

[CUL Staff Web Site](#) | [Hot Topics Archive](#)

Cornell University Library

**FOR ADMINISTRATIVE PURPOSES, THE PROJECT IN THIS REPORT WAS NOT
IMPLEMENTED**

Task Force on Electronic Theses and Dissertations

Report and Recommendations

Prepared for:

Walter Cohen, Vice Provost and Dean of the Graduate School
Ann Stundem, CIT Director of Academic Technology Services
Sarah Thomas, Carl A. Kroch Librarian

By:

Martha Crowe (chair), Susan Currie, Minnie Empson,
Elaine Engst, Richard Marisa, and John Saylor

September 17, 1999

CONTENTS:

[Summary](#)

[Recommended Procedures](#)

- Scope of the ETD Collection
- Preliminaries
- Submission
- Processing

[Rationale for Recommended Procedures](#)

- Scope of the ETD Collection
- Preliminaries
- Submission
- Processing

Issues and Proposed Resolutions

- Restricted Access
- Preservation
- Printing
- Relationship to UMI and Publication Rights
- Membership in the Networked Digital Library of Theses and Dissertations (NDLTD)

Evaluation

Considerations for the Future

Conclusion

Budget

Task Force on Electronic Theses and Dissertations

Report and Recommendations

In agreement with the decision of representatives of the Cornell Graduate School, the University Library, and CIT, Sarah Thomas selected and charged the Task Force on Electronic Theses and Dissertations (ETD) with planning a trial program for electronic versions of theses and dissertations that would run parallel to the current processing of printed versions. In addition, the task force was to prepare a budget for that activity and suggest evaluation criteria to determine the future of the service after the pilot program.

SUMMARY

This report contains a set of recommended procedures, the rationale for the recommendations, a discussion of issues and suggested solutions, suggestions for evaluation, considerations for the future,

and a budget. The task force emphasizes that it sees the ETD project as a work-in-progress and that the recommendations it makes at this point should not foreclose changes at any point in the development process. Indeed, it recommends in the rationale that certain suggestions be reconsidered during the pilot period.

RECOMMENDED PROCEDURES

- *Scope of the ETD Collection:*

The Cornell ETD collection will consist of scanned copies of printed dissertations, which means that digital TIFF images corresponding to the printed pages will be the canonical format of the collection. In the prototype stage it can include dissertations with foldouts, photographs, and color images. It will also incorporate in the course of the first year the 114 dissertations previously scanned in the DAISY project (described in "Publication of Electronic Dissertations," at <http://www.library.cornell.edu/staffweb/ETDSTUDY.HTML>). Electronic submission may be added after the prototype period.

- *Preliminaries:*

- The Graduate School revises its submission instructions to include the electronic publishing option and inform students how to participate.
- The Graduate school makes an announcement on its Web site.
- The Graduate School announces the new opportunity to all appropriate fields and requests that faculty and committee chairs encourage degree candidates to take advantage of it.
- A notice is included in the Grad Bulletin in the Cornell Chronicle.
- If feasible, the Graduate School publishes an announcement in the *Cornell Daily Sun*.
- The Graduate School identifies those departments, such as Music, which tends to have oversize and irregular formats, that are not considered appropriate candidates for prototyping. These will not be targeted for initial inclusion.

- *Submission:*

Students submit exactly the same materials as currently, with the addition of:

- a signed release allowing the dissertation to be mounted in the public archive
- an electronic abstract and bibliographic data to be submitted by filling out an online Web form.

- *Processing:*

- The thesis advisor checks to verify that the electronic abstract and metadata have been submitted.
- A third copy of the dissertation, to be used in the scanning process, is produced by the Olin Copy Shop on non-archival paper and charged to an account set up for that purpose, after which the two original copies are submitted to the Graduate School as they are currently. The third copy goes to CIT.
- CIT scans the third copy to TIFF images at 300 or 600 dpi and then recycles it.

- A student assistant does minimal document structuring (such as page numbers, author, title, date, and abstract).
 - CIT completes a dissertations home page (the basic design has already been created).
 - The files are mounted on a library server (probably Library 5) using the Dienst document archive.
 - Dissertations are processed as they are received during each conferral period, but the electronic files are accessible only via a private, unpublished, and access-protected URL (e.g., for quality control) until they become publicly accessible after formal conferral.
 - The electronic metadata and abstract submitted by the student are used directly by library staff to create the bibliographic record, which includes the abstract, for the online catalog.
-

RATIONALE FOR RECOMMENDED PROCEDURES

- *Scope of the ETD Collection:*

The decision to scan paper copies rather than ask for electronic submissions is based on several factors, the first being the desire of the Graduate School to introduce electronic dissertations with as little inconvenience to the students as possible. Because a number of faculty stated a preference for direct electronic submission, the task force did examine the possibility of accepting electronic formats, as well as scanning print copies, and concluded that we do not have the staff to process and verify electronic formats during the pilot project.

In addition, electronic file formats are now in flux, and the most logical one for ETD purposes, PDF, can be unreliable in some instances. The task force expects to accept and disseminate electronic formats in the future.

