

# REPORT OF THE AD HOC COMMITTEE ON ELECTRONIC THESES AND DISSERTATIONS

June 2000

As we progress into the electronic communications and information age, availability of scholarly material in electronic format will become increasingly important. Several universities have begun accepting, and even requiring, submission of graduate student theses and dissertations in wholly electronic formats. Virginia Polytechnic Institute and State University, for example, has *required* that all theses and dissertations be submitted in electronic format since 1997; West Virginia University since Fall 1998. Universities who are planning or have instituted the use of Electronic Theses and Dissertations (ETDs) join the Networked Digital Library of Theses and Dissertations (NDLTD), an umbrella group that works together to maintain and make available to the entire world their networked archives of ETDs. As of May 2000, the NDLTD currently has 64 member universities, including 20 Research Class I universities and 8 of UK's current benchmarks – Florida, Georgia, Iowa, North Carolina State, Penn State, Texas, Virginia, and Wisconsin. Twenty-eight of the members are international – in Australia, Canada, England, Germany, Hong Kong, India, Italy, Korea, Mexico, Russia, Singapore, Spain, and Taiwan.

As a result of discussions with a number of graduate school deans and with UK faculty and directors of graduate studies, the Graduate School decided to embark on a pilot project that will provide graduate students at the University of Kentucky the *option* of submitting their research and scholarly work in the form of wholly electronic theses and dissertations, commonly referred to as ETDs.

After consideration and approval by the Graduate Council, Michael T. Nietzel, the Dean of the Graduate School, formed an *ad hoc* committee in January 2000 whose objective was to discuss, design and begin to implement a system that would permit graduate students at the University of Kentucky the option of submitting theses and dissertations in wholly electronic format. The charge was open-ended, and had very few boundary conditions:

- Participation of graduate students was to remain voluntary
- Both master's theses and doctoral dissertations were to be included
- The system designed was to be as simple, easy, and inexpensive for the student as possible
- This Pilot (or initial) Project should be in operation by the start of the Fall 2000 Semester

This Committee met frequently during the Spring 2000 semester. This document contains the broad recommendations of the Committee (sometimes referred to as "we" in the text) to the Dean of the Graduate School for the system to be implemented at the University of Kentucky. We are cognizant of the fact that administrative decisions and the availability of resources will greatly affect the details and procedures of the overall system to be implemented.

We further note, at the start, the full and active participation of the University Libraries along with the Graduate School, without which this initiative would not be possible or would have taken considerably longer to achieve. At a couple of universities, the initial impetus for ETDs came only from their library system, and encountered passive or apparently even active resistance from academic administration. At a couple of others, the reverse happened. At the University of Kentucky, we have been fortunate to have both of the central core groups working cooperatively, pulling in the same direction at the same time to bring this important advancement in scholarly communications to the University community.

## Advantages of ETDs

There are a number of exciting advantages to ETDs.

- **Broader exposure** of the research and creative efforts both of the individual student and of the entire academic program involved. The thesis or dissertation is not housed in the dusty corner of a university library.
- **Worldwide, any-time/any-place access.**
- **Faster availability.** Theses and dissertations can be displayed on the Web in a matter of days or a couple of weeks, not months, as is the case now with the paper and UMI microform versions.
- **Master's theses also** can readily be displayed, not just doctoral dissertations, as is currently the practice.
- **Electronic searchability of the full text** of the thesis or dissertation, depending on the format used, not just of the keywords and abstract.
- **Enhancement of job opportunities.** The added student experience of preparing wholly electronic documents and a working familiarity with several different types of software are increasingly critical skills in an electronic age for any professional.
- **Less expense for the student,** depending on the infrastructure support. No rag copies, at a duplicating cost of 20 cents per page, are prepared. Hi-resolution digital images, maps for example, can replace very costly large-size color photographs.
- **Optional multi-media enhancements** such as pictures, film or audio clips, interactive elements, and spread sheets can at times provide a much richer and more valuable scholarly document.
- **Opening scientific and other scholarly information to the three-quarters of the world that can't afford it now.** For example, the total budget of Tunisia for scientific journals is about \$50-60 thousand per year; this will buy only some 50-60 periodical journals across all of science and technology. The University is able to expand its service to all of the resource-poor nations of the world.

## Concerns about ETDs

There are a number of concerns in different academic constituencies about ETDs, particularly among those who are as yet unfamiliar with this form of scholarly communication.

