Using ETDs to Overcome Obstacles to Teaching Digital Literacy and Content Management

Gabrielle V. Michalek
Principal Librarian/Archivist
Carnegie Mellon University Libraries
Gabrielle@cmu.edu

Abstract

Digital literacy and careful content management are essential prerequisites to success in higher education today. Intellectual prowess and solid study habits are necessary, but no longer sufficient. Students collect and produce many files of digital content in their journey toward a degree, and being able quickly to locate, understand, and open the right version of the file when the pressure is on can be crucial to their careers.

Research libraries have added data management training to their instruction curriculum to increase graduate students’ digital literacy and content management skills. The Educopia Institute has collaborated with twelve research libraries to develop and pilot the ETD+ Toolkit to support this initiative. Attempts to deliver data management training, however, immediately encounter obstacles. How does the data management expert capture and hold the attention of busy graduate students? Who can bring students together as a captive audience? Surely not the data management expert alone. Formal connections and collaboration with campus faculty and senior management are required to motivate and focus student time and attention on data management, which presents another challenge. How do we forge these connections?

At Carnegie Mellon University, the ETD+ Toolkit was the catalyst that captured the attention and enthusiasm of the Assistant Vice Provost for Graduate Education. Now, working in partnership with the Office of Graduate Education (OGE) and using the ETD+ Toolkit, we can teach graduate students best practices for managing their research outputs from inception through deposit of their thesis or dissertation in the University Libraries’ digital repository. The partnership has also inspired much-needed conversation with students and faculty about scholarly communication, copyright, and research data.

In this session, I will discuss how Carnegie Mellon University Libraries used the ETD+ Toolkit to improve access to graduate students and to create meaningful relationships with them. I will offer effective strategies for assembling audiences of graduate students and imparting to them skills that improve the quality of research data files associated with ETDs. For research libraries, the urgency of the students “need to know” translates into the urgency of forging collaborations with those who can influence and motivate their behavior.

Keywords: ETD Toolkit Development Project, Research Data Management, Marketing Library Services
Carnegie Mellon University’s ETD Landscape

Carnegie Mellon University (CMU) is a global research university that offers many outstanding education and research programs. Graduate programs span the fields of the arts, humanities, engineering, sciences, and public policy. CMU offers world-class graduate programs based on an interdisciplinary approach to education. Faculty and researchers from across the university work together to create dynamic programs that support the evolving needs of society (e.g. Robotics, Artificial Intelligence, and Engineering & Arts).

While students acquire a timely and relevant education, the CMU environment presents certain challenges to the management and administration of these programs. For example, Carnegie Mellon does not have a centralized Graduate School overseeing its graduate programs. The separate colleges and departments manage their own graduate programs with some guidance and support from the university’s central administration Office of Graduate Education. Other programs and offices are also available to meet the needs of graduate students (e.g. Office of International Education and Graduate Student Assembly).

In CMU’s decentralized environment, each college establishes the policies and workflows for the creation and management of electronic theses and dissertations (ETDs) in their degree programs. No two workflows are alike. One college may have a couple of staff members who are responsible for guiding the ETD workflow for all of the departments within the college. In another college, individual departments within the college are responsible for managing their ETD workflow. Policies and guidance documents likewise differ. For example, some colleges address how supplemental files are to be submitted, while other colleges or departments do not mention how these files are to be treated. There is no single university policy or workflow for ETDs and their supplemental files.

The handling of supplemental files associated with ETDs is a significant problem in many ways. To complicate matters, supplemental files in the various academic fields are created in different formats. Until recently Carnegie Mellon did not have a platform capable of supporting a variety of formats of the supplemental files associated with ETDs. In the past, even if a student submitted supplemental files there was no easy way for the university to maintain and provide access to those files. In 2007, the university’s institutional repository, Research Showcase, began to be managed by the University Libraries. This platform primarily supported PDFs but could also ingest audio and video files without streaming capabilities.

Prior to 2014, the Libraries received some ETDs and supplemental files, but standard practice was to submit a paper manuscript, which was sent out to be bound and then shelved. In 2014, the Libraries created a new workflow that would encourage the creation of ETDs and allow the electronic files, rather than paper manuscripts, to flow into the Libraries. As the de facto central warehouse for all ETD’s and their supplemental files, the University Libraries is responsible for administering and making available the research datasets and complex digital objects (multimedia files, software, etc.) that often accompany these scholarly works as supplemental files. While the previous institutional repository platform fell short of supporting discoverability and access to supplemental files, the University Libraries used alternative approaches to providing these services. ETDs with supplemental files contributed to the Libraries were placed on a server by a system administrator, normalized, and given a DOI for persistency. The works were cataloged and made available through the library catalog. This process revealed several
structural and technical problems concerning supplemental files. It was these shortcomings that motivated CMU Libraries’ participation in the ETD+ Toolkit Development Project hosted by Educopia.

