where theses and dissertations are financed by public funds, authors are expected to make their work public, and ETDs are the easiest way to accomplish this.

• An ETD program introduces digital libraries in the university allowing other projects to bloom.

In a SPARC position paper titled, “The case for institutional repositories,” Crow (2002) discusses other benefits that institutional digital repositories, which include ETDs, bring to libraries:

Establishing an institutional repository program indicates that a library seeks to move beyond a custodial role to contribute actively to the evolution of scholarly communication […] Institutional repository programs promise libraries an extraordinary level of visibility within the university […] The library’s relevance to the faculty – and, consequently, the institution overall – will increase. (Crow, 2002, Impact of institutional repositories section, para. 1).

How ETDs get started
ETDs are basically a subset of an institution’s local digital repository. An institutional repository (IR), as defined by Crow (2002), is:

[...] a digital archive of the intellectual product created by the faculty, research staff, and students of an institution and accessible to end-users both within and outside of the institution, with few if any barriers to access (Essential elements section).

Academic libraries usually host the institution’s local repository within their digital library, such as the California Digital Library (CDL), which is the IR for the libraries of the ten University of California (UC) campuses.

Having an existing digital library, which most large research universities have, makes it easier to institute an ETD repository. Since 2002, George Mason University’s Fenwick Library has maintained the Mason Archival Repository Service (MARS).
According to D. Salo, Digital Repository Services Librarian, the library is currently in the process of creating an ETD (personal communication, April 19, 2006). Fenwick Library has submitted a proposal to the university administration to institute and maintain an ETD through MARS. Their proposal reviews the limitations to the current method for submitting theses and dissertations and recommends a new policy that requires one electronic copy in addition to one paper copy stressing that for the present the new policy will not eliminate the traditional paper version of the scholarly paper (D. Salo, personal communication, April 19, 2006).

Johns Hopkins University, Sheridan Libraries manages the Library Digital Programs (LDP), which includes digital services, digital collections, and research and development initiatives. One initiative is an ETD. The Sheridan Library web site explains that they have initiated a pilot project to study the various electronic publishing systems that host ETDs, the current systems being: DSpace, ePrints, DPubS, and DiVA (Services, collections, & projects, 2005, para. 4). The site supplies a sign-up sheet for those interested in participating in the pilot project, saying that students from all departments may participate but that participation does not replace the obligation to submit a paper version (Pilot program for ETDs, 2006).

Some universities populated their ETD repository by instituting an electronic copy theses requirement in certain departments only. Vanderbilt University conducted a pilot program in which certain department participated on a voluntary basis, but within those departments electronic submission was mandatory (Smith, 2002, p. 5). University of Kentucky also gave their students the option of submitting in electronic or traditional format (Smith, 2002, p. 5). As of this writing, University of Kentucky still offers the electronic thesis or dissertation as an option.

Other universities have made electronic copy submission mandatory for students and have done away with the paper copy all together. Virginia Tech began requiring all graduate students to submit electronic theses and dissertations in January 1997 (Seamans, 2003, p. 56). West Virginia University implemented a mandatory submission policy based on the idea that ETDs “contribute to worldwide graduate education and provide a means of ‘unlocking’ the under utilized results of academic graduate research” (Smith, 2002, p. 6). Smith, quoting the Library Technical Consultant for West Virginia University Libraries, writes:

The success of the ETD program has helped to create a “heightened sense of awareness on campus of the profound effects of information technology,” and that this has helped to bring a whole host of IT developments to the West Virginia campus (Smith, 2002, p. 6).

Smith (2002, p. 6) stresses that all the universities he researched offered workshops to students to help them deal with the technical issues related to submitting theses and dissertations in electronic format. He says that the methods established by NDLTD and the Virginia Tech UNESCO Guide reduced the difficulty of developing from scratch the technological expertise needed to teach faculty, students, and administrators, and libraries about ETDs (p. 6).