- **Will Publishers publish** articles, chapters, and whole books that are derived from material that is freely available on the Web? Could economic, copyright, and prestige issues result in reluctance or outright refusal on the part of publishing houses or professional societies to also publish this material?
- Possible **loss of intellectual property rights**. Will students, their advisors, and any funding sources lose some of their full rights to the material and possible benefits derived therefrom?
- **Plagiarism is easier** with electronic originals, both in the ETD itself and from it in other's works. (But it is also easier to detect.)
- **Long-term electronic archival standards and software convertibility are not yet established**. Will the ETD still be accessible in 50, or even 10, years in the future? This is clearly a problem that has not been solved.
- **All fair-use and copyright requirements must be observed very strictly** by the student. Although these requirements are legally no different from those for theses and dissertations submitted in paper medium and are obligations of any scholarly activity, the much easier and greater accessibility of ETDs makes these issues of heightened concern.
- **Intermediate paper copies** are almost always necessary for the Advisor and Advisory Committee to read and edit. This could result in a greater, not lesser, burden of work for the student at perhaps the most critical time in his or her graduate career.
- **Insufficient infrastructure support** could result in higher, not lower, expenses for the student and/or the academic program in purchase of needed software and equipment, as well as in increased work and frustration in learning how to use new softwares.

The Committee understands and appreciates these concerns on the part of many individuals and groups and that they cannot be blithely dismissed. However, we feel some that some concerns are not based in reality or have diminished in importance over the past few years as the concept of ETDs has spread more widely. Others can be eliminated or minimized in a well-designed and institutionally supported ETD system. A few are indeed quite valid, and of concern and active discussion within the ETD community. Therefore, the Committee strongly **recommends** that the concept and practice of ETDs be adequately explained to the University community, especially to the students, faculty, and academic programs who will be involved. ETDs must be willingly accepted and embraced by the academic community. Education is the key, not coercion.

## Major File Formats

No single other issue raised so much discussion as the issue of which major file formats (sometimes termed *schema* or *platforms*) would be accepted during the initial pilot project. That is, what software systems would be acceptable for the body of the dissertation? The discussion was quite lively and even intense at times. This is an exceptionally important choice because choosing the allowable format(s) essentially defines the nature of the entire ETD system. In this case, the medium is the message.

Choice of file format is based on several factors including ease of creation of the document, searchability, and long-term archiving (migration) issues. The two most common file formats used by universities with active ETD programs are PDF and XML. Universities that permit or require ETDs vary in terms of which format is permitted and whether more than one format is permitted.

Documents produced in or converted to Portable Document Format (PDF) are easy to create, easy to serve via the web, and maintain pagination when printed. They also maintain the physical formatting or layout intended by the author for items such as graphs and illustrations across different platforms used by those accessing the document. One prepares a document using a word processor for which there is a readily available converter (such as Adobe Acrobat Distiller) to PDF format. For example, Microsoft Word allows for such conversion, and the output from other Windows- and Macintosh-based word processors (such as WordPerfect and Canvas) can also be converted. In UNIX, the software for conversion of PostScript to PDF is also available. This produces essentially an electronic version of a paged document.

The main advantages of PDF are its simplicity, ease of use, and the low cost of the software (currently about \$80 for the full Adobe Acrobat suite). One can take files generated on the latest versions of several popular word processing programs and, almost, but not quite, “push a button” and the file is converted into PDF format ready for uploading onto the Web. Students already using modern word-processor software need only one or two 2-3 hour instructional sessions to learn how to convert documents into PDF format. Its main disadvantages are limited searchability of the full text (relative to markup formats), that it is not truly “electronic” in nature (it largely emulates paper pages, but with some added features), and that it is proprietary software. Adobe does provide a free “reader” application so that users do not need to purchase software to read PDF files, which makes using this proprietary format sensible. Adobe has also assured the ETD community that it will be around for the long haul, but we have no guarantee. The U.S. Library of Congress supports PDF, which gives us reasonable certitude it will not disappear. Thus it seems to us highly unlikely that PDF will “go away” at some point and we will have a major migration problem to contend with.

Searchability is enhanced for ETDs in PDF format by requiring the student to prepare and submit a “metadata file” in SGML format. This file contains the bibliographic description of the ETD (author, title, date, keywords, etc.). The metadata file is created by a web-based form page (template) that the student completes when submitting an ETD.

PDF is the most common file format accepted by current ETD operations. Because of reduced training requirements and fewer technical start-up requirements, PDF also enables a more rapid initiation of a new ETD operation. West Virginia University, for example, has required that all theses and dissertations be submitted in electronic format since Fall 1998, after only a 1-year test period, and will accept only PDF documents.

