Carrots or Sticks: Getting Students to Submit Electronic Theses at MIT

Keith Glavash, MIT Libraries
Bill Comstock, Harvard College Library*
Larry Stone, MIT Information Systems

Abstract

MIT completed a pilot project in the fall of 1999 to collect doctoral theses in electronic format from students in three departments. Despite the technical savvy of most MIT students, only about a third of the sample actually chose the option and succeeded in submitting their PDF document. We discuss what happened and why electronic thesis submission isn't yet required at MIT.

Introduction and Background

For at least the past half dozen years there has been talk – ranging from casual to serious, from nearly every constituency which ever would have reason to touch an MIT thesis – to explore and move forward in establishing electronic theses at the Massachusetts Institute of Technology. No one with whom we have discussed it has ever voiced opposition to the fundamental idea. In fact, quite the opposite response has consistently led us to believe this was not a question of "if", but rather "when". The potential advantages – from access and searchability to physical space savings to new forms of content and construction – are plain to see. To no one's surprise, working through the devilish details to make this happen has been both interesting and challenging, and promises to occupy a good deal more of our time in the future.

The most fundamental challenge in trying to make a change to MIT's thesis requirements has been to identify common goals and realistic practices amongst the vastly different sub-cultures of its five schools. There is no overarching "Graduate School" as such. Each school – architecture and planning, engineering, humanities and social science, management, and science – operates its graduate degree programs with a great deal of independence, a hallmark of MIT's culture. The current requirements for thesis documents are, in fact, one of the few common threads that run through the Institute's graduate programs. This, however, does not imply that a change can easily be negotiated. Despite the atmosphere of high technology associated with MIT, variations in computing resources and practices are wide.

The MIT computing environment is a complex landscape of UNIX, Windows, and Macintosh. Since the only standards ever imposed on theses pertain to paper, authors are free to use whatever software they like for word processing. It would not be acceptable here to place any limits on the students' choice of word processing software. The challenge for us is to design a process through which *most* electronic document formats can be converted to a single standard format *easily.*

Despite these long odds, there is considerable motivation to work toward this goal. An e-thesis will, for example, be a much more elastic data repository, able to capture data-sets, simulations, video, etc. It can be searched, indexed, hyperlinked to other research publications, and made

---

available to remote researchers without first needing to be digitally reformatted. Our test project was motivated by mounting pressure over several years from both inside and outside MIT to begin to collect and display theses in electronic format. This most often was relayed in comments from students, faculty and outside researchers that suggested MIT was behind in this area. Besides our fundamental attraction to an electronic option, we had a strong desire to create a substantial digital library of theses, ultimately to increase the dissemination of MIT research to the world.

MIT is one of the only American universities which continues to publish and distribute its own theses. We submit abstracts to Bell & Howell's *Dissertation Abstracts International,* but distribute the theses ourselves on a cost-recovery basis. Sustaining this business model depends on our ability to market and distribute the documents effectively and economically. Given the usefulness and growing demand for theses in electronic format, capturing digital versions at the time of submission is a very attractive alternative to scanning.

Although we have significant concerns about our ability to substitute electronic theses for traditional paper at MIT in the near term, there is little disagreement about the inevitability or the potential advantages of this trend. And although the initiative we will describe did not address the larger issue of redefining the nature of an MIT thesis in an electronic medium – what will surely be a very interesting but time-consuming process – it was the all-important first step into the future of theses at MIT.

Start-up
In May of 1998, a group of staff from the MIT Libraries and MIT Information Systems, with the support of two faculty members from the Lab for Computer Science, proposed an e-thesis pilot project. The proposal called for a small, controlled, optional program, running parallel to the paper-based procedure. Students would be allowed to submit their thesis documents to a secure Web site in Adobe Portable Document Format (PDF). The proposal was presented to MIT's Committee on Graduate School Policy where the response was generally positive, although concern was voiced in several specific areas: the potential to spoil commercial publication prospects for students in certain disciplines, particularly in the School of Humanities and Social Sciences; authentication and certification; patent holds; and exceptional formats.

