

Taking the Plunge: Requiring the ETD

Kimberly Douglas

Betsy Coles

George Porter

Eric Van de Velde

California Institute of Technology

kdouglas@caltech.edu

MC 1-43, Pasadena, CA 91125

<http://library.caltech.edu>

Keywords: Liaison librarian faculty graduate dean copy of record voluntary submission file size LaTeX template Frame-maker thesis regulations persistent URL unique identifier

Abstract

It made sense for Caltech, the California Institute of Technology (a private, technically focused, U.S. university, <http://www.caltech.edu>), to go electronic when it comes to theses. It took, however, more than three years: From March 1999 when Prof. Ed Fox of the Virginia Technical University spoke at Caltech to July 2002 when ETDs became required for all PhD candidates. How was it done and what are the lessons learned?

Unfortunately, the University Microfilms has taken a non-exclusive position and is committed to peaceful coexistence with the NDLTD.

Phase One: Voluntary Submission of the ETD

The Caltech Libraries successfully launched its ETD effort to the Caltech community in the spring of 2001.

Introduction

The California Institute of Technology is a small, highly focused private research university in the United States that values being on the leading edge of new technologies. The administration and faculty expect individuals, campus organizations or departments to achieve that objective by self-initiated entrepreneurial methods. This is instinctive in the research culture on campus and is therefore not necessarily consciously acknowledged nor even understood by the practitioners. Therefore, it behooves the library administration to so thoroughly understand this culture that innovative projects are launched with a minimum of bureaucracy and process. Yet it must still be done well and successfully.

Caltech produces about 150 PhD theses per year. These breakdown as follows: Engineering = 42%; Chemistry 23%; Physics 12%; Biology 10%; Geology 8% Math 2%; Political Science 2%; Economics 1%. In the spring of 1999, Ed Fox came to campus and made a presentation to enlist increased participation in the electronic thesis (NDLTD¹) effort. Both the Vice Provost and the Graduate Dean attended that presentation and with continuing agitation and encouragement from the library administration, the Graduate Dean agreed that Caltech should embark on this initiative. The one issue that concerned him was the impact this project would have on the contractual relationship that Caltech and most U.S. universities have with University Microfilms now Proquest. For-

Preparation

In the two years between March of 1999 and June 2001 library staff prepared the basic platform for the ETD:

1. Selected and implemented an applications program. A single applications developer was identified who worked on this project part-time in combination with other duties. The ETD-db (NDLTD) software was selected largely because it was free and was openly supported by the developers at Virginia Tech. In addition it was OAI compliant; it met design criteria necessary for a repository of theses and provided a functional user interface. This software met our requirements of no cost and it supported the capability to restrict access to specified files. The latter was not supported by other applications, such as Eprints.
2. Established the policy and syntax for a unique and persistent identifier and in connection with the library's other digital collection projects developed a permanent resolver mechanism, the PURR.² Each thesis is assigned a unique identifier by the ETD-db (NDLTD) software. The identifier is exported in the OAI_dc metadata (OAI Dublin Core metadata subset) and is used in the URL to point to the service page. Since it is unique to the archive, it can be used by the resolver to make an ETD Persistent URL. Even though the identifier number is created by this specific application software, the Caltech Libraries will permanently maintain

¹ National Digital Library of Theses and Dissertations

² Sponsler, Ed (2001) PURR - The Persistent URL Resource Resolver. <http://resolver.caltech.edu/CaltechLIB:2001.003>

the same unique identifier for the thesis even if the application software were to change.