The UK Department of Computer Science has already begun to display the work of its students on the departmental web site <[www.cs.engr.uky.edu](http://www.cs.engr.uky.edu)>. To date, two dissertations have been posted in PDF and postscript formats.

The second type of major format, the markup languages, comes in greater varieties and degrees of richness and complexity. Virginia Tech, for example, will accept documents in Standardized Generalized Markup Language (SGML), which they have been developing for about 5 years or in eXtensible Markup Language (XML), which is a somewhat simpler subset of SGML. A consortium of German universities, the country from which dissertations arose historically, has decided that their dissertations will be submitted in XML. The main advantages are a greater degree of searchability and “mining” of the body of the ETD, greater richness of options and usefulness, truly electronic nature, and non-proprietary file format. The major disadvantages are the greater degree of complexity and

difficulty of use, the more extensive training needed to prepare a dissertation in SGML or XML, and the fact that there is as yet no “industry standard” for the markup definition file.

Documents marked up in SGML or XML must have an associated file called a Document Type Definition (DTD) that specifies the markup rules for that document type. The DTD defines the standard parts of the document (title, author, keywords, *etc.*) and specifies the structure of the document (how the parts are organized). A document that uses a more detailed DTD is more difficult to mark up but searchability of the document is greatly enhanced. **There is currently no accepted standard DTD for the ETD document type.** Current ETD programs that have decided to accept XML are either using a DTD for a similar document type or have developed their own ETD DTD. Either way, it is unlikely that their DTD will match the eventual universal standard. Other universities that have decided that XML is the best option, but many are postponing implementation of their programs until the DTD standard is developed and accepted.

After extensive discussion and consideration, the Committee **recommends** that the major file formats that will be accepted during the first-year Pilot Project will be PDF and HTML. HTML (HyperText Markup Language) is also a subset of SGML, but the DTD is built into standard web browser software. HTML defines the way a document “appears” rather than how it functions. But it is non-proprietary, can be read with standard web software, will allow the author to create a truly electronic document (with nonlinear navigational links and multimedia enhancements) and can be created with any text-based editor. HTML is not a typical file format for other ETD projects. But we believe that accepting both PDF and HTML is an appropriate compromise to give students the broadest possible range of expression while still enabling our pilot project to be activated quickly and maintain the best options for migration to future formats. No other major formats will be accepted.

One notable consequence of allowing HTML as a format is that current Graduate School appearance formats (margins, *etc.*) will no longer be enforceable. For this reason, we **recommend** that theses submitted in HTML format also be submitted in PDF format for the present. It requires a greater level of skill to prepare an HTML document, so those who do so should have little additional difficulty also preparing a PDF version.

We further **recommend** that the Committee continue working on the option of using XML in the future, and to promote the advantages of XML over PDF. Clearly use of XML will require a more extensive training and technical support system for students than is the case with PDF. There is a strong practical judgement by the majority of the Committee that we could simply not have a viable XML option in place for Fall 2000, and that it is highly questionable for Spring 2001.

In June 2000, a major international meeting will be held in Germany on the XML format; our hope is that this will prove to be a significant watershed in clarifying and standardizing a markup language for ETDs.

### **Optional Multimedia Enhancement Formats**

The richness and scholarly use of at least certain types of dissertations can be greatly enhanced by the addition of components available through specialized, non-text software. For example, consider the information and richness that can be provided by inclusion of high-resolution maps, scanned images,

three-dimensional rotatable figures, audio and video clips, and animations showing changes over time. The problem is that there are myriad software programs that will perform these tasks, their lifetimes are short, some programs are upgraded frequently and some disappear. The half-lives of software programs and even whole software companies are short. How then does one preserve the availability and usability of an ETD over, say, the next 100 years? We have some 3000 years of technology to support the putting of marks on paper, making paper, storing paper, and transporting paper. This is an enormous and unsolved problem with electronic information in general and ETDs in particular. For example, the U.S. is currently involved in mapping the entire face of the earth, to a resolution of a meter or so, via satellite. However, all the Landsat data of some 20-30 years ago is in serious danger of being permanently lost because of the costs involved in “migrating” the data to more modern electronic storage and reading systems.

