Marjorie Hassen

Content Recruitment and Development: A Proactive Approach to Building an Institutional Repository

EDUCAUSE

11 October 2006

Good afternoon. Before I begin I need to make a confession. I am actually here this afternoon under false pretense. I have only recently become involved with the Penn Library's institutional repository and while I am the sole presenter today, this paper is really a result of a group effort. It was created in large part by several of my Penn colleagues who have worked more closely with the repository than I. These include Richard Griscom. Mary Steiner, and Sandra Kerbel. That said I will do my best to provide a coherent picture of our strategy for recruiting participants for Penn's institutional repository, which is known as *Scholarly*Commons@Penn

To put our repository effort in context, I'd like to start by giving a quick overview of the academic environment at the University of Pennsylvania. Penn is a mid-sized, private university with twenty-four thousand students and four thousand faculty members. The campus is served by a library system consisting of a central library and fourteen branches that employ a staff of four hundred FTE.

The university itself consists of twelve schools, including Engineering, Education, Medical and Law Schools, and the Wharton School of Business. The campus is decentralized, meaning that each of these schools runs somewhat independently: they have separate budgets and in many cases, they maintain their own school-centered support services in areas such as information technology. Penn employs responsibility-centered budgeting, which means that the schools pay for services provided by other campus units. (The concept of responsibility-centered budgeting extends into areas you might not expect: for example, units bill other units for use of meeting spaces in their buildings.)

You can get a sense of how this process works at Penn through this chart. An algorithm determines how much each school contributes to a central allocated cost pool, and the provost negotiates the distribution of the pool to the library and other central services.

The library is funded through a fee assessed to each school. For some deans, this fee is seen as a "library tax." However the fee is viewed, though, this budget practice puts the library in the position of being perceived as a service provider to a set of paying customers, and it's only natural for the schools to have expectations that might not exist in other budgetary environments. This environment at Penn had a direct impact on our approach to building an institutional repository. It would have been impossible to have all twelve schools agree on a single approach, so instead we looked for partner schools.

Before reviewing the history of the repository, here's a snapshot of where it now stands:

Seventeen units are represented, including four institutes and centers, and the units are associated with four of the twelve schools. The content covers a broad range of document categories including:

- faculty scholarship (including journal articles, conference papers, working papers, technical reports)

- master's theses and capstone projects

- selected undergraduate scholarship

- lecture series

and

- unit-sponsored conference content.

As of the end of September, there were a total of 1,250 items in the repository. (I should mention that the actual content is more extensive. There are also 5,500 Penn doctoral dissertations supplied by Proquest as a part of our contract with them for their Digital Commons software.)

Based on statistics provided by the Commons software, our weekly use includes an average of 9,800 cover page hits, 1,500 full-text downloads, and 3,400 searches. For 69 % percent of the downloads, Google was the means of discovery, and more than 80% of use originates from off campus.

Here is one example of a popular entry – this for a paper in our Department of Electrical and Systems Engineering papers series that has received over 3,200 downloads since it was deposited in August 2005.

So, how did we get started? The institutional repository at Penn had its origins in a proposal made by the head of the Engineering Library, in December 2002. She urged the Library administration to consider hosting a repository, and she offered to take responsibility for exploring the possibilities and pursuing its development.

During 2003, while the library administration was considering how a repository might be funded and supported in Penn's environment, a group of library staff, known as the Repository Management Team—or "Repoman"—kept abreast of trends in institutional repositories and completed a system analysis and comparison. When a funding opportunity presented itself, the library was ready to move ahead with a pilot project.

The partner who came forward was the School of Engineering and Applied Science, whose dean was interested in highlighting faculty scholarship and raising the profile of the school. The dean agreed to fund the software costs for one year, and the Library, in turn, agreed to fund the labor costs of setting up and developing the repository.

We chose the Digital Commons system, marketed by Proquest and based on the Berkeley Electronic Press (or "bepress") platform. Two factors influenced our decision: first, the results of a system comparison of DSpace and bepress completed in 2003 at our request by a library-school intern from Drexel University, and second, our need to have a highly visible product up and running quickly to meet the expectations of the School of Engineering. Choosing the Digital Commons product removed the necessity of providing staff and resources to implement and maintain a server. Leasing a service made it possible for us to move forward relatively quickly.

Of course, there's more to maintaining a repository than software and technical support. There's also content recruitment, verification of permissions, creating metadata, and processing the actual submissions. As part of our decision to start small, we staffed the project initially with existing Library personnel and did not commit resources to create positions specifically for the repository. We wanted to wait until we were sure the project was securely on its feet before considering the prospect of hiring new staff.