Although the project was purposely small in scale, we wanted to gain experience working with several departments of different character if possible. Three departments were ultimately chosen: Brain and Cognitive Sciences, Chemical Engineering, and Electrical Engineering and Computer Science. The scale of the test was further limited by the fact that we chose the February 1999 degree period, i.e. only those doctorates granted at the mid-academic year stage. In all, there were about 35 potential doctoral theses to be submitted among the three departments.

The process of convincing the three departments that this would be an attractive option to offer their students emphasized to us the extremely stressful nature of the thesis submission period for students. Graduate administrators were insistent that we make the process as simple and fast as possible and that no additional pressure be placed on their students. While not entirely skeptical, they communicated their reservations about our ability to gain the students' attention for this.

Attracting Participants
Our search for incentives with which to attract students to this option was difficult. Unfortunately, submitting an electronic thesis would not yet eliminate the need for a paper document. Although the electronic version was meant to be an exact duplicate of the paper document, there was nothing in place other than author certification to assure that the electronic

and paper thesis versions were identical. We eased the paper requirement to a single copy – down from two – but it carried little value in a campus computing environment that offers free printing.

In addition, many issues surrounding the collecting and processing of e-theses had to be resolved before the Institute Archives could accept an electronic substitute for a document they were responsible for adding to their permanent collection. Therefore the paper copy would continue to be the "official" thesis with the required signatures, submitted to and held by the MIT Archives. So by asking the student to submit an e-thesis, we were adding another step to an already long process of thesis approval and submission. We learned that asking for one more event to occur during the final stages of thesis submission – even if innovative and attractive – is considered too much by the majority of MIT students.

Some of us were of the opinion that e-theses would present a strong attraction to students simply because of their novelty at MIT, or because students would see value in getting their research disseminated so quickly and widely. In repeated messages urging them to use this option and offering technical support, we pointed out that they would be among the first MIT students to submit an electronic thesis, something of near certain historical significance in the Institute. Alas, very few agreed or saw any value in that prospect.

One added feature did seem to improve our chances of getting a submission: we created a submission site for students who had already gone through the traditional paper submission process and moved away from campus. Some of these students were still interested in submitting an electronic version, but just had not had an opportunity to consider the e-thesis option before the flurry of graduation and relocation. However, because only current students hold MIT Web certificates, this site had to be "insecure". Unauthenticated recent graduates were allowed to use it because it was easy manually to certify such a small number of authors, but such a practice would not scale up.

Project Results
Twelve of a possible 35 theses were successfully submitted in PDF. Another half dozen have been added since that time by special request, which we have welcomed. These numbers represent a reasonably good return, but took months of labor-intensive personal appeals to realize – an approach which, once again, simply does not scale. More importantly, we were surprised by the large number of technical problems encountered by students in converting their documents to PDF. For example, some PDF's came in with broken fonts because they were generated by TeX (see below) and needed to be reformatted with scalable fonts during conversion. In a few cases we had to upgrade to the very latest release of Aladdin Ghostscript in order to process the submitted PDF file successfully. Ghostscript was chosen to validate the PDF and convert it to page images because it is the best free software that can run unattended on our Solaris server. We still consider PDF a far better choice than PostScript, because different word processors take different liberties with PostScript. But many of the "successful" submissions still required hand tooling. This is really the crux of the problem: if the process of PDF conversion had been as straightforward as we had anticipated, many more students would have submitted e-theses. When the conversion did not go smoothly – and it rarely did – students decided it was not worth the trouble.

MIT students – especially those in technically-oriented disciplines, the vast majority – construct their thesis documents with a wide variety of software applications, from standard word processing programs to relatively obscure graphics tools, frequently using a combination of several to produce the thesis. Tying such a document together in a single PDF file can be a challenge. LaTeX and

TeX – word processing programs frequently used in equation-rich documents – are particularly troublesome, and both are used extensively here.