Clearly, some compromises must be made on what software formats may be included in an ETD at UK. As Virginia Tech has been the leader in ETDs, the Committee considered permitting all the software formats that Virginia Tech currently will accept, which are shown in the table below. However, Virginia Tech and the NDLTD are quite aggressive in the numbers and types of software enhancements allowed. This, we feel, must be balanced by the dangers involved in many of them disappearing with time and serious migration problems in the future. The multimedia enhancements will be lost to the reader. Further, as a general rule, it is very difficult to take something away that was once provided or allowed; it is more prudent to begin with a smaller list and, with experience, add to it.

We therefore **recommend** that only those software formats listed in the table of Approved Multimedia Formats following *with an asterisk* be accepted during the Pilot Project, and that any others be strongly discouraged. In the latter and other cases, the student may well have to provide and imbed the software needed to perform the function desired, for example, a video reader. The student would thus also need to secure any needed copyright releases or licensing agreements for the software. Further, there is no guarantee that the operating systems of the future will be able to run the imbedded software.

Students must be clearly informed that any optional enhancement file, even those in the currently approved formats contained in the following list, will not necessarily be migrated to future file formats. These are optional, and they are enhancements. The “core” of the dissertation in the major file format must contain all the important and critical facets of the document. The University can fully commit only to migrating the base text of the ETD (the PDF or HTML document itself).

## APPROVED FORMATS FOR OPTIONAL MULTIMEDIA ENHANCEMENTS

---

### IMAGES

---

- \*PDF (.pdf) use Type I PostScript fonts
  - \*TIFF following version 6.0 or later, including CCITT G4 (.tif)
  - \*JPEG (.jpg); also requires the TIFF file
  - \*CompuServe GIF (.gif)
  - CGM Computer Graphics Metafile (.cgm)
  - PhotoCD
- (Recommend a *minimum* resolution of 600 dpi for images of pages with text, and suggest that imbedded images be available in several possible resolutions)

### VIDEO

---

- \*MPEG (i.e., MPEG-1, MPEG-2) (.mpg)
- QuickTime - Apple (.mov)
- Audio Video Interleaved - Microsoft (.avi)

### AUDIO

---

- \*MPEG-2
- \*WAV (.wav)
- CD-DA
- CD-ROM/XA (A or B or C)
- AIF (.aif)
- SND (.snd)
- MIDI (with timing information) (.midi)

### AUTHORING

---

- Authorware
- Director (MMM, PICS)

### SPECIAL

---

- Spreadsheet - Excel (.xcl)
- AutoCAD (.dxf)

### REFERRING

---

- \*"Handles" such as URNs (URLs do not last very long and are discouraged)
  - \*ISBN, ISSN
-

## **Optional or Required Participation**

In keeping with charge to the Committee, the Committee concurs that submission of an electronic thesis or dissertation should remain an *option* for the student, at least for the immediate future.

## **Theses and Dissertations**

In keeping with the charge to the Committee, the Committee heartily agrees that both master's theses and doctoral dissertations be accepted in electronic format. The ETD projects at a number of universities include master's theses, and this is an excellent way to showcase the fine research and creative work done by our master's students. The question of whether or not theses should also be submitted to UMI is a separate one, and addressed later.

## **Membership in the NDLTD**

In February 2000, the University of Kentucky joined the Networked Digital Library of Theses and Dissertations (NDLTD) in a status appropriate to the stage at which we were – an institution that has begun a pilot project involving ETDs. The NDLTD is an umbrella group that works together to maintain and make available to the entire world the networked archives of ETDs. Participating member institutions maintain their collection of ETDs on their own computer networks. The NDLTD serves to coordinate and assist member institutions, and as the starting point for electronic search and retrieval from the distributed collections. Further information is available at <<http://www.ndltd.org/>>.

## **Submit Paper Copy Also?**

Some universities also require the student to submit an archival paper copy of the thesis or dissertation along with the ETD. This Committee **recommends** that the eventual system to be designed *not* require that paper copies also be submitted. One of the objectives in switching to ETDs is to eliminate paper, not to continue it. This is also one of the incentives for students to switch to ETDs.

However, during the early phases of the pilot project, before the full system is in place and fully ruggedness-tested for safety and security, the Committee **recommends** that the student participants be required to submit an ordinary paper copy or a copy on disk (Zip or CD). We feel this is a necessary precaution to guard against loss or destruction of a student's thesis or dissertation. Further, this will be a necessity before the procedures for electronic submission of the ETD via the Internet and reliable, multiply redundant backup procedures are in place and fully tested.