Since the School of Engineering was providing funding for the software, it was clear that the Library's wish list of outcomes would be secondary. We would need, first and foremost, to meet the expectations of the

school. Through conversations with the dean and faculty, we learned that they were hoping to accomplish several goals:

- Increase the use of School of Engineering research

- Provide a showcase to highlight and publicize the school's research

- Offer individual branding for departments in the school

and

- Allow individual departments to determine what type of content would be appropriate for their areas of the repository

From the outset, the dean was an enthusiastic supporter of the pilot project. In fact, he distributed email to Engineering faculty promoting the new initiative, which of course lent weight to its importance in their eyes.

Our approach to building the repository was to provide a proactive service model. With the support of the school's dean and department chairs, we were granted permission to begin harvesting materials identified through existing websites, such as the school's listing of faculty publications, and working through faculty members' personal websites.

Instead of waiting for faculty to deposit, we gathered citations, then checked to determine whether an individual publisher's policy would allow posting, with the assistance of SHERPA (**Securing a Hybrid Environment for Research Preservation and Access)** that tracks publisher's copyright policies. Only after we had gone through these steps did we contact the individual faculty member about our plans to include their research in the repository.

In addition to searching resumes and lists of research on Penn sites, we set up SDI alerts in journal article databases—such as Science Citation Index, Engineering Index, and INSPEC—so that we could keep abreast of new publications for deposit without having to make individual requests for lists of new publications from over one hundred faculty.

By taking this approach, we made participation simpler for the faculty. Instead of asking them to take the time out of their schedules to determine which of their papers could be deposited and then locate the appropriate versions and upload them, we did all of this preliminary work and contacted them to report that certain papers would soon be available on the repository. In situations where we needed an alternative version for posting, we'd contact them to request the files for specific articles.

Because the dean had already emphasized the importance of the repository in his email to faculty and in his meeting with department chairs, faculty were cooperative. Having the weight of the support of the dean, and in turn, the department chairs, was essential to the success of the pilot.

We also took an activist approach to securing permissions. If the information provided by the publisher on the SHERPA site was unclear or incomplete, we initiated discussions with the appropriate professional societies and publishers.

For example, we contacted the Materials Research Society and expressed our desire to archive Society papers. We explained that we would provide attribution where material was originally published and as a result, they agreed to allow us to include publisher PDFs and updated their SHERPA entry accordingly.

In cases where a society's policy was restrictive, we worked with faculty and library colleagues to continue to press the issue, and at least one faculty member wrote a publisher requesting permission to post his publications in the repository.

Consequently, we did not work in isolation from the academy on issues of scholarly communication, but rather, when appropriate, informed faculty and senior administrators about the issues. As part of our marketing, we developed flyers about how to get started with depositing documents and noted which societies and publishers would allow posting. We have also maintained a separate website, called "Winning Independence." This site covers issues surrounding scholarly communication.

Rather than market the repository as a product to change opinions on the culture of scholarly communication—which is the function of the Winning Independence site—we have kept the focus on what the repository can do to make Penn research more visible and accessible.

At the end of 2004, with the conclusion of the pilot project, we decided to extend the scope of the repository beyond the School of Engineering to the campus at large. Again, we felt it important to start small and to focus on units that were of a manageable size and showed enthusiasm for the project, building on our pilot experience.

Realizing that we wanted departments and faculty to buy-in to the repository voluntarily, we focused on identifying incentives and promotional tools to motivate participation rather than trying to secure a departmental or even institutional mandates that would have made participation compulsory. While mandates may generate content, we were concerned about the risk of inciting resentment and antipathy toward the project.

When we officially opened up the repository to units outside the School of Engineering in February 2005, we blitzed the campus (and beyond) with publicity:

- We designed a brochure about the repository and the pilot with the School of Engineering and Applied Science. This brochure was so well received that the Dean of the School sent it to his colleagues at other Schools, thus emphasizing the value of a professionally-produced publication.

- A press release was prepared and distributed to national library publications.

- Our library director sent a personal letter, along with the brochure, to campus deans and department chairs. Some departments then asked for more brochures to distribute to their entire department.

and

- The campus newspaper, the *Daily Pennsylvanian,* published an article on the repository

Our most effective publicity, however, was a grassroots effort involving the Library's bibliographers, the subject specialists who have the closest relationships with academic units on campus. We asked these bibliographers to serve as missionaries for the project by going out with brochures in hand to help spread the word and to gauge interest among their constituencies.

Before sending bibliographers on the road, however, we made sure they were fully up to speed on the repository project—the concepts behind it, how to use the public interface, how the submission process works, and so forth. We did this through various meetings with staff and library department heads.

