The Communicator
COUNCIL OF GRADUATE SCHOOLS
ONE DUPONT CIRCLE, N.W., SUITE 430 WASHINGTON, D.C. 20036-1173 (202) 223-3791

VOLUME XXXI NUMBER 1 JANUARY 1998

# THE ROLE OF ELECTRONIC THESES AND DISSERTATIONS IN GRADUATE EDUCATION

John L. Eaton, Associate Provost for Graduate Studies and Professor of Entomology, VA Tech

Edward A. Fox, Professor of Computer Science and Associate Director of the Computing Center, VA Tech

Gail McMillan, Associate Professor and Head of Scholarly Communications and Special Collections for the University Library, VA Tech

In Spring and Summer of 1997, NPR, the *Chronicle of Higher Education*, the *Roanoke Times*, the *New York Times*, EDUCOM *Edupage*, and other media publicized work on electronic theses and dissertations (ETDs). These news outlets, in their desire to report on controversy, largely missed the point of why it is desirable to produce theses and dissertations, and indeed many other publications, in electronic (digital) format. Hence, in this article we will seek to correct some of the misinformation released and show the benefits of the worldwide ETD initiative.

Few would dispute the following statements that describe the context for our initiative:

- Electronic publishing and related digital technologies (digital libraries), will be widely used to produce, store, and access information.
- Most scholars and researchers at institutions of higher education have an obligation to share their findings.
- Theses and dissertations constitute a great and largely hidden cache of primary data and information, much of which does not make its way into refereed publications.
- By the use of digital technology, theses and dissertations can be both improved and made more accessible to scholars and researchers.

Accordingly, we aim to:

- have graduate students learn about electronic publishing and digital libraries, applying that knowledge as they engage in research and build and submit their own ETDs;
- improve graduate education and knowledge sharing, through more effective, more complete, faster, and more widespread access to the information in ETDs;
- give graduate students access to electronic publishing and multimedia tools, to improve the expressive capability and quality of theses and dissertations;
- improve access to knowledge, reduce staff effort, and save shelf space in university libraries;
- help universities learn about digital libraries, as they collect, catalog, archive, and make ETDs accessible to scholars worldwide; and
- aid universities to learn how to unlock the potential of their students intellectual capital, individually and through federated collaboration, by participating in a Networked Digital Library of Theses and Dissertations (NDLTD - see <http://www.ndltd.org>).

The concept of ETDs was first openly discussed at a 1987 meeting arranged by University Microfilms Inc. (UMI), and attended by representatives of Virginia Tech, University of Michigan, SoftQuad, and ArborText. In 1991, Virginia Tech decided to pursue the development of the ETD concept. Ed Fox (Computer Science) and John Eaton (Graduate School) have collaborated on this project since that time, investigating problems associated with production, archiving and access, initially with a local faculty committee. Since 1992, they have worked with the Coalition for Networked Information (CNI), the Council of Graduate Schools (CGS), UMI and other interested organizations, helping run a series of design and discussion meetings. Additionally, the Virginia Tech University Libraries' Scholarly Communications Project, under the leadership of Gail McMillan, began in 1993 to develop procedures and systems for processing, archiving, and providing public access to ETDs. At Virginia Tech, the Commission on Graduate Studies and Policies agreed to require the electronic submission of theses and dissertations beginning in 1997.

The ETD initiative spread to the Southeast with the help of $90K funding by the Southeastern Universities Research Association (SURA), and to the nation with about $210K funding by the U.S. Department of Education (for three years starting September 1996). In the first year of that project, corporate support of over $1M in donations was provided by IBM, Microsoft

*continued on page 3*

## Inside *The Communicator*

ETD's *continued from page 1*

and Adobe. The origianl goals of the numbers of universities joining have been exceeded, and scores of other universities have been visited or expressed interest. Efforts are coordinated by a steering committee including members from Adobe, Canada, CIC (Big 10), CNI, CGS, IBM, NSF, OCLC, SOLINET, SURA, UK, UMI, Western Area Graduate Schools, World Bank, etc.

Why should theses and dissertations be submitted in digital format? The main reason is that people learn best by doing, and if students submit their ETDs, they should learn many of the concepts and skills of electronic publishing. At the same time, a valuable digital library can be constructed in an economical, scaleable, and sustainable manner—and students and others can benefit from using it. We aim to improve graduate education, wherein students become information literate; this can be partly accomplished if all graduate students become electronic publishers and can use digital libraries. Since in the United States alone about 400,000 students receive a graduate degree each year, the ETD initiative could involve very large numbers of students, who will be the next generation of educators, researchers, and scholars. If they can publish electronically and add to digital libraries, the future works they produce will not have to be scanned or re-keyed to become available through digital libraries.

Toward this end, we continue to develop text and multimedia documents, extensive WWW training materials, and a distributed education and evaluation program in which universities accept responsibility for local support of the initiative. Graduate students will be able to find the full texts of related ETDs easily, to read literature reviews prepared by their peers, and to follow hypertext links to relevant data and findings. Their professors can point to the best examples of research in their area, even to the level of an interesting table, an illustrative figure, or an enlightening visualization. ETDs already include color images or graphics, illustrate concepts with animations, and explain processes with video, and could use audio when dealing with musical studies.

Access begets access, so having more graduate works in the NDLTD is likely to stimulate greater interest in ETDs. Records from January through August 1997 show that the Virginia Tech Library ETD collection was accessed nearly 100,000 times. One document was accessed 7,000 times and several others were accessed over 1,000 times Accesses come from all over the world; Twenty one institutions have now started pilot ETD projects. Three of these are in other countries.