It is to be noted that unless or until the University receives permission from the State Archivist to maintain theses and dissertations in electronic format, we will need to maintain a paper copy because these documents are classified as "permanent" public records. This requirement is not expected to be a long-term impediment, if at all.

## Relationship with University Microfilms International (UMI)

UMI, formerly University Microfilms, is now a part of Bell & Howell Information and Learning Company. Additional information is available on their web site: <[www.umi.com](http://www.umi.com)>. The University has a many-decade history of sending doctoral dissertations and abstracts to this company for “publication” and sale of paper or microform copies.

Many people use the term “publishing” an ETD when an institution uploads one onto its Web site to make it available to interested scholars. This may be an unfortunate term because it seems to make many faculty and groups nervous that the student may have lost rights to the material or that a publisher or professional society will then refuse to publish a book or article derived from the dissertation; but the term is in common use. A very important point to be made here is that a university “publishing” an ETD on the Web is legally no different from that same university “publishing” a paper dissertation by binding, cataloging, and depositing it in its library system for access by scholars or the general public. The common wisdom is that the student enjoys no less rights or legal protection with ETDs; the main difference is the faster and much wider availability of the latter. UMI, for example, makes no essential distinction between dissertations submitted on paper or in electronic format.

Currently, doctoral students at the University of Kentucky are required to submit two copies of their dissertations on 100% rag paper and to pay a \$57 dissertation fee; master’s students pay only the \$7 fee. [Throughout this document, the term “dissertation” is meant to include DMA projects also. The term “thesis” would also include any Plan A educational specialist program that might be established.] Of these amounts, \$50 is sent to UMI as a “publishing fee” for dissertations, and \$7 is a “binding fee” to help defray the costs involved in binding, cataloging, and archiving the document in University Libraries.

The doctoral student fills out a booklet entitled *Survey of Earned Doctorates* for the National Opinion Research Center at the University of Chicago and a *Doctoral Dissertation [Publishing] Agreement Form* for UMI. The latter essentially grants UMI the non-exclusive rights to reproduce and distribute the abstract and dissertation in paper, microform, and electronic format. The student, however, retains the original copyright to the dissertation and derivative rights thereto. For an additional \$35 fee, UMI will *register* the student’s copyright of the dissertation with the Copyright Office. UMI asserts that this confers additional legal rights on the student and protection of the rights to the work. The University does not send master’s theses to UMI; thus the student does not pay the “publication fee”, but might possibly have a somewhat diminished right to or protection of the material. The UMI booklet containing the Agreement Form also lists their requirements for electronic dissertations.

In the initial days of ETDs, some professional societies, publishers, and even UMI did indeed balk at publishing materials that were freely available on the Web. However, with time and with the increasing interest in ETDs, most, if not all, seem to have backed off from their initially negative reactions. Some recognize ETDs as the wave of the future, and there is money to be made from them. Full books derived from initial dissertations are so changed and edited that they essentially comprise a “new work”; there do not seem to be copyright problems encountered. Derived articles published in respected and refereed journals are bestowed with added peer validation over an ETD. Societies and publishing houses have come to realize that they will continue to make money in this enterprise.

The Committee **recommends** that the University continue our relationship with UMI for doctoral students. This provides for an alternative, more traditional, venue for display of the student's work, and continues a long-term, reliable safeguard against loss of the student's work. There was little, if any, inclination among Committee members to recommend that master's theses also be submitted to UMI; few universities do this.

The Committee further **recommends** that the Graduate School re-visit the issue of the dissertation or graduation fee for students who submit an ETD. If it is a "fee", then it is a fee to cover (some of) the costs of the services rendered. What are the current components of and services rendered for the fee? Will the services rendered be significantly more or less costly in the case of ETDs? With a view to fairness, we feel this issue should be studied.

As an example, Virginia Tech charges master's students a \$20 archiving fee for a thesis and doctoral Students a \$70 fee for a dissertation (\$20 archiving and \$50 UMI charge). West Virginia University charges master's and doctoral students graduation fees of \$55 and \$65, respectively; apparently, both theses and dissertations are sent to UMI. (There is also an optional \$45 fee to register the copyright.)

## Copyright Issues

Copyright arises automatically when a work is first fixed in a tangible medium such as a book, film, videotape, or electronic medium. The student owns the copyright to a dissertation or thesis; there is no legal difference whether the work is in paper or electronic format. For example, the student's agreement with UMI grants them the *non-exclusive* rights to publish the dissertation and abstract in paper, microform, and/or electronic formats. The student does not sign over the copyright, as is the case for an article published in most professional journals; and the student still has the right to publish a book or article(s) from material derived from the dissertation. Some additional information on copyright, research, and dissertation issues is available on UMI's web site <[www.umi.com](http://www.umi.com)> at <[www.umi.com/hp/Support/DServices/prepare/copyright.htm](http://www.umi.com/hp/Support/DServices/prepare/copyright.htm)>.

