The authors discuss the development of a related set of institutional repositories among several liberal arts college libraries. Contrary to the usual focus on faculty publications, the primary goal of these repositories is the promotion of student work, especially undergraduate theses. Discussion of issues concerning selection of materials and archival policies is included along with practical considerations of workflows and reflections on the advantages and disadvantages of the particular software platform (Digital Commons). Marketing the repository and the subsequent addition of other materials, including e-journals, are discussed in light of ambiguity about its purposes among campus faculty and students. Serials Review 2006; 32:92–98.

© 2006 Elsevier Inc. All rights reserved.

Background of the Project

When discussions about institutional repositories (IRs) at universities and colleges arise, the focus frequently falls on faculty output, such as preprints and postprints of articles and classroom materials. However, as Trinity University’s librarians discussed the concept of the IR with colleagues from other liberal arts schools (initially Carleton, Dickinson, and Middlebury Colleges), we thought that student work might be a more interesting place to start. Each of our schools admits high-performing students, and students work closely with faculty. Many of these undergraduates produce senior theses, and students work closely with faculty. Many of these undergraduates produce senior theses, and the schools rightly point to their quality as a sign of the education they have received. In the past, these theses have languished in the stacks and probably received little use. Acquiring digital versions of the theses and placing them prominently in an IR seemed to be a logical first step for us.

Archiving and promoting our students’ work are sufficient reasons to justify an IR. But we have an ulterior motive as well. Much of our interest in beginning an IR resulted from discussions we had on the Trinity campus concerning scholarly communication. We wanted students – and eventually faculty – to develop some conception of the issues surrounding copyright, fair use, licensing, and alternative publishing models. A student thesis digital archive supplies a place for them to publish their theses to a wide audience and participate in the larger scholarly world without going through the typical journal route. We encourage them to use a Creative Commons license for their work, which allows them to retain their intellectual ownership rights while making educational use of their materials easier for others. The students are thus exposed to a new model for publishing academic work and will hopefully be in a better position to understand these issues when they enter their careers, many in professions where intellectual ownership is important. Faculty, seeing what we hope is a successful publishing model for students, will perhaps become more aware of the options that they have with the scholarly information that they produce.

The seeds of these earlier discussions developed into contracting with Berkeley Electronic Press to form an institutional repository for theses called Digital Commons® CDMT (after the first letters of the original colleges’ names). This publisher had first interested us in its efforts to produce electronic scholarly journals in a timely and cost-effective manner. Using the software it
had developed for e-journal article submission, review, and publishing, this publisher has created an easy-to-use platform for institutional repositories. We were impressed by the success of the University of California/California Digital Library eScholarship repository, which uses this platform, and decided that we could use the same software successfully for our student work repository. Our vendor hosts the server and manages the software for the IR service. (ProQuest is now marketing and supporting this product.) Each participating library has its own Digital Commons site, which it manages independently; the vendor then provides an umbrella site, allowing users to search the various collections as one entity.

The repositories at the colleges developed more slowly than expected during their first few months. This type of project was quite new for many of us, and as we elaborate below, selling the concept to other campus constituents has not been easy at times. Most of the IR projects to which we could point were concentrating on faculty publications. Many faculty had no concept of a student digital archive and its possible applications. We should also not minimize the difficulties of committing to this sort of project at a liberal arts college library. None of our institutions has the level of staffing to throw at this effort that a research library might be able to do. For example, Trinity University has ten librarians serving its 2,600 students, so finding the time to work through the various policy and procedural issues and begin any sort of production can be difficult.

Of course, most of our libraries have been working on other digital projects. At Trinity University, while we are pursuing the development of the thesis project, we are simultaneously ramping up projects involving the digital archiving of two photographic collections (early university history and theater productions). Our staff selected CONTENTdm as the platform to support these projects; that software is superior for the archiving and manipulation of image files compared to the textually focused Digital Commons platform. Does that mean we have two institutional repositories? We need to consider whether our users will find multiple systems confusing. Our current solution uses a library Web page to point to our various digital collections by topic, rather than concentrating on the platform. In this we parallel the way libraries usually point to their bibliographic databases by topic and not by software vendor. But as digital collections grow, we will no doubt need to develop other discovery tools for our users that will effectively integrate multiple collections with our other resources. Federated search tools will likely be one way to attempt this integration.