The bibliographers then began their work. Some met with constituencies without the repository project leader present. In retrospect, those meetings would have been more productive had the Project Leader attended, given her better understanding of the project and its scope.

Also, to assign this kind of work to the bibliographers, the library administration needs to send a clear message that this type of work is a priority for the library and is to be considered part of the bibliographers' responsibilities. On the other hand, the Project Leader needs to be sensitive to the bibliographers' workloads and provide support for them through other staffing to take care of permissions, posting, and harvesting content.

As we considered our prospects, it was clear that we did not have the staff or the experience to take on the entire campus, so we targeted individual units that would be visible and enthusiastic contributors. We wanted to establish a representative cross section of disciplines in the repository, and since we already had a foot in the door with the hard sciences, we identified several social science programs, including the Graduate School of Education and the Penn Institute for Urban Research. Our attempts to make inroads into the humanities departments were less successful. To date, we have only one humanities series established: a series of lectures in the field of religious studies.

We also chose not to start with what is arguably the most well-known school on campus, the Wharton School, because we felt they might reject the repository out of hand as perceived competition with their Knowledge@Wharton site. This is a general interest site that reports on recent Wharton research. We have made initial inroads into Wharton by working with an individual faculty member who was keen on getting his papers into the repository. And now, because of the success of an undergraduate research journal begun by the School of Arts and Sciences, which is now available in the depository, through word of mouth we have laid the groundwork for inclusion of content from a similar journal sponsored by the Wharton School.

Expanding the project obviously has required a greater commitment of personnel resources. Again, we have proceeded with some caution and committed resources only as the project has demanded it. In fact, just a few weeks ago, we filled the first position devoted solely to the repository.

The staff currently assigned to the repository include:

- A project leader

- An operations manager

- A policy and planning working group

and

- Subject liaisons

The Project Leader coordinates the administration and planning for the repository, chairs the Repository Working Group, supervises the work of the operations manager, and is the main conduit to the subject liaisons. So far, the responsibilities of Project Leader have been appended to an existing librarian's position—first the head of the Engineering Library and now the head of the Music Library—but it could eventually evolve into a full-time position.

The Operations Manager is our new position, the first assigned specifically to the repository, although it remains a one-year appointment at this point. Again, we are approaching the dedication of resources to the project with some caution, and our level of staffing is evolving as we observe patterns of interest and growth.

The Operations Manager deals with the set up and maintenance of new communities and series, provides under-the-hood administration, reviews metadata and document integrity, and processes submissions for units that don't have an internal coordinator or a library liaison.

The Institutional Repository Working Group is a deliberative body charged with setting the strategic direction of the repository and setting policies. It is composed of the

- The Project Leader

- The Operations Manager - who we have just hired

- A staff member from the Library's Information Technology and Digital Development Department

- A digital content specialist, currently a staff member from our Center for Electronic Text and Image

- A librarian from our Information Processing Department

- A subject bibliographer

and

- A liaison from the library administration

As I mentioned earlier, the library subject liaisons are key to our outreach to the various academic units on campus. They inform their constituencies about the repository and become involved when one of their units expresses interest in participation. The Project Leader makes sure leads are followed up, either personally or in conjunction with the subject liaison. Through work with the subject specialists, the Project Leader ensures continuity and good customer service to the units.

After the pilot year with the School of Engineering had been completed, the Library decided to assume full responsibility for funding the repository. The responsibility centered budgetary model Penn employs can

work against the Library's ability to initiate projects that require financial support when schools see only the dollar amount they are asked to contribute. Since we did not want money to be a deterrent to participation, we decided to bear the brunt of the cost and funded the project through the Library Director's Special Initiative Fund.

An area where adjustments had to be made when we extended the scope of the repository was content recruitment. For the pilot, we were working with a school whose dean had thrown his full support behind the project, and as a result, faculty felt, if not enthusiastic, then at least compelled to contribute content.

As we thought about extending the scope of the repository to include other departments and schools in the university, we realized that buy-in and eager participation by faculty wouldn't necessarily be a given.

Based on our successful experience with Engineering, we decided that the path to increased participation was through securing the backing of schools and units rather than through approaching faculty individually.

In our presentations to unit administrators, we pitched the repository as a vehicle for publicity—a showcase for the research of their schools and departments. After securing the support of a unit, we then relied on its administrators to provide the backing to encourage faculty to contribute.

We have been assertive in identifying prospective communities and providing support for getting content into the repository, but we have been careful to give the units complete control over deciding what type of scholarship is appropriate for their section of the repository. Our only policing of content is to verify that we have permission to post a particular submission.

We have tried to make it easy for faculty to deposit by following the model of the University of Glasgow, which offers authors the option of either self-deposit through the repository interface or mediated deposit by email.