Massachusetts Institute of Technology’s (MIT) Theses in DSpace (2002) repository contains “selected theses and dissertations from all MIT departments [and] contains approximately 10,000 theses completed at MIT between 1879 and the present” (para. 1). The web site explains that after 2004 “all new Masters and PhD theses will be scanned
Choosing software and implementation models

ePrints is the original, open-source, repository software developed at the University of Southampton in Great Britain. The ePrints solution, as explained by Tennant (2002), “is squarely focused on the faculty working paper, (called preprint or e-print) . . . and assumes that faculty will directly upload their own prepublication . . . via an institutional . . . repository,” a model that is currently being used at CalTech and the Digital Library of the Commons at Indiana University (Software section). MIT developed DSpace as a digital repository technology that is considered by experts to be more flexible and robust than the ePrints software because “it makes fewer assumptions regarding what type of object is being uploaded” (Tennant, 2002, Software section).

DSpace is currently being used in institutional repositories all over the world, as shown on the DSpace Instances web site. A quick look shows ETDs at the following institutions: DSpace@SLU (Saint Louis University, Phillipines); DSpace at Ural State University, Russia; ETD of Indian Institute of Science, Bangalore (etd@IISc), India; MSpace at the University of Manitoba, Canada; Nagoya Repository, Nagoya University, Japan; OdinPubAfrica (15 African nations); Oregon State University (USA); QUEprints Cranfield University (UK).

Tennant (2002), at the CDL, discusses implementation models as a next step after choosing software in creating an IR. “There are nearly as many models as there are institutional repositories,” Tennant (2002, Implementation models section) states. The three examples he focuses on, distributed, semidistributed and semicentralized, have to do with management and assigning uploading responsibilities.

In the distributed model, the original self-archiving model that Harnad and others had in mind when starting the Open-Archives Initiative, individual faculty upload their own scholarly output into the repository through a software platform. Institutions that follow this model, such as Virginia Tech, provide support in the form of workshops, tutorials and submission instructions at the web site. The DSpace software platform has the broadest application in terms of being able to handle any educational material in digital format, such as lecture notes, visualizations, simulations, original graphics, datasets, images, thereby enabling faculty to make full use of the repository. The ePrints platform was designed specifically for prepublished manuscripts. In terms of ETDs, the distributed model enables students to manage their own formatting issues, such as making PDF files, and directly upload their electronic copy.

The semidistributed model “assigns management responsibilities to organizational units . . . that . . . assist faculty [and students] with uploading their papers” (Tennant, 2002, Implementation models section). The CDL’s eScholarship repository, explains Tennant, uses this model. CDL describes its eScholarship Repository as a “free, open-access repository infrastructure [that] provides UC departments direct control of publishing scholarly materials such as postprints, journals and peer-reviewed series, and seminar papers” (Collections & services, 2006). The eScholarship web site lists the library liaisons who are available to assist faculty and students for each UC campus (Campus eScholarship liaisons, 2005).
The semi-centralized model assigns responsibility to the library to set up and manage a repository site for any university unit and to upload papers on behalf of faculty. California Technical Institute (CalTech) uses this model for their Collection of Open Digital Archives (CODA, 2006), their generic IR that holds faculty research papers and more, and runs on the ePrints software platform (Tennant, 2002, Implementation models section).

The CalTech ETD repository requires students to submit theses and dissertations online and walks them through all aspects of the submission sequence from thesis regulations to faculty approval (CalTech Library Services, 2006). The CalTech ETD repository uses the ETD-db open-source software platform developed for theses-specific repositories at Virginia Tech. Students are provided with information and instructions for walking through the entire submission process. An ETD database developer’s resource is available for students to implement their own ETD databases and includes information about the technologies used to implement the ETD-db (CalTech ETD-db, 2006). The developer’s information and instructions are identical to those at the ETD homepage at Virginia Tech and appear highly technical to the novice.