The new workflow that increasingly brought ETDs and their supplemental files into CMU Libraries presented the curation and preservation challenges in terms of the viability of formats, file backup, management of intellectual property rights, and dissemination pathways. The urgent need for students to know how to name their files consistently, how to organize them, how to control the various versions, how to select viable formats and use persistent URLs, among other issues, quickly became clear. A system administrator was initially responsible for addressing these issues, but given the scale of the problem this approach was unsustainable.

Recently, the Carnegie Mellon University Libraries entered into a contract with Figshare to become the new university institutional repository. The intention is to migrate all ETDs from the previous platform into Figshare and place new ETDs with supplemental files into the new platform moving forward. The good news is that the Figshare platform will support a broader range of file formats and makes ETDs with supplemental files more accessible. The bad news is it will also introduce additional challenges when it comes to long term preservation and access to those files because it requires the author to engage in some basic data management practices in order to improve the files viability over time.

As the 2014 ETD workflow directed ETD files to the Libraries, the Libraries was also embarking on a new program directed at Research Data Management. The work of the ETD+ Project was seen by CMU Library Administration as a springboard into identifying better data management practices for a whole range of different research data content types including ETDs.

ETD + Project and Carnegie Mellon University Libraries Collaboration

At a Steering Committee meeting in 2013, Katherine Skinner of the Educopia Institute reached out to its member communities asking if they would be interested in participating in a grant sponsored by the Institute of Museum and Library Services (IMLS) to collaboratively study curation and preservation for ETD research data and complex digital objects. At the time, CMU Libraries was very concerned about the preservation of born digital content, and belonged to the MetaArchive Cooperative, one of Educopia’s communities. Given our growing experience with ETDs and their supplemental files, we agreed to participate in the project. The grant proposal was prepared in the spring of 2014 and by September of that year, IMLS announced it had selected the project for funding. Work on the project officially began in January 2015.

Having vast knowledge and experience with ETDs, members of the Educopia network decided to focus on creating a number of deliverables that would support preservation of graduate student research associated with ETDs. These deliverables included guidance briefs, workshop materials, software tools and evaluation activities that can be used by students and staff to manage supplemental files associated with ETDs.

Creation of all project deliverables was based upon data gathered from two surveys conducted by each of the universities participating in the ETD+ Project. (A partial list of those participating include: Carnegie Mellon, Indiana State, Morehouse, Oregon State, Penn State, Purdue,
University of Louisville, University of Tennessee, the University of North Texas, and Virginia Tech.) The project partner at each university distributed one survey to a representative sample of graduate students from the Masters and PhD programs across the university, and the other survey was distributed to university staff responsible for the management of theses and dissertations. The surveys sought to gauge the use, abilities, and needs of these two distinct populations in regards to supplemental ETD materials. At CMU, prior to conducting the surveys, the project partner submitted the survey instruments to the university’s Institutional Review Board (IRB) for approval because CMU treats surveys as qualitative research on human subjects that must comply with federal regulations. After receiving IRB approval, the surveys were distributed to representatives from each college, department and program.

Data from the surveys was helpful for the project, but also proved to be informative on the local level. The full dataset was analyzed by members of the Educopia staff to support the project’s preservation and curation goals. In addition, each participating university received survey responses from their constituents. The responses revealed student and staff perspectives on local ETD practices, and helped CMU Libraries personnel understand the attitudes, practices and perceived challenges of both groups.

One of the more interesting findings from the surveys involves student and staff understandings of whether supplemental files associated with the main body of the thesis and dissertation can even be submitted to the institution as part of the submission process. Carnegie Mellon’s result concerning this question matched the responses from other universities working on the project. That is, neither students nor staff clearly knew whether submission of supplemental files was an option. Despite CMU Libraries having a clear policy and workflow for the submission of ETDs and their supplemental files, the university’s decentralized environment and lack of a central graduate school exacerbates the problems of establishing a single set of policies and then communicating those policies to the relevant audience.

Another important finding from the surveys was the strikingly different attitudes among staff and students at each of the participating institutions concerning the student’s preparedness to execute data management practices. Students believed that they were already competent at data management practices and that they needed minimal intervention while university staff believed that students were not well trained in data management and required additional training and support. This finding did inform the creation of the project’s deliverables, packaged as the ETD+ Toolkit and resulted in the creation of a slide presentation directed at ETD administrators.