Content Issues

The schools participating in our IR consortium, now called the Liberal Arts Scholarly Repository, all are interested in obtaining and archiving student theses and dissertations. However, as work progresses on each IR, the individual institutions have added other types of materials to their IRs, including:

- electronic journals;
- faculty working papers and presentations;
- commencement and convocation addresses;
- departmental planning documents and annual reports;
- university policies and governance documents (e.g., the faculty handbook);
- videotaped senior student project presentations; and
- education students’ curriculum units.

One might look at the variety of approaches taken by the different libraries in this group and wonder if there are any overall collection guidelines. The quick answer is no; each library determines what to add to its own IR. Our approach recognizes the novelty of the IR and the hesitation over how to use it. Experience has shown us that when describing the repository to faculty, administrators, and even some library staff, we often get blank looks for the first few times. This seems to be one electronic product that does not immediately grab people’s attention. Rather, it takes a particularly relevant example for the average person to see the usefulness of the IR for his or her purposes. Consequently, it is entirely reasonable to begin the IR with a small, specific collection focus – in our case, student theses – and then look to find appropriate problems for which the repository can provide a good solution. Just getting started and learning what can be done with this new tool is a worthwhile activity.

That said, running an IR can be expensive in both funds and staff time. Library staff must obviously do some planning in order to determine which software suits the materials that will be archived, how the collections will be displayed to users, and so on. The following questions and issues should be considered fairly early in the development of a project:

1. Purpose: for instance, regarding our focus on student work, do we want to create a comprehensive record of student efforts at our schools (including both the exemplary and the less successful work) or is the IR meant to be a showcase of the best student work, useful for promotional and admissions purposes? Within our small consortium, librarians have taken quite different positions on this.

2. Scope: what materials should be collected to meet the purpose of the IR? Although we at Trinity started with the idea of including just honors theses, we have considered other student work as well, such as non-honor level theses and senior group projects in computer science. Currently we are happy to add any of these student projects, but there are concerns about lower quality work that might be considered more ephemeral and not worthy of long-term preservation. Some of our consortium colleagues feel quite strongly that the student IR should be selective in nature. Regarding theses, however, we have been questioned about whether students should be required to include their theses in the IR. At Trinity, some students have been advised by their faculty...
advisors to decline inclusion (much to the library staff’s consternation), because the thesis work is considered potential journal article material, and some publishers currently may construe dissemination via a digital repository as “prior publication.”

3. Ownership and copyright: it may be reasonably clear that students own their intellectual work and maintain the copyright for anything they produce in tangible form, but most academic institutions also assert some rights to their own ability to use that work for institutional purposes; historically, requiring a copy of a thesis to be added to the library collection is one example. Additional copyright issues must also be considered carefully with student work; students are often less savvy about the rights of image creators, for example, and may include unauthorized reproductions of images in their papers. When only one professor reads such a document, there is not any serious issue; when the student’s work is digitally available through an IR and indexed in Google and Yahoo, there are greater concerns for liability.

4. Formats: archives that have always dealt with a variety of forms of expression – printed works, manuscripts, photos, art, video, and even more unusual “realia” – have an additional host of issues when digital preservation and dissemination are concerned. We now must consider what formats we will accept, based largely on what we think we can support in the future. Currently the majority of our student theses and other materials are submitted in Microsoft Word format and converted into PDF files by our Digital Commons software. We expect that most people who wish to read the theses will be able to handle these formats. But what happens when a contribution to the IR includes an unusual digital media type that may not exist in a few more years? It is our responsibility to see that we can convert that file into a format that can be viewed by our users some years in the future.

5. Length of preservation commitment: early in the process of soliciting materials from campus authors, the library staff must have an agreement with these authors about how long the library will maintain access to those materials. More than likely, the library will be committing to long-term, ongoing maintenance of those materials. We intend to provide perpetual access to the digital theses we archive, just as we have the older printed ones. But the implications of long-term storage and maintenance, possible future file conversions or providing access to legacy software, and the like, should be discussed before agreements are made to include specific types of materials in the IR.