The Committee discussed the apparent perception that worldwide Internet access may make any copyright issues a more visible issue and possible problem. Although we are unaware of any data or other reports to validate this hypothesis, we are led to suggest that students be provided enhanced education on copyright and fair-use issues.

The Committee **recommends** that students be provided copyright information, including their rights, the use of 3<sup>rd</sup>-party copyrighted material in their theses, and possible problems with later publication or other copyright issues. It seems logical to broaden this effort to fully informing *all* students, including those preparing paper versions, because the rights and obligations are the same regardless of the medium.

## Access Restrictions for ETDs

Because of the immediate and ready accessibility of ETDs, most universities seem to offer access options for ETDs. These restrictions can roughly be classified as "geographic" or as temporal in nature. Initially, Virginia Tech, arguably the world's leader in ETDs, allowed two levels of geographic

restrictions. These were (1) by password only or (2) from the campus network only. The latter roughly emulated the same situation as with paper dissertations. One had to come to campus to view or sometimes even to find a dissertation, although Dissertation Abstracts makes most dissertations discoverable and most university libraries make them available via inter-library loan. Theses, in general, languish in even greater obscurity. Temporal restrictions seem to range from 1-3 years prior to worldwide access; sometimes it is 1 year, with provisions for limited renewals for cause. Sometimes the restrictions are for the whole work, sometimes only for certain parts thereof. However, with time and the more widespread submission of ETDs, many universities have backed off from excessive restrictions on immediate worldwide availability. They have found that students and their advisors seem to quickly be unconcerned about geographic restrictions.

The Committee pondered these issues for some time. We realize that there are sometimes quite valid reasons, such as for patent and/or proprietary purposes, in which some delay is desirable or needed. The Graduate School already does informally hold back from sending a paper thesis or dissertation to the University Libraries (and UMI) for a period of time in such situations, although there are typically very few such documents involved. Weighing the facts that participation in UK's ETD system is voluntary, that the issue of access restrictions seems to have much less importance than a few years ago, but that ETDs are much more quickly and widely accessible, the Committee has opted for some (minimal) access restrictions.

The Committee **recommends** that there be *NO* "geographic" access restrictions, such as by password only or from the UK campus only. Worldwide access or none at all.

The Committee **recommends** that the option of a temporal restriction be permitted from the date of submission of the ETD: A *one-time* restriction with a limit of *two years maximum* and only for valid reasons *with specific permission of the Dean* of the Graduate School. [For ease in administering this restriction, perhaps the student/advisor could have a choice between 1 or 2 years for the one-time restriction.] If someone has no substantive reason for withholding access or wishes a longer or open-ended period, they can default to the current paper system.

## Unique Identifier

Those involved with ETDs agree that each ETD should have a unique identifier. One idea is that the NDLTD should design a numbering scheme whereby each ETD will have its own identifier number – an identifier for each participating institution in the distributed system plus a systematic identifier for each ETD at that institution. Another idea is that each ETD be assigned an ISBN number and treated like a book. There is as yet no industry standard or clear consensus, but that will undoubtedly come. In the meantime, however, the Committee feels that our ETDs should be assigned a systematic identifier within the university that will permit unambiguous classification, tracking, and eventual conversion over to whatever standard identifier system is eventually selected by consensus.

The Committee **recommends** that each ETD submitted be assigned a unique identifier by the Graduate School on submission. The structure of the identifier, taken from the Virginia Tech system, recommended is – [University identifier]-[Program identifier]-[Year]-[Thesis or Dissertation]-[Sequential number]. It is not clear whether a classification by calendar year or academic year would be better for cataloging, reporting, and assessment purposes. At present, most reporting and

assessment figures at the Graduate School are required to be done by academic year, whereas the Library catalogs paper theses and dissertations by calendar year.

For UK, this would translate into – UKY-[4-letter Program Code]-[Year]-[T or D]-5-digit sequential number. A unique 5-digit sequential number avoids ambiguity and would carry us through the first hundred thousand ETDs, about 20 years even if ETDs were to be made mandatory at UK.