The processing of masters' theses is different from that of dissertations. They are submitted already bound and are not scannable in that form. Rather than deal with changing procedures for theses during the initial project, the task force prefers to exclude them.

The 114 dissertations scanned in the DAISY project are already structured and require almost no preparation to be added to the database.

- *Preliminaries:*

There should be few limits on the number of departments participating; in fact, the more items processed during this prototype period, the greater will be the effectiveness of the trial in developing a smooth and effective work flow for the future production stage. As scanning requires no particular level of computer expertise on the part of the student, the door is immediately open to participation by many fields.

Since there are no disincentives to student participation, generating broad public awareness of the

opportunity may be key to a large response.

- *Submission:*

The student continues to contribute two archival copies, rather than reducing to one, because it is the experience of library public services that students like to come and "see" their dissertations on the library shelf, making the retention of the second paper copy a public relations matter. As well, smaller institutions that are not as well equipped electronically will not have access to the Web version, and it is the second copy that can be sent to them on interlibrary loan.

- *Processing:*

The Graduate School expressed some concern for the security of the copy that will be sent for scanning. In the previous DAISY project, procedures were very informal, and it was easy for various accompanying documentation to be lost. Initially this project will create a third copy to travel the scanning route and minimize the changes to current procedures for handling the two archival copies. Once a proven mechanism is established for scanning, the third copy may be eliminated.

Scanning is at present not a production service at CIT, but it is retooling its scanning ability and is willing to do the scanning for the pilot project. This arrangement will develop an effective procedure and set standards. For future production, scanning will be outsourced. A reasonable step would be to approach Challenge Industries, which now does the microfilming of dissertations, and suggest that they set up a scanning operation. There is no suitably efficient and cost-effective vendor on campus, and Challenge would likely attract business from other clients also. In addition, Challenge has expressed interest in scanning and indicated that they are concerned about reduced income from microfilming in the future. This timing could provide an incentive for both the ETD project and Challenge to investigate initiating a service there.

A student will be hired to do the document structuring. Although CIT itself has no organization to do structuring, Rich Marisa has offered to manage and train the person.

By university regulation, dissertations may not be made public until conferral is approved by faculty vote; therefore, scanned dissertations will reside on the server but remain inaccessible to viewers until conferral.

ISSUES AND PROPOSED RESOLUTIONS

- *Restricted Access:*

The Graduate School permits restrictions on access only for patent protection. Such dissertations are physically retained in the Graduate School office until they can be released for distribution. Thus the question of restricting electronic access is moot.

- *Preservation:*

The library will be responsible for archiving and backup. The electronic files will reside in a library server located in the CIT server farm. Each time the library files are backed up, the dissertations will be included.

The archival paper copy will be housed in a secure location in the library as part of the university archives. The TIFF images are currently as close as technology can come to providing archival quality for digital files. Microfilm is still the accepted national standard medium of archival preservation, and for the present Cornell will continue to microfilm all dissertations and submit the microfilm to UMI.

- *Printing:*

CIT has the capability of providing printed and bound copies of dissertations ordered directly from the ETD Web site. Because of the sensitivity of UMI toward such competition, however, it would be politic not to offer print-on-demand at this time. The Cornell library receives a number of services through UMI, and maintaining a good relationship is important. We will demonstrate printing from the ETD archive via CUPID/NetPrint, but for the time being UMI will continue to be the supplier of full copies of dissertations.

- *Relationship to UMI and Publication Rights:*

No changes are expected in Cornell's relationship to UMI (now owned by Bell Howell) at this stage. Challenge Industries will continue to microfilm all dissertations and send the film to UMI, which verifies that the quality is acceptable.

Students sign a dissertation publishing agreement with UMI granting to UMI the non-exclusive right to reproduce and distribute their dissertations in paper, microform, and electronic formats. Other publication rights may be granted as the students choose. University Counsel searched its files and found no record of a contract between the university itself and UMI. The office manager of the Graduate School, which would have been the keeper of the contract, was also unsuccessful at finding any documentation. When the Graduate School repeatedly requested a copy from UMI, the company finally responded only with a sheet stating rates. Counsel concludes that it is possible there is no contract with the university, as the students sign the copyright agreement directly with UMI, and it is unlikely that the university would have ever allowed UMI exclusive rights. Counsel sees no reason not to proceed.

The dissertation home page will post a notice such as "Reproduction or distribution of these dissertations in any format is prohibited without written permission of the author." In addition, Counsel recommends that the permission form signed by the students who choose electronic publication contain the following elements:

- a statement that no parts of the dissertation have been previously published (as may be the case in scientific fields), or, if any have been, the student must supply documentation that

further publication is permitted by the original publisher notification that electronic publishing of the dissertation may be considered prior publication by some publishers and would preclude later publication by them in the event that at some time in the future the university wants to provide a print-on-demand service, a box to check if the student wants to give permission to have the dissertation printed and sold Counsel has agreed to review any permission statements that are drawn up.