The Toolkit components were created by team members from all of the participating institutions. The primary components of the Toolkit are guidance briefs and workshop materials. The guidance briefs are short (3-4 page) “how-to” documents designed to assist student researchers in understanding how their approaches to data and content management impact credibility, replicability, and general long-term accessibility: knowledge and skills that will impact the health of their careers for years to come. The workshop materials include a slideshow and handouts that serve as in-person training materials that guide students through a series of six modules, each related to a different facet of research data management, including copyright, data organization, file formats, metadata, storage and version control. Instructors use the workshop materials and reference and share the guidance briefs when teaching the modules. The guidance
briefs are stored in a central location that students can access when they need additional information. Members of the Libraries and Office of Graduate Education staff make sure that students are aware of these materials.

The ETD+ Toolkit materials also includes pre-workshop and post-workshop surveys where members of the project team request feedback that could be shared with our funder and that could be used to improve the Toolkit components. Surveys are distributed by a staff member in the Office of Graduate Education to those students who have signed up for each workshop. The pre-workshop survey measured what students knew about the topic prior to the workshop and the post-workshop survey measured what students learned as a result of attending the workshop. Distributing the surveys was somewhat problematic at CMU since library personnel do not have access to graduate student signup sheets or listservs and have to rely on staff members from the Office of Graduate Education to share requests for feedback. Student responses to the surveys provided meaningful information that continue to shape how the modules are marketed and taught at the university. For example, the instructor at Carnegie Mellon discovered that students learn best when provided examples of university generated research data that demonstrate best and worst data management practices that impact the research’s long-term accessibility and interoperability. These examples have been incorporated into the instruction program.

Throughout the project the University Libraries has had to rely on the kindness of other units at the university to collect feedback, test and redesign the content found within the Toolkit modules. After the project was completed, seeing the value of teaching students research data management skills, library administration and the Office of Graduate Education agreed to offer the workshop series on a more formal basis. The next challenge was to build an audience for the content.

**Libraries Lack Captive Audiences**

There was a time when every person working on a degree needed to visit the library for access to information and research support. As more trusted content became available online, the number of students and faculty making in-person visits to their library significantly decreased. Meanwhile, the U.S. government issued unfunded mandates to researchers to publish federally funded research and the associated digital data via open access. These mandates compel universities to provide support to help researchers comply with the mandates. Sensing an opportunity to expand their role and use of their information management skills to support the needs of the university, academic librarians stepped forward to help shape policies, deliver services, and develop and manage infrastructure – thereby, aligning the role of the library with broader university missions and further enhancing the library’s reputation.

Of the three main areas of support that included policy creation, service development and infrastructure management, many academic libraries chose to leverage their information management skills to develop new research data management services. These programs, including the EAD+ Toolkit Project, are an attempt to train faculty and students to use sound data management practices to help ensure preservation, discoverability and interoperability of research publications such as ETDs and their associated digital files. To comply with federal mandates, researchers need to master these skills. However, delivering these new services is
challenging because it requires librarians to engage with new audiences in different, more creative ways.

When a faculty member feels that students need to acquire a skill set to be successful in their field, they can integrate this training into the course syllabus and require students to complete the work. Most academic librarians do not have this level of control over students or faculty in their departments. In order to attract members of the academic community, librarians must find ways of integrating themselves within the teaching and learning framework of the university. This is not always an easy task. At the CMU Libraries a two-pronged approach was used: direct contact through Social Media and mediated contact through other campus organizations.

Direct Contact through Social Media

One of the first and easiest approaches CMU team members did to build an audience was to become more active in Social Media to seek out students that we were trying to forge a professional relationship. By following faculty and students on Twitter and LinkedIn we gained a better understanding of the fields in which they are working, and it allowed us to shape our conversations with them when meeting one-on-one. Research indicates that use of social media increases a scholar’s citation count and impact score, suggesting that retweeting an announcement of new research published by the faculty and students is a good practice to community building. These people will appreciate your efforts at community building and help as they establish their careers and reputations. In this manner, you will build relationships with them based upon trust, respect and mutual support.

Another step that was taken to build an audience was to work with other members of the CMU library community that share a common concern for ETDs and research data management. This team provided additional necessary skills and offered fresh insights to solving the problem of audience building. Other librarians, especially those assigned to support learning and research activities to schools, departments and centers across the university were willing to collaborate since they experience the same challenges of audience building for their community. Each of these librarians are serving researchers representing a different discipline, or even subdiscipline, that have differing needs, expectations and desires of the librarians in their fields. Some faculty may be very willing to collaborate with the librarian, while some may be resistant to working with library personnel and grant access to “their” students. As a result, some librarians themselves may be willing to be more assertive about integrating themselves into the teaching and learning framework within their departments while other librarians may be more reluctant to take on these activities. Identifying those librarians that are willing to engage in audience building and those faculty who are willing to invite a librarian into their teaching and learning process is key to building an internal team that will work together on outreach activities.