6. Metadata: just like other library collections, those materials we are adding to our IR collections require access points to be discoverable by other users. Some IR software provides better mechanisms for supporting metadata creation and use than others. We discuss the support for metadata creation and searching in the Digital Commons software below. Metadata creation can be an expensive part of these projects, and the level that will be desirable for a particular collection should be evaluated at the beginning of a project. There are clear differences, for example, in what is required for an image collection without accompanying text versus a typical textual document that can be full-text searched.

7. Migration: institutional repository software and best practices are still in considerable flux at this time, so we must make the assumption that we could migrate to a different IR platform in the future as needs and products change. Library staff should be considering possible migration plans for their IR materials and metadata right from the start of any new project. Contracts with IR software vendors should include stipulations that a library’s information can be exported in appropriate formats for later import into a new system. Our consortium, for example, has had annual discussions of the suitability of our present platform in both functionality and costs and has continued to keep aware of competing products.

8. Authority for decisions: the political and economic factors that are attached to many collections considered for the IR indicate the need to establish firmly the location of authority for decisions. In our situation with students who did not want to submit their work to our Digital Commons, we have had to work out policy issues with both an academic administrator and a few faculty, recognizing both our limited role in determining campus policy on thesis requirements and our leadership role in determining the way submitted items will be handled once in the library.

Features of the Digital Commons

Much of our workflow for processing materials is based on the particular IR software we use. Below is a discussion of some of the workflows we have developed around the Digital Commons as well as some interesting features of this software.

Types of Publications

To accommodate different types of publications, the Digital Commons offers different publication templates. For monographic documents like working papers, research papers and theses, there is a monographic publication template. For continuing resources like e-journals, there is a peer-reviewed journal publication template. A template for selected works of individual researchers is also available. All documents are published into series and have settings for appearance, workflow, and policies; we have organized our publication series along departmental lines. For example, the Computer Science Honors Theses series currently houses ten documents; the series itself has links to Web sites and policies and is searchable (Fig. 1).

Types of Workflows

The Digital Commons offers different levels of access for different types of users (administrators, editors, reviewers, authors) in order to facilitate workflow...
management. For example, in a peer-reviewed journal, an author submits an article through a form on the journal’s Web page, the editor is notified via e-mail, and using the system, matches the new submission to a reviewer. The system continues the automated process by sending the reviewer a request via e-mail, tracks reviewer responses, follows up automatically on deadlines, and notifies the editor about the status of the review process. The editor makes a decision on the article and the system notifies the author and the reviewer. Although we have an electronic journal on our site, we have not yet utilized the peer-review software. It does take time to learn, and as of yet, faculty have not been interested in learning the system; thus far, once an issue is finished, the library has loaded it into the Digital Commons on their behalf. We are hoping to have more success using the peer-review software with student publications.

Another component of the Digital Commons we have not utilized is author self-submission. A submit link can be created on a site so that an author can submit her document and metadata directly. Although fairly simple to use, it still does take some time for students and faculty to learn and this may prove too much of a barrier for adding content to the Digital Commons. A secondary issue with utilizing the self-submission process on our campus is that, even if a system is easy to use, if participants are required to process their own materials, it may seem like the library is taking away a service instead of providing one.

**Staffing**

There are two staff members who work directly with the administrative module of Digital Commons at Trinity: the head of cataloging who creates publication series and organizes the site, and a cataloging assistant who loads in documents and metadata. Loading documents is very easy; a thesis is used to illustrate our workflow. Currently, the student e-mails the thesis to the cataloging assistant, and, if one has not been created previously, the head of cataloging creates a departmental community (e.g., History Department) and a publication series for housing documents (History Honors Theses); the head of cataloging assigns series permissions to the cataloging assistant. The cataloging assistant then loads the thesis and any associated files along with metadata into the series. Once a thesis is loaded, the system automatically generates an e-mail notification to the author. In addition, the system automatically generates an e-mail to the author describing any download statistics for that particular thesis on a monthly basis.