Example:

```
Unique Namespace Identifier:CaltechETD
ETD-db Unique Local Namespace Specific
String or identifier:
etd-11252001-103911
ETD Persistent URL:
http://resolver.caltech.edu/CaltechETD:etd-
11252001-103911
```

- Continued developing interest on campus and gaining permissions of the Graduate Deans. Caltech had undergone turnover in Graduate Deans. The Grad Dean of 1999, who was present for Ed Fox's presentation in the spring of 1999, was not the same Dean in 2000-01 nor the same in 2002-03. The primary rationale was the necessity of advancing Caltech's operations into the networked digital environment. Caltech's deans are also active researchers with graduate students and a teaching requirement. The library's arguments resonated with the perspective that the students were also bringing to the table.³

Experience

In June of 2001, the then Graduate Dean announced that PhD candidates who wished could voluntarily submit an electronic version of their theses to the Caltech site. The library provided only very minimal support at that time. The student needed to have already prepared a PDF file from whatever source application they used. A PhD candidate who wished could connect to the Caltech ETD site and submit the electronic version of their thesis.

Of the 299 theses of 2001 and 2002, sixty-three (63) were submitted voluntarily from the following disciplines: Engineering_22_; Biology_9_; Chemistry_12_; Geology_8_; Physics_11_; Math_1_. None were submitted from Economics or Political Science.

The students who participated the library's effort experienced very few problems. Many students also voluntarily submitted positive comments referring to their desire to make their work more generally available. The students saw the posting of their thesis as a constructive step in making themselves known within the research community.

Phase Two: Requiring the ETD

Now that interest had been raised and the library staff were gaining practical experience at a volume that al-

lowed for constructive learning, it was time to address the issues necessary for a successful mandatory submission.

Preparation

Over the following months, September 2001 to April of 2002, library staff addressed the issues that essential to making the ETD submission mandatory. The Library established a team of four: Head of Technical Services, two subject librarians and the SYSTEMs application developer maintaining the ETD-db software. In addition, one of the senior library managers provided guidance and encouragement as needed to negotiate new relationships with interested parties on campus. Together this group worked through the various implementation issues: Refining policies and procedures with the Graduate Office and guiding users to appropriate tools and instructions via a website and in briefing sessions in the library;

- A Website <http://library.caltech.edu/etd/> with advice, guidance, requirements and instructions was created. PDF was established as the primary acceptable format. For multimedia files the library chose to accept the same list of formats that University Microfilms accepts. The official document of the Graduate Office, "Caltech PhD Thesis Regulations" as revised to include the ETD submission preparation and process. This was done by library staff rather than await the availability and inclination of the Graduate Office staff
- The team developed a briefing session for the students providing the essentials of what they needed to know and do to prepare an electronic thesis. In these sessions issues of acceptable fonts, preparation of images; size of single file theses and how to segment a theses for optimal usability; creation of multiple versions to address various resolution and color options; naming the files for optimal presentation were covered. In addition copyright issues are explained both from the authors and the publishers' point-of-view. Subject liaison librarians were also educated to these issues to lead orientation classes for graduate students and to participate in trouble-shooting responses.
- Library staff established a cooperative relationship with staff of Caltech's Digital Media Center who provide technical support for digital publishing tools. With the launching of the required electronic thesis, the DMC offered a specific class instructing authors on the mechanics of creating a PDF for Long Documents and Dissertations.⁴ In collaboration with past graduates templates for Word, LaTeX and Framemaker were created and offered to the students.
- The ETD-db interface for author submissions was revised to meet Caltech needs and to provide additional guidance in the process.

³ Graduate students in 2001 were, on average, about 26-27 years old. They were about 20, juniors in college, when the WWW gained ubiquitous following.

5. Library procedures for handling print theses were revised. Only one paper copy is now needed and it is often printed from the submitted etd. Since information about the thesis committee for the thesis is entered into the metadata for the etd though it is not actually carried in the work itself, catalogers add that information to the catalog record.
6. The library continued to require the printed theses, though it was expected to be a bound copy of the printed PDF file to ensure that the printed copy and the electronic copy were the same. Furthermore the printed theses would continue to be the "copy of record" or the archival copy.
7. The Library acquired a campus site license for the Adobe Acrobat package greatly facilitating production of PDF on Windows and MacOS workstations. Further, it provides the necessary tool for combining and segmenting PDFs for optimal access management.