Library personnel with marketing or design responsibilities also helped significantly with outreach and audience building by creating a marketing campaign. Posters, post cards, and digital signs in campus buildings were used to promote the workshop service and help build audience. The promotional materials were also distributed at various events.
Project members from CMU used the promotional materials to aggressively market the workshops at any event directed towards graduate students. These include the university-wide graduate student orientation, which is offered in the fall and spring. The orientation program exposes students to all of the services, programs and resources available to them. The Dissertation Bootcamp is another event where the marketing materials are used. This boot camp is open to all CMU Ph.D. students who are in or near the dissertation writing stage of their course of study.

**Mediated Contact through Other Campus Organizations**

The need to partner with academic units external to the Libraries is well documented. For the ETD+ Workshop series the Libraries primary collaborator was the Office of Graduate Education. The Assistant Vice Provost who leads this office, convinced that mastery of the skills taught in the ETD+ workshop modules was a necessary element of a digital literacy program, endorsed the workshops. She saw this program as being mutually beneficial to the OGE and the Libraries. Her endorsement opened the doors to administrative and financial support for each workshop. As the umbrella academic office for many of the programs that support graduate students at CMU, having the support of OGE helped to facilitate connections to ancillary graduate student programs and services.

A staff member from the OGE identified and scheduled locations for the workshop on campus, which is not a trivial task at Carnegie Mellon. Since the workshops were marketed by OGE and library personnel to every graduate student across the Pittsburgh campus, it was important to find a meeting space in a central location to reduce travel time. The workshops were taught by the Director of Connected Scholarship at Carnegie Mellon in a classroom auditorium that was equipped with modern technology and Wi-Fi access. This facilitated teaching and learning by allowing the workshop instructor the ability to respond to student questions that may be tangential to the topic and go “off script” by accessing websites or databases that were not part of the original lesson plan. Responding to student’s immediate needs in real time encouraged them to be more engaged in the workshop and made for livelier discussions.

Students were enticed to attend the workshops by free food. Students are always hungry and often low on cash. Providing pizza, salads and drinks immediately guarantee a larger audience. CMU Libraries does not typically budget for food for students; partnering with the OGE provided the funds to cover these expenses.

The OGE staff member also controlled the distribution lists for graduate students on campus, and was responsible for announcing times, dates and location for each workshop, and distributing to graduate students the promotional materials created by the Libraries’ marketing team. In addition, this staff member emailed to registered students the links to the pre-workshop and post-workshop surveys. Receiving these announcements from the Office of Graduate Education increased their level of importance, which helped to motivate students to attend. This collaboration allowed each of the units to focus on what they did best, and allowed the instructor to focus on teaching rather than administrative work.
Working with the Office of Graduate Education staff also created opportunities to network with the Graduate Program Committee, the Graduate Student Advisory Committee and the Doctoral Mentoring Committee. These committees are responsible for providing graduate students with feedback and support to help ensure they meet the requirements of their degree programs. As these committee members become aware of the University Libraries’ workshops in support of the ETD, they can direct students to the workshops and to the instructor for additional support concerning research data management issues.

While marketing a new service to Deans is helpful, at Carnegie Mellon we have found that the Associate Deans Group provides greater return on investment since the Associate Deans have more direct contact with the faculty and the students in their areas. Having an Associate Dean announce and provide a brief description of the service brings it to the attention of all of the departments within the college at the same time. This is also a great opportunity to distribute the marketing materials for the workshop series.

**Conclusion**

In 2018, ETDs are more than just a PDF. Today, ETDs consist of supplemental files representing a variety of formats. At Carnegie Mellon, new policies and workflows encourage students to deposit their ETDs into the institutional repository managed by Carnegie Mellon Libraries. The repository also facilitates researcher compliance with federal mandates requiring those with federal funding to make their research and associated data files available open access. Seeing the need for the development of tools and teaching resources that support research data management, the Carnegie Mellon University Libraries participated in the ETD+ Toolkit project. When the resulting series of workshop modules were ready to be delivered to graduate students, the University Libraries faced several challenges such as the lack of a central graduate school and the distributed approach to managing graduate programs on campus, the lack of a single policy and workflow concerning ETDs, and the lack of a captive audience for the workshops. To overcome this challenge, members of the University Libraries worked together to create a marketing strategy and promotional materials in an attempt to attract students to the workshops. Partnering with external academic offices proved to be both mutually beneficial for each other as well as the students. By using these approaches, the ETD+ Workshop Series has evolved into a new library service that is well attended and well received by the campus community.


The White House, President Barack Obama. This is “historical materials” and frozen in time. Website is no longer being updated. (https://obamawhitehouse.archives.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research).