Wong (2006) discusses the software platforms the City University of Hong Kong Library (CityU) considered for their ETD repository and their IR. For the ETD repository, options included the E-thesis-specific ETD-db, developed at Virginia Tech, or generic repository software, such as DSpace or ePrints (Wong, 2006, 5. E-theses in CityU IR section). Virginia Tech’s software, ETD-db, contains a submission module, a theses-specific metadata set, and a workplace for faculty supervision and approval. DSpace, a generic repository platform, is “better supported ... in terms of data structure, security, administration and future expansion” (Wong, 2006, para. 6). CityU Library chose DSpace for their ETD repository based on Wong’s conclusion that “the ETD-MS [Metadata Standard] supported by ETD-db [was] not totally OAI compliant [which] may create an obstacle in metadata harvesting,” the platform they also chose for their IR (Wong, 2006, para. 6).

CityU Library made use of a comparison study between ETD-db and DSpace platforms conducted by Edinburgh University Library while in the process of developing Theses Alive! “a software solution for institutions in the UK to implement their own E-theses or ETD online submission system and repository” (Jones, 2004, Introduction section, para. 1). The Library compared the two platforms in the following areas: general overview, submissions procedures, archiving and access, administration and security. They conclude, “DSpace is clearly ahead of ETD-db. It is a well-supported package with a future that is being planned now, while ETD-db has been dormant for some time” (Jones, 2004, Overall, Conclusions section, para. 1). The ETD-Metadata Standard current version, dated May 6, 2001, is available at the NDLTD web site (2004, Document history section).

Protocol and interoperability
Whichever software platform a repository chooses, all adhere to standards of interoperability and metadata harvesting laid down in 1999 at the Santa Fe Convention as the Open Archives Initiative. At this convention, various organizations, including NDLTD and the Networked Computer Science Technical Reference Library, met to develop a solution for large-scale interoperability between digital repositories (Suleman and Fox, 2003, Early efforts section). Today the Open Archives Initiative
Protocol for Metadata Harvesting (OAI-PMH) is an interoperability framework “based on HTTP, which facilitates the … transfer of metadata among networked systems” (Suleman and Fox, OAI protocol section).

The OAI-PMH framework allows for two types of participation: the data providers at the input end and the service providers at the output end. Each institution’s ETD functions as a data provider. However, unless these disparate repositories are joined in some way, each one must be searched independently. The protocol is ideas for created distributed ETD repositories searched through a centralized networked digital library, the original vision for the NDLTD. The metadata harvesting capability, built into the OAI-PMH standards, enable OAI service providers to collect or “harvest” metadata from all repositories into a central site and then republishing it as a single collection. The NDLTD Union Archive is the centralized site for NDLTD members. Any local ETD site, then, can harvest from the NDLTD Union Catalog using its own search and browse capabilities (Suleman and Fox, 2003, NDLTD Union Archive section).

Interoperability requires that repositories use standardized metadata tags, such as SGML/XML and other mark-up languages that assign unique identifiers to items within a repository. OAI-PMH requests use the identifiers to extract metadata from the item (OAI-PMH: Definitions and Concepts, 2004). Service providers, like the NDLTD Union Archive, are able to harvest, index, and search metadata from remote sites, store it, and then republish this metadata through their own OAI data provider interface (Suleman and Fox, 2003, NDLTD Union Archive section).

Registered service providers (2005) lists registered providers of search service based on harvesting metadata using the OAI metadata harvesting protocol. Registered data providers (2007) lists the 610 repositories that conform to OAI standards. These are the repositories that the service providers search. ARC – A cross archives search service (2004), the first federated search service based on OAI-PMH, is an experimental service supported by Old Dominion University, Digital Library Research Group. As of this writing ARC is not currently running. All service providers like ARC and Citebase, another experimental search service run by the University of Southampton, harvest metadata from the OAI the data providers, including institutional repositories, ETD repositories, and open access journals (Citebase, 2007).

OAIster, a service provider based at the University of Michigan Digital Library Production Service, makes accessible “collections of freely available, previously difficult to access, academically oriented, digital resources” (OAIster, 2006). OAIster’s homepage says it provides access to 11,003,628 records from 754 institutions’ open archive repositories, some of which are ETDs. OAIster allows users to search and browse by citation, keyword, or by IR with each one named and its contents briefly described, whereas ARC allows searching by citation or by IR without names given in full or any descriptions. OAIster is one of the three official search services that extracts thesis and dissertation records from OCLC’s XTCat Experimental Theses Catalog (XTCat, 2002). WorldCat and the NDLTD Union Catalog are the other two (XTCat, 2002). As discussed earlier, the NDLTD Union Catalog also enables federated searching of the ETD repositories of the 174 member libraries of the NDLTD consortium (Browse/search ETDs, 2006).