**Metadata**

The Digital Commons allows for simple metadata input: title, author, abstract, keywords, and date of submission. Initially, we planned to use a controlled vocabulary for subject terms but for the time being have decided against it. One reason for our hesitation is that the software does not effectively support controlled lists of subject terms nor provide for any “see from” or “see also” references. While it is possible to create a list of subject terms for data entry, it is only of use to the person assigning the metadata and loading the paper; the searcher is not given the option of a drop-down menu of possible vocabulary to choose in order to specify a search, but rather must directly input the subject term. The catalogers would be taking valuable time to assign good subject headings/series/genre and other collocation devices for little value to patrons. Also, because of the wide range of documents housed in the
Digital Commons, it is difficult to find a controlled vocabulary that suits the content of all the documents. Perhaps because one of the focuses of the Digital Commons is author self-submission, the metadata structure must be simple. However, since we have not utilized the self-submission process at Trinity but instead have involved the catalogers with inputting metadata, the system seems to lack some robustness. As professionals, we know that there is much more we could be doing to organize and create access to these collections via metadata. What we would prefer to see in a system is the ability to easily provide specialized lists of vocabulary terms to accommodate different disciplines that patrons can browse. We would also like to have the ability to control this list with “see” and “see from” references.

Format Issues
The Digital Commons works best with textual documents: theses, reports, journals, etc. It is possible to upload associated non-text files (images, data sets, video clips) so they may be presented alongside the main document. In the past two years, there has been a marked increase in student theses with accompanying files: data sets, images, video clips, etc. Previous to the Digital Commons, these associated files were not terribly accessible as the CD-ROM or some such container was simply attached to the binding of the thesis. Presenting these theses in the Digital Commons has made these key files much more available (Fig. 2). However, the Digital Commons is not geared towards managing image collections as are CONTENTdm and other databases. For example, the Digital Commons is unable to create thumbnails or to provide multiple resolutions of images; in addition, the display of images is limited to one view instead of offering users multiple ways to arrange search results of images.

Interface Characteristics
The Digital Commons supports full-text and descriptive metadata (authors, titles, keywords, abstracts) searching of textual documents. However, it is worth noting that only descriptive metadata are searchable with non-text documents. For example, if you conduct a search for words within a PowerPoint presentation, the results are null. All metadata and full text can be searched simultaneously in a general search box. More specific searching can be done in the advanced search box; however, it is not possible to browse by subject as the subject terms must be input directly by the user. The system can limit a search to a single publication series or the entire collection. There are RSS Feeds and e-mail notification options for reports of newly published content.

The Role of the Consortium
At first glance, forming a consortium to work on institutional repositories might not seem necessary; most IRs by their nature focus on the output of a particular organization. However, we have found several advantages to belonging to a consortium for our thesis project. First, vendors who are hosting repositories see an opportunity to leverage their costs by encouraging groups of colleges to sign up for their services, so they are willing to offer discounts to consortium purchasers. Second, the schools in our group are all high quality liberal arts colleges and are proud of the work that their
students do. Yet schools like ours generally are less well known to the public than large research universities. When we create a consortium IR that showcases the research that undergraduates can do at our type of institution, we promote one of the key benefits that we offer prospective students. Although the schools in our consortium may be considered competitors for the same top students, we also profit from the added exposure of being seen in the company of other high caliber schools. Third, this collaboration with other recognized institutions offers useful promotional value internal to each school. For instance, those faculty and administrators who may not catch on to the value of an IR initially can be influenced to consider it more seriously when they can see that other highly regarded colleges are also pursuing similar projects.

Other potential advantages to our consortial repository project have not materialized. Many of us expected that the library staff at the various schools might share the development of procedures, collection policies, and metadata creation. However, these have turned out to be areas that have been dealt with individually by each library at this point. We have a vendor that can provide support when we encounter problems using the software, and the vendor has set up a discussion list for our consortium to share some of the solutions among us. Since our software platform is relatively easy to use and was almost completely a turnkey system, we probably have not experienced the same need for support as those with other platforms. For instance, those who are providing their own server and open source software without vendor assistance may find the pooling of library expertise to be substantially more useful than we have up to this point.

Marketing the Institutional Repository

As we mentioned above, the IR is not a concept that immediately sells itself to campus users. Faculty in particular appear to take longer to understand what the library is trying to do and how their work might be supported by us. Our experience, which has doubtless been duplicated elsewhere, is that a substantial amount of library staff time must be allocated to marketing the IR, both to acquire content and to justify the expense of the project.

Our student thesis project has a substantial advantage over faculty-oriented archives: students understand the benefits of the online repository much faster than faculty and staff. When we first approached some faculty about the concept, they were lukewarm on the idea at best, but they did suggest some students who were writing good theses. When we contacted the students, they immediately liked the idea of their work being online, discoverable in Google, and potentially visible to future graduate school admissions officers, employers, and even relatives.