How has Virginia Tech approached the submission of ETDs? Since 1994, the ETD submission process at Virginia Tech has students creating documents using common word processors and submitting them as Portable Document Format (PDF) files. These PDF files may be moved across computer platforms and operating systems and still retain their formatting. (The electronic documents look just like the paper copy—indeed a paper copy can be printed from the PDF file!) Software for producing PDF is provided in campus computer labs, so using PDF costs the students nothing. The Acrobat Reader software needed to view the document also is free and may be downloaded from Adobe. Students submit PDF and multimedia files via a WWW submission form. To help students with this process the project team conducts frequent workshops.

When the Graduate School receives the PDF file, it is reviewed on-line for proper format. Then the library catalogs the ETD and archives an electronic copy. An OpenText search engine supports full-text searching of the bibliographic data, abstract, and PDF files. Alternatively, digital library patrons can use the on-line library catalog to locate ETDs and use the given addresses (e.g., a persistent URL) to download them.

What are some of the challenges faced by an institution wishing to join the NDLTD? The Graduate School, library, and computing center must collaborate for the project to succeed. It is crucial to get faculty and graduate student support, and to develop an advisory committee of technologically adept faculty to help develop local policy. The governance system also should be informed during the development phases so they can guide full scale campus implementation.

Student training is of central importance. Creation of the PDF file and subsequent submission is relatively easy. A more serious challenge is improving the low level word processing skills of students, who typically use a wide variety of word processors and related tools (e.g., for mathematics, diagrams, or scanning images). We help students learn how to create electronic documents, which requires many to understand concepts such as resolution, fonts, quality of compression, bitmaps, and hypertext. Moving to requiring ETDs necessitates that some thought be given to enhancing the training, support staff, and programs, and arranging for electronic archiving.

What about relations with publishers? Since the ETD initiative is a federated effort involving many universities, and since there are thousands of publishers, this is complicated. Virginia Tech has contacted a number of publishers, as well as publisher's associations such as AAP and AAUP. After detailed discussion with the American Chemical Society or AAAS, cordial and amicable solutions were developed that satisfy all parties. We hope that such publishers will formalize and make clear statements about these solutions so that students clearly understand how to proceed, and be assured that their careers and research will benefit rather than be harmed.

Part of learning about electronic publishing involves understanding copyright laws and issues regarding intellectual property rights. Through extensive discussion with concerned parties, Virginia Tech developed an approval form, available on the WWW, that is currently printed, completed, and signed by the student and all faculty committee members, specifying who can access their work and when access might be extended. This allows a student several options—to make the work available worldwide, to make parts available worldwide and parts available only to the local campus, to allow access only to the local campus, or to secure the ETD in the case of patent or proprietary concerns.

Restricting access to the campus (of the whole work in the rare cases when a book version of the thesis will be published, or of a chapter that is very similar to an article that will appear in a journal) satisfies concerns of publishers that require there be no prior publication of works they accept, but puts greater restrictions on

*continued on page 4*

ETD's *continued from page 1*

traditional library access services such as interlibrary loan. Accordingly, we hope that publishers will agree to broader access after some time has passed following the release of their publication. While some publishers seem content with a delay of three months or one year, we hope that all will agree to some reasonable period of time. We expect that such an arrangement will have no real negative impact on publishers or their revenues, and argue that having authors of journal articles more knowledgeable about electronic publishing and digital libraries will save them considerable operational expense.

Readers of journals are accustomed to high quality articles that have benefited from peer review and editorial processing. In many cases, these are based on ETDs that are much longer and more detailed. Thus, we expect ETDs to evolve into an important genre that supports journal publishing. Readers of ETDs will realize that these have been subject only to the critical review of a faculty advisory committee. ETDs do, however, contain much information which is not included in derivative manuscripts, e.g., lengthy literature reviews, extensive bibliographies, details on experiments, and student resumes. Because of this content, we would prefer to see them be freely available, especially to graduate students who want access to the latest works and who may model their own theses after high quality ETDs. CGS and many other organizations regularly give awards for outstanding theses and dissertations; clearly some do value these documents highly!

What is the future of the ETD initiative? Virginia Tech shares its submission and archiving software and provides training to any institution in the NDLTD. Virginia Tech staff (reachable by email at <etd@vt.edu>) will provide a videotape about NDLTD upon request, visit interested universities, or engage in video conferences as desired. Extensive on-line information is available at <http://www.ndltd.org/>. Virginia Tech also is developing tools so students can submit ETDs using the Standard Generalized Markup Language (SGML, the "parent" of HTML) as well as in PDF. SGML documents are more easily archived, more easily searchable, more reusable, and therefore are more valuable to scholars. SGML also makes it easier to interconnect rich multimedia documents, such as those using standards like JPEG for images, or MPEG for video.

Virginia Tech is coordinating development and implementation of a distributed digital library system, so ETDs from all participating institutions can be locally stored but still be easily accessed. This will allow browsing and searching (based on institution, date, author, title, keywords, full-text, image content, etc.), as well as downloading. OCLC and IBM are assisting with these developments, helping apply the latest standards and technologies.

Hopefully this editorial will help graduate deans and others understand the potential and real benefits of this project, and to realize that, contrary to what some have claimed, it is not a threat to the employment of graduate students in academic positions, not a threat to faculty promotion and tenure, and not a threat to the publishers who through the peer review process improve derivative manuscripts that are based upon the rich mine of information contained in ETDs.