Process Design
Designing the electronic submission process had two overriding challenges: first that it be simple and fast enough to be attractive to students, and second that it produce an identical digital rendering of the paper document. In the course of overcoming those requirements, many specific issues – including student authentication, format conversion, metadata collection, timing, and thesis "holds" – were worked through by the project team. From all indications, the user interface presented no problems for students. The forms were clear and the uploading process worked well.

Thesis holds – most often due to the time-consuming patent filing process – are a sensitive issue at MIT. The Institute's policy is to release as many theses as possible to the public as soon as possible in order to disseminate research results quickly. However, MIT's Technology Licensing Office, which is responsible for filing patent applications, is not always aware of potential patents prior to thesis submission. If it suddenly becomes apparent that a thesis may have patent possibilities, there is frequently a scramble to locate copies and place them on hold. The traditional paper submission process is slow enough to allow most of these situations to be identified and rectified prior to release. But a dramatically speeded-up submission process can spoil patent opportunities, especially when linked to online publishing (disclosure) of the document, as it is here. Hence for the purposes of our project, it was necessary to build in a one-month delay between submission and online display. We hope to improve the patent identification process in the future so that most documents can be released online as soon they are available. We recognize that this is not necessarily ideal for some students who have commercial publishing plans. Most likely, we will allow students to exercise some control over online access for a limited period.

E-theses are added to MIT's Digital Library of Theses (DLT) http://theses.mit.edu, a freely-accessible collection of page images including nearly 4,000 thesis titles spanning the past 100 years. When an e-thesis is submitted, the digital library software (Dienst) converts it into 100dpi page images with 32 levels of grayscale for optimal Web viewing. We have found this to be the best apparent resolution at a reasonable download time (average 40Kb/page). The PDF's are available upon request for a fee.   MIT's Digital Library of Theses is the topic of a separate paper – *Harvesting the Low-Hanging Fruit: World Wide Web Access to a Collection of MIT Theses* – also being presented at this Symposium. Its popularity, and the growing demand for additional titles, is motivating us to look for alternative ways to incorporate new theses immediately upon publication. Ultimately, finding solutions to the electronic submission problems we have encountered will eliminate the need to carry out any digital conversion, and will allow us to build a much more valuable online collection.

What next?
At the outset of this project we were hopeful that electronic submission could be scaled up to include most departments over the relatively short period of two to three years. While that time frame is not out of the question, we are now convinced that until we can offer much improved conversion technology, we will not gain significant voluntary participation in electronic thesis submission.

We will continue to support electronic thesis submission passively by allowing doctoral students from any department to do so if they choose. And we are eager to work with other institutions in the areas of new authoring tools, format conversion, authentication, digital watermarking – all

areas which soon must be addressed here as we proceed toward our ultimate goal of fully implemented electronic thesis submission. MIT's recent joining of the Networked Digital Library of Theses and Dissertations (NDLTD) will help us to keep involved and informed along those lines. On the local front at MIT, we need to help facilitate and shape a dialog within the academic ranks about the nature of a thesis in the electronic age.

Just as importantly, we believe the growth and development our retrospective Digital Library of Theses presents the strongest motivation to move forward with electronic submission – both for the developers and for the students whose work will occupy its virtual shelves – because its increasing value and popularity attract more attention and demand more content. We simply must find easy ways to build this collection with new theses. The success of the DLT is already pushing us towards two promising enhancements: abstracts and tables of contents for all our online theses will soon be incorporated into the bibliographic records, resulting in an enormous improvement in subject searchability. And we are making plans to digitize whole groups of new theses as they are submitted, in order to disseminate the very newest research as quickly as possible. Although these actions may appear to facilitate a delay in the route toward electronic thesis submission at MIT, we feel they ultimately will help strengthen our case for it, and avoid a loss of valuable content in the mean time.