- Membership in the Networked Digital Library of Theses and Dissertations (NDLTD). The task force refers the decision on joining NDLTD to the Graduate School and the library, since it is more appropriately an administrative decision. As it is independent of material progress in the implementation of Cornell's ETD project, it can be considered at any time. NDLTD is an international federation of member universities, centered at Virginia Tech, that have joined in an initiative to provide free online access to as many dissertations and theses as possible. Currently there are fifty-seven member universities and six member institutions. Membership in the initiative entails:
 - A letter of commitment.
 - Collaboration with other members in establishing standards and sharing of information to ensure interoperability.
 - Sharing all ETD MARC bibliographic records.

Member libraries are responsible for serving and maintaining their own dissertation files, as well as for implementing their preferred method of archiving. Although hod of archiving. Although the NDLTD welcomes participation in, for example, developing templates and testing software, there is no obligation to do so. The letter of commitment is basically a statement that the institution intends to establish a pilot program for serving ETDs with the ultimate goal of requiring electronic submission at some time in the future. There are many options-choices depending on local politics, practice, economics, and other individual institutional needs.

Almost all the member schools are smaller institutions with the exception of the University of Texas, which joined because "It just seemed the thing to do." There appear to be two main reasons why the larger and more-prestigious institutions have not joined. First, they see no real benefit in doing so (except for increased visibility, as explained below). Large institutions have the resources and technology to publish dissertations for themselves, but the smaller schools need the support of each other and the organizational expertise of NDLTD. Second, the larger schools are asking the same question: Why aren't the others joining? Everyone is hesitant to go first. In sum,

- Cornell's technology is compatible with the NDLTD. Cornell would need only to give the URL of its dissertation site to the NDLTD, and Virginia Tech would set the connection.
- Cornell would have no obligation other than to contribute its MARC cataloging records.
- Cornell Library users and staff would derive no particular benefit, as they already have free access to the NDLTD database via the Web.
- The NDLTD would benefit from listing Cornell as a member; it would gain credibility

and impetus from attracting major universities.

- Cornell and its graduates would benefit from increased visibility for, and access to, the university's ETD collection. NDLTD would provide a link on its Web site to Cornell, and Cornell's site would be indexed in LTD's central index so that anyone doing federated searching on the NDLTD site would automatically be searching Cornell's database along with the others.
 - Three representatives of the NDLTD at Virginia Tech are available on invitation to visit interested institutions and make a presentation: a professor of computer science (the director of the project), the former dean of the Graduate School and current associate provost for graduate studies, and the scholarly communications librarian.
-

EVALUATION

- Evaluation of the pilot project should occur after a year.
 - The server will collect usage statistics. It is possible to identify the host computers of users from academic institutions or with distinct addresses.
 - A brief survey form will be designed for the ETD site that will ask visitors to indicate their organization, position, and reason for visiting the site without identifying the user personally.
-

CONSIDERATIONS FOR THE FUTURE

- At the end of the pilot year Cornell should examine accepting electronic submissions and discuss acceptable formats, especially for students in the sciences and engineering, some of whom (and whose faculty) have already expressed a preference for submitting digital materials.
- UMI is now accepting electronic submissions and archivally microfilming all it receives. Once Cornell's ETDs have progressed from pilot to production, it would be technologically possible to cease microfilming them here and arrange a procedure with UMI whereby UMI could use Cornell's electronic files directly. Currently UMI's charges remain the same whether an institution submits film or an electronic version.
- By summer 2000, the library will be operating with a new library management system, a Web-accessible catalog with developing image-serving capabilities. Cornell should consider serving the dissertations through the library catalog, which will integrate them into the same environment as the other digital resources that the library provides. In the new catalog, the dissertation record will contain the abstract and a live link to the full text of the dissertation.
- At some point, the Graduate School could consider changing the processing procedures for masters' theses so they could be included.
- If Cornell wants to expand its ETD collection retrospectively, its "historical dissertations," those conferred from 1869 to 1925 (no longer under copyright) could be scanned and added.

Dissertations conferred from 1925 to the present could also be included via a notice to alumni on the dissertation Web site, giving the name of a contact person here with whom they could arrange to order their dissertations scanned at cost (or even for a contribution).

- It is advisable to revisit the technology within five years.
-

CONCLUSION

The technology for the capturing, processing, and distribution of electronic dissertations is proven. It is the procedural elements that need to be implemented and tested over a period of a year. Financial conditions for this project are also advantageous in that there are few start-up costs for hardware or equipment: the server, hardware, and software are already available, necessitating only the purchase of extra disk space for the database. Thus, start-up outlays are mainly for staff. The other costs for the pilot year are actual production costs.

The task force recommends this experimental phase as an exploration—an ongoing revision of procedures, policies, documentation, and other support to arrive at a successful electronic dissertation service.

Accompanying Budget: [*Financial Estimate for E-Dissertations Project*](#)

The following budget presents the costs for three scenarios—for each of the low (50), medium (100), and high (150) estimates of the number of submissions per year.

[CUL Staff Web Site](#) | [Hot Topics Archive](#)

05/28/02 vwb