In April of 2002, the Library staff made a formal presentation to the Graduate Studies Committee at which the new Website for the ETD was debuted. Later in May 2002, the Dean of Graduate Studies made the ETD submission required for all PhD candidates as of July 2002.

Experience

The most serious concern was raised by the Graduate Student representative, who, though recognizing that most Caltech graduate students would have no difficulty preparing a PDF document, was concerned that there would be a few who would encounter difficulty. Given the deadlines that a Grad student has to meet, there was concern that students not be faced with more steps. The committee did discuss without conclusion the nature of the printed theses versus the electronic theses. The library decided that it would continue to add a printed copy of the theses to the collection. In fact, the library decided to keep the printed theses as the version of record thus avoiding a potentially controversial issue. At the meeting, it was clear that the faculty still felt very strongly that the research results in a thesis needed to be "written-up" in a proper manner. It was considered still very much necessary that a student articulate in English the nature of their work. Media, images, datasets would not suffice to describe and document the work, although such could be added to augment the work.

The primary issue had to do with the size of files. Some astrophysics theses with images could be as much as a few thousand MB⁴. This was not reasonable for downloading. Guidance for dividing a single thesis into separate files needed to be established. In addition, questions were raised about how to restrict access to portions of a thesis for patenting purposes. There were also a few questions having to do with making editing corrections to

the thesis after it had been uploaded. Some students followed up to be sure that the abstract to the thesis was presented well in the NDLTD environment.

Conclusion

Two deliberate decisions greatly assisted quick implementation of the etd requirement: The subject liaison librarians were involved in the implementation. This spread the workload over a larger public service staff resulting not only in a manageable additional burden to each but also in gaining their energy and excitement toward making this project a success. Their liaison responsibilities and experience facilitated reaching out to the various departmental groups on campus. Also, the premier role of the bound printed thesis in the traditional launching of an academic career was not challenged. The library decoupled the concept of requiring an electronic thesis and establishing the copy of record. Students and faculty are more likely to try new approaches if the risk is low. Faculty, particularly, are conservative by nature and are legitimately concerned about the career impact on their students of some of the changes in the digital era. By maintaining the print thesis as the copy of record, the electronic version while including the text of the printed theses can contain additional content in different formats or even different versions of the thesis to accommodate more adventurous presentation media. We are just beginning to see this occur as we conclude the first year of required electronic theses.

The Caltech Libraries implementation of the ETD requirement was definitely one that required planning and an organizational willingness to learn from concurrent experience. Not all issues were solved or were even explicitly addressed. Archiving procedures remain to be thoroughly worked out. To a large extent it is our assessment that all perspectives on true archiving are somewhat preliminary and await the promulgation of ubiquitously followed standards and robust tools for both author and digital archivist. These are evolving certainly. In the meantime, we do encourage authors to submit the source files of their thesis whenever possible. These are kept hidden from the database user and are intended to be used solely for file format migration purposes. Additionally, as more students responded to the requirement of an ETD, more faculty have been tangibly touched causing challenging feedback in regard to the bureaucracy and formalities involved in the awarding of a PhD. Nevertheless, the overall commitment by the faculty is great enough to assure that the questions will be overcome.

4 Carolyn Patterson of the Digital Media Center at Caltech prepared extensive explanations on how to package large documents in pdf. These instructions are available at: <http://morel.caltech.edu/classes/pdfs/0612PDF%20for%20Long%20Documents.pdf>

5 To be specific, a biochemistry thesis came in at 7500 MB. The student did segment it so that all portions were smaller than 300 MB.

Not all problems are completely solved. It is absolutely not a turn-key process. We realized that some issues will need to be resolved over time. The voluntary ramp-up period allowed library staff the window to develop the essential tools and procedures and to identify the critical issues. The most important result is that we have managed to create a climate wherein all students must partic-

ipate yet with little or no risk. This allowed the campus community to demonstrate to themselves that creating and submitting an ETD is neither onerous nor impossible. We've been successful by being flexible and nimble to respond aggressively, knowledgeably and quickly when needed.