Starting this spring UMI ProQuest will institute an Open Access publishing option for dissertations and theses, in addition to their traditional publishing option. For an increased fee, graduate students can choose to publish in PQDT Open, a new ProQuest
Key issues and concerns
Virginia Tech’s Fox (2001) discusses five key concerns about ETDs that each institution has to address: ownership of property rights, what access is allowed, how ETDs relate to publishers, the issue of plagiarism, and the matter of cost (what are key concerns section). Similar concerns are brought up by other institutions. The University of Kentucky’s concerns include ownership issues including how ETDs relate to intellectual property rights lost to publishers, plagiarism issues, costs of software and hardware infrastructure, and long-term preservation issues (Smith, 2002, p. 12).

Institutions are concerned that publishers will not publish articles, chapters, or books derived from ETDs, and that students, advisors, or funding sources will lose some of their full rights to their material. Fox (2001) discusses ownership rights of ETDs as resting with authors in most institutions but some institutions request or claim ownership. Where research has been sponsored, the funding agency may claim rights (What are key concerns section, para. 2). Property rights for ETDs would not be different than print theses simply because they are deposited in a digital repository owned by a university; in other words authors do not give up property rights because the university owns the repository in which the TD is deposited.

The issue of how ETDs relate to publishers has been under study. Seamans (2003) researched “whether or not ETDs would be viewed as prior publications and would . . . be ineligible for consideration for publication in traditional journals” (Abstract). She reports on a survey of publishers that asked editors about ETDs as prior publications. The survey discovered that “94 percent of . . . respondents stated that the journal had a policy on prior publications . . . but that 68 percent . . . stated that these policies did not specifically refer to works . . . posted on the Web or made available electronically” (Seamans, 2003, p. 56). Publishers most frequent comment was about the differences between a thesis or dissertation and a work derived from one, “regardless of whether the [original] work was in electronic or paper format” (Seamans, 2003, p. 60). Publishers reported that authors usually had to rewrite dissertations or theses before they could be accepted for publications as books or articles (Seamans, 2003, p. 60). So publishers do not necessarily see an ETD as an obstacle to a future publication based on the earlier work.

Publishers’ attitudes toward ETDs directly relates to the access level allowed, whether access would be limited in any way. Seamans (2003, p. 56) suggests four levels of access ranging from “worldwide access to the entire document to securing the entire work with no access allowed to any part of the document.” Students are afraid that if publishers considered an ETD a prepublished work, this would hurt their chance of publishing their work later in a journal or as a book. Fox’s (2001, What are key concerns section) suggestion to students who are working on a book is to limit access to the university community and to discuss the issue with a publisher before posting
the ETD. Fox claims little evidence that “public access to an ETD will hurt future sales of an eventual published book” (What are key concerns section, para. 4). Fox also discusses students’ responsibility to get permission from publishers for content in their ETD that is similar to an earlier published work (What are key concerns section).

Another concern is plagiarism. It is easier for students to plagiarize within their ETD and from others’ ETD because of increased access to electronic documents and the copy and paste features. Search features, however, make detecting plagiarism easier as well. Every university has policies in place regarding plagiarism and these will have to be enforced as well as fair-use restrictions (Fox, 2001, What are key concerns section).

The concern with costs includes increased financial and work burdens on students and academic departments. Additional university personnel may be needed to “propose, publicize, initiate, refine, and institutionalize the [ETD] activities” and preparation (Fox, 2001, What are key concerns section, para. 6). If “ETDs instead of paper TDs are required, there should be net savings relative to old processing methods,” claims Fox (2001, What are key concerns section, para. 6). But if paper copies are still needed by theses advisory committees, ETDs increase rather than decrease the burden of work for students and increase expenses for students and departments if software and equipment must be purchased (Smith, 2002, p. 12).