Specific examples are: UKY-HIST-2000-D-00027 or UKY-CHEN-2001-T-00102

## **Submission and Backup Procedures**

Without going into many details that will have to be worked out over the next year or two, the Committee broadly **recommends** that the current division of authority and responsibility that exists between the Graduate School and the University Libraries be maintained, and further, that the University follows the systems established by Virginia Tech when possible.

Essentially, the student will submit (a) the Examination copy (dissertations only) and (b) the Final copy of the ETD to the Graduate School who will approve each in turn. When the Final copy is approved, (c) the Graduate School will transmit it to the University Libraries (or release a hold on the copy). The Libraries will (d) catalog and upload the thesis or dissertation to the UK ETD web site, store and archive it, and (e) submit appropriate materials to the NDLTD and to UMI (the latter for dissertations only).

The Graduate School will be responsible for carefully securing any ETD for which a temporal time-delay has been approved, and ensure against its premature release to the ETD web site, NDLTD, and UMI.

The Committee strongly **recommends** that the Graduate School and the University Libraries institute a reliable and multiply redundant set of backup procedures to ensure against loss or corruption of a submitted ETD.

Until a tested and reliable system is devised for Internet submission of an ETD, the Committee **recommends** that the student be required to submit one (or two) electronic copies on disk – floppy disk, zip disk, or CD – and possibly also a paper copy on ordinary paper. Until a fully tested and reliable system is in place for backing up and storing ETDs on the UK web site, the Committee **recommends** that the student be required to submit an electronic or ordinary-paper copy of the work.

The Committee further **recommends** that the student be strongly advised to also keep at least two electronic copies of his/her ETD in physically different locations as a safety measure against loss. (Perhaps one copy could be kept by the Advisor.)

## **Format Migration**

Although it does not appear to be written down anywhere, the University of Kentucky seems to have historically committed itself to the availability and preservation of the (paper) theses and dissertations

submitted to it. This seems to be a common practice among all universities. While it may not be a contractual agreement, there appears to be a long-term tacit understanding between university and student. In a sense, these documents are “published” by the university when it accepts, binds, catalogs, deposits, preserves, and makes them available to scholars. The University of Kentucky lists Theses and Dissertations as documents for “Permanent Retention” in its public-records retention schedule.

We must therefore do no less than the *equivalent* for ETDs. Ensuring the long-term availability and readability of ETDs as software evolves (and disappears) is commonly called “migration” of the ETD. It is not a trivial matter, and is by no means clearly resolved. It is an issue of pressing concern for all involved with electronic information and documents. We clearly cannot guarantee with certitude the 100-year survivability of ETDs, any more than we can with paper copies: Libraries burn down; insects eat paper dissertations; floods, mold, and air-oxidation destroy them.

From the Virginia Tech web site:

“The library shares with the university the responsibility to guarantee that ETDs will be available to researchers, both within and outside the scholarly community. To keep ETDs reader-friendly and to maintain accessibility will mean migrating the current file formats to new, standard formats not yet known. This will be done through the cooperative efforts of the library (who maintains the submission software, the database of ETDs, and the secure archive) and university computing expertise.”

The Committee recommends a three-tier migration system as being reasonably possible and roughly equivalent to the situation with paper theses and dissertations:

1. Every possible effort will be made to migrate the *major file formats* – currently PDF and HTML, eventually to include XML and others. In the worst possible situation, paper copies can be printed, bound, and stored. This comprises the body of the dissertation.
2. Effort will be made to migrate the *optional multimedia formats* on the University’s Approved List of such. It is not possible at this time to guarantee with high levels of certitude the successful migration of all these formats for 20-50 years into the future.
3. No effort will be made to migrate multimedia formats not on the Approved list and evanescent items such as URLs. These change and disappear too quickly.

The Committee **recommends** that students and advisors be clearly informed of these different levels of permanency, and be cautioned to prepare carefully the main body of the ETD (that in the major file format) with this factor in mind. For example, data in an optional Excel spreadsheet must be adequately explained in the text, and/or a table in the body of the text could contain a sufficient sample of the total data from the spreadsheet to support the main points being made. Large, high-resolution, multi-page maps can be accompanied by a smaller, lower-resolution map in the body of the thesis.

### **Digital Signature and “Watermarking”**

Ideally, the ETD submission procedure should be totally electronic, including electronic approval of the Final copy by the Advisor and DGS. How can the identity of these two and the student be verified in the case of electronic approval? Currently, we rely on written signatures on the two Final copies of the thesis or dissertation that are submitted. Secondly, electronic documents can easily be altered or

corrupted, by design or by accident. How does one guarantee that a copy of an ETD being read by a scholar in the future is identical to the original, approved copy? At this time, one reasonably relies on the security procedures of the library in which the paper thesis is deposited or on those of UMI.