Today’s students are tomorrow’s alumni. We had a number of urban studies theses completed many years ago when we had a graduate program in that discipline. Due to their in-depth focus on San Antonio, they are among our most requested theses. We contacted the alumni authors and asked for permission to scan their theses and add them to the Digital Commons. Those alumni we were able to locate and contact were almost universally excited about the opportunity. Combined with the generally positive comments we hear from our recent graduates, these reactions lead us to believe that the inclusion of student work into this type of digital institutional showcase will help create a stronger tie between them and their alma mater.

When we talked to faculty about our project, we tried to emphasize how the Digital Commons would showcase the fine work that their students were doing. Surprisingly, many of them saw no utility in making this work available to others through the Web. We have responded to this reaction in part by pointing to the usage of the documents in our IR. Although we do not have a large selection of documents at this point, we have had a large number of downloads (the current 110 contributions have been downloaded nearly 7,000 times in the past ten months). Obviously, interest in this material does exist.

The capturing of student work is our primary goal, but we also hoped that we could add faculty materials to the IR. This has been much less successful, with a few contributions of research reports and one e-journal. To be more convinced of its usefulness, faculty need to see that the management and promotion of their intellectual work are enhanced or made easier by using the IR. Also, it is quite likely that faculty are not as enamored of putting their materials in a repository with their institution’s brand as they might be if they could showcase their work under their own name. We see this when faculty create their own Web pages; they typically emphasize their own personal accomplishments rather than subsuming them as activities of their academic departments and universities. (Students and administrators are more likely to view an institution’s brand on their work as a sign of its value, we believe.) Our vendor has done some basic work in creating templates for personal researchers, though we have yet to develop any such pages with our faculty.

Our experiences with marketing our IR projects have led us to believe that it is very important to start small, choosing projects that have usefulness to our constituents, can be completed successfully, and can serve as seeds for additional projects. Much of our additional content has come because we saw an archiving problem, determined how our resources could solve it, and then sold the idea to the information creator based in part on our earlier success. This method allows greater tailoring of our marketing message, offering specific benefits on well-defined issues rather than promoting a mass storage project. By moving deliberately from one defined project to another, we are also able to avoid promising more than our staff resources can provide.

For example, the Education Department was interested in an online portfolio program to house their lesson plans and planned a project independently of the library; however, the Education Department soon realized that it was in its best interest to allow wider
access to these materials than just within the department and to let someone else organize and load them into a database. The library stepped in as a natural resource for this project.

Impact of the Digital Commons and Future Projects

It is hard to judge the impact of our IR at such an early period; however, a regular review of usage statistics can tell part of the story. The Digital Commons makes this easy as the system records and arranges statistics in an effective way. Both hits on document cover pages and full-text downloads are recorded, and usage reports are generated at the site, publication series, and document level. As already mentioned, e-mails citing usage statistics are automatically generated to authors on a monthly basis, and students in particular seem to be very excited to see hits on their work. Our statistics show that our publication series of lesson plans created by students in the Education Department are receiving the most traffic, followed closely by honors theses from the Computer Science Department.

Recognition by the campus that the library’s IR can serve an important function is growing, but our Digital Commons is by no means a well-known resource. Most of our campus administrators support our move towards digital archiving of important university materials, but that has not meant any increase in staffing to work on large projects. We anticipate that some of our current cataloging staff time that is spent on processing physical materials will be reallocated to these sorts of digital projects in the coming years.

We are currently investigating other projects to add to the Digital Commons: a student-run literary review; an annual publication of student essays; the student newspaper; and administrative documents like faculty and staff handbooks. These projects may be better suited to utilizing the peer-review software that is an important component of this system. Ultimately, we expect to work on projects not yet envisioned, as our campus constituents increasingly realize their need for safe, digital storage and future access to many of their intellectual creations, and their acceptance of the library’s role in providing this service.

Suggested Readings

Institutional Repositories


Trinity University, Digital Commons@Trinity, http://digitalcommons.trinity.edu (accessed March 7, 2006).

Copyright

Creative Commons, http://creativecommons.org (accessed March 6, 2006).


Faculty Participation


E-portfolios


Copyright of Serials Review is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.