The UNESCO Guide (2001, p. 87) states start-up costs as relating to infrastructure and training, such as whether or not an institution already has a digital library in place and if the network infrastructure is provided by the host institution. Start up involves costs in terms of human resources, infrastructure, and training. Human resources include professionals, technicians, and management, with the professionals creating the procedures and developing the tools, the technicians uploading theses and assisting students, the management supervising day to day and communicating with university administration (p. 88). Infrastructure includes the actual server and software with the server site usually provided by the university, and materials, such as a workstation PC, printer, and software provided by the library (p. 88). Training involves training team members, organizing training, and developing training manuals, tutorials, and documentation (p. 88).

There is the issue of preservation and the lack of long-term electronic archival standards. Will the ETD be accessible in the long-term, or even the short-term, future? Digital preservation is a national concern that is being worked on at the highest levels of government. In 2005, Lockheed Martin was awarded a $308 million contract by National Archives and Records Administration to build a permanent archives system that will preserve and manage electronic records created by the federal government. “The Electronic Records Archives (ERA) system will capture electronic information – regardless of its format – save it permanently, and make it accessible on whatever future hardware or software is currently in use” (Martin, 2005, para. 3).

The Government Printing Office (GPO) has accepted Adobe PDF as the preferred document format. Most ETDs have also accepted “PDF as a low-cost solution for the delivery of electronic documents” because of the reduced training and minimum start up costs involved (Smith, 2002, p. 7). Nobody is sure that Adobe can guarantee the long-term retention of its PDF format. A Preservation 2000 International Conference statement by the GPO claims, “publications . . . that rely on a proprietary format or
commercial software . . . pose serious challenges . . . since backward compatibility in newer technology will depend on market forces and demand” (Smith, 2002, p. 8).

The GPO suggests transfer of all publications to a single . . . open standard format such as HTML for text and TIFF for images. ETD operators have begun looking into SGML and XML (extensible markup language) as alternatives to PDF but there are disadvantages (Smith, 2002, p. 8). SGML was developed first but is argued to be too complex to be easily adopted (Smith, 2002, p. 8). XML, developed in 1996 by the World Wide Web Consortium (W3C) as a new standard, takes the parts of SGML that are relevant to the internet (Smith, 2002, p. 8). XML has become as the language of choice for transmitting data in a standardized format on the Web; e-commerce sites for example use XML (Smith, 2002, p. 9).

Several institutions have explored XML for ETDs. Virginia Tech, for instance, created an XML tagging schema for encoding ETDs, but there are still major concerns. One is universality – there is not one schema that covers all the possible elements that may be found in theses and dissertations in all disciplines, each element such as the math and science symbols requiring a unique tag. Another problem is the steep learning curve for XML when compared with PDF. Departments would need to invest resources and time in training graduate students to write an XML ETD (Smith, 2002, p. 10).

Summary
This paper attempts to introduce ETDs and to build the case that ETDs add value by enhancing graduate education, expanding graduate research, increasing a university’s visibility, and instructing students, faculty, administration, and librarians about digital technology. Issues discussed include the history of ETDs at various universities, implementation models and software choices ETD administrators must make, concepts involved in understanding ETDs such as protocol and interoperability, and key issues and concerns such as intellectual property rights, publishers’ views of ETD as prior publications, increased ease of plagiarism, costs in terms of human and technical resources, and long-term preservation.

References


Further reading


Corresponding author
Kristin Yiotis can be contacted at: kyiotis@yahoo.com

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints
This article has been cited by:


6. Stephen Pinfield, Jennifer Salter, Peter A. Bath, Bill Hubbard, Peter Millington, Jane H.S. Anders, Azhar Hussain. 2014. Open-access repositories worldwide, 2005-2012: Past growth, current characteristics, and future possibilities. *Journal of the Association for Information Science and Technology* n/a-n/a. [CrossRef]


8. Qurratu'Aini Adibah Ahmad Fazil, Zuraidah Abdullah, Shahrul Azman Mohd. Noah. Applying Zachman Framework to determine the content of semantic theses digital library 1596-1600. [CrossRef]