The details of electronic signature authentication and of digital “watermarking” have not yet been worked out. There is no accepted industry standard.

For these reasons, the Committee is currently investigating and **recommends** that the University monitor the evolution of electronic authentication and incorporate these desirable elements into the ETD process when possible. For the time being, hard (paper) copy will be needed.

### **Additional Resources Needed**

In the first year of the ETD Pilot Project, those who will be doing the work involved in implementing the project seem to agree that it can be started with relatively little additional resources. Personnel would devote time to the Project as a part of their normal workload. A temporary (shared) server computer that will house web information and archived ETDs has been secured. There will be some costs involved in purchasing needed software for student and project personnel use and backup media. However, it is clear that as students increasingly turn to ETDs and our collection of archived ETDs grows, that significant additional (or redirected) personnel and equipment resources will be required for the long term.

With adequate infrastructure support, it seems clear that there is a cost benefit for the student with ETDs, realized primarily from eliminating the cost of duplicating two copies of the thesis or dissertation onto 100% rag paper (now 20 cents per page). It is not at all clear whether an ETD system would mean a net increase or decrease in costs for the University as a whole. Full-time students, however, now pay a \$40 per semester Technology Fee, so it seems appropriate to provide some level of infrastructure support to the student for the preparation of ETDs.

Cornell University has done a detailed, if now somewhat dated, cost analysis for implementing ETDs, which can be consulted at <[www.library.cornell.edu/staffweb/ETDSTUDY.HTML](http://www.library.cornell.edu/staffweb/ETDSTUDY.HTML)> for details. Virginia Tech also provides (1998) cost estimates on its web site at <[scholar.lib.vt.edu/theses/data/setup.html](http://scholar.lib.vt.edu/theses/data/setup.html)> of about \$24,000 for Staff, \$36,000 for Equipment, and \$15,000 for Software. To summarize from these reports, the resources needed at UK for a moderate level of ETD implementation are estimated to be:

- Site Equipment – An Operations Server, a Web Server, backup equipment, software, and supplies (ca. \$30,000)
- Library Staffing – A combination of Librarian, Programmer/Technical Consultant, and Student Assistant equal to about 1.0 FTE (ca. \$35,000 per annum including benefits)
- Graduate School Staffing – Added electronic Thesis/Dissertation checking of perhaps 1/8 FTE (ca. \$3,000 per annum)
- Student ETD Software and Equipment – Additional supporting infrastructure at one central or several distributed sites (ca. \$12,000 per site)

The total estimate is about \$40,000 annually in recurring personnel costs and about \$10-15,000 per year in equipment and software costs assuming 5-year lifetimes for each, but does depend on the number of student-support sites.

The Committee therefore **recommends** that the University Libraries and the Graduate School work closely together with other appropriate University units to ensure that the needed levels of resources for personnel, software, equipment, and student and faculty training are available on a timely basis to meet the anticipated increase in ETD submission. Budgetary recommendations will need to be made in time for the 2001-2002 Fiscal Year. As more students become aware of the increasing importance of acquiring the skills needed to prepare electronic documents in general, not just theses and dissertations, the need for additional University support of the technology infrastructure for both undergraduate and graduate students will arise anyway. If ETDs are to become a facet of scholarly communication at the University of Kentucky, the resources, training, and support system needed must be institutionalized. The Committee notes with pleasure the strong interest of the Teaching and Learning Center in participating in our ETD project in ways consistent with its own mission.

### **Committee Membership**

The *ad hoc* Committee consisted of faculty, staff, and a student representing The Graduate School, the University Library, Legal Consul, and several academic disciplines:

Katherine A. Adams, Associate General Consul, Office of Legal Counsel  
Suzanne L. Allard, Doctoral Student & Instructor, College of Communications and Information Studies  
Gary J. Anglin, Professor & Chair, Department of Curriculum and Instruction  
Mary K. Kelly, Web Coordinator, The Graduate School  
Kevin S. Kiernan, Professor, Department of English  
Elizabeth G. Kraemer, Electronic Resources Librarian, University Libraries  
James E. O'Reilly, Associate Professor, Department of Chemistry, and Associate Dean, The Graduate School  
Victor W. Marek, Professor, Department of Computer Science