This model, however, assumes the authors will take the initiative, which is not necessarily the case, even when there is encouragement from the administration. Instead of relying solely on faculty deposit, we decided to extend the activist approach that had been so successful with the School of Engineering.

- We locate CVs and publications lists by searching personal and departmental websites,

and

- We set up notifications for new publications appearing by faculty so that they can be reviewed for eligibility and subsequent deposit

Once we have identified an author's publications, we harvest eligible content ourselves whenever possible. When an article turns up through one of the alerts we have set up, we check the publisher's policy, and if the article can be deposited as is, we move ahead.

When we can not find full text of a publication, or when a search of the publisher's policy tells us that we can post only a pre- or post-print, we approach individual faculty with specific requests for particular papers. Depending on the unit, the email might come from library repository personnel directly, or we might draft a note and have it come from administrator in the unit to give it added weight. We don't hear back from

everyone in response to these requests, but we do hear back from some and that usually leads to the deposit of the publications.

We have found that for faculty, one incentive in particular has helped our cause: the receipt of personalized monthly activity reports generated by the Digital Commons software, which provide them with feedback on how much their publications are being used. This has encouraged some authors to submit—or have us submit—more publications on the site, so we create a positive feedback loop.

Clearly, faculty members are intent on getting published in the places that are most highly regarded in their field. They want to place articles in the best journals, and they want to sign book contracts with prominent publishers. Given this environment, we assume the question that comes to mind when they first hear about the repository is, "What's in it for me?" and that's a question we have to be able to answer convincingly.

For this reason, pitching the repository as an alternative venue for publication tends to carry no weight, since most faculty members—especially those on a tenure track—have no real interest in being published outside the traditional venues in their field. Arguing that the repository is helping to transform the culture of scholarly communication also carries little weight. Faculty may agree with the argument in principle, but generally, only established, tenured faculty are in a position to feel comfortable challenging the status quo. The scholarly communications argument might appeal to a faculty member's conscience, but because it really offers nothing of direct benefit to the faculty member, in the end, it is weak.

So, in our pitch to departments and individual faculty, we have focused on the repository as a vehicle for increased access and visibility for scholarship rather than as an opportunity for publication or as a vehicle for improving the state of scholarly communication. We point out that colleagues, students, and beyond will be able to discover research through the most popular search engines, and because the library has a long-term commitment to the repository, access to the research will be ongoing. And we emphasize that the repository can serve as a single, convenient location for them to retrieve copies of their research and to refer their research to others.

And rather than pursuing heavy-handed tactics such as mandatory deposits, we have been working under the assumption that exposure to the repository—through word of mouth, monthly reports, and campus publicity—will increase participation. Despite our attempts at publicity, however, the repository doesn't become "real" until a faculty member receives email that an article has been deposited and then that it has been viewed 72 times in the past month. And often, when faculty realize how much the depository is contributing to the discovery of their research by others, they are eager to make further deposits

It is worth noting that in some cases, we have seen faculty members first become interested in the repository as the result of the deposit of papers they had coauthored with colleagues from a participating unit. While taking a faculty member by surprise may not be the way to go in all cases, it will, at minimum, provide the opportunity for discussion.

The repository's growth has been organic and at times happenstance. For example, the creation of the College Undergraduate Research Electronic Journal (CUREJ) on ScholarlyCommons@Penn was the result of a visit of the dean of the College of Arts and Sciences to speak at a meeting of public services librarians. He mentioned in passing that the college was planning to publish an undergraduate research journal, and we seized the opportunity to host this journal using the repository interface.

And months later, after the journal was live on the Web, the dean of the college mentioned the site to the director of the Research Scholars program in the Wharton School, and as a result of their lunchtime conversation, the Wharton School has now asked us to set up a similar journal, the Wharton Research Scholars Journal.

What lies ahead? As we look to the future, we are considering a number of possible new initiatives.

Some faculty have expressed interest in having citations for their entire list of publications available on ScholarlyCommons, not just citations for the publications we've deposited. Our working group has been discussing how to act on this request. On the one hand, it runs counter to the purpose of the repository. By adding a large number of citations without content, we would dilute the repository and perhaps reduce the incentive to deposit actual content. One option would be to build a separate database to fill this need and retain the focus of the repository as it stands. The University of Wisconsin is working with this approach by building a separate bibliography database, with selected records added to the repository with full-text.

We would like to work with others to develop a submission process that allows authors whose content is eligible for deposit in multiple repositories—for example, institutional, disciplinary, and funding-agency repositories—to deposit a document only once, and then have the content fed automatically into the various repositories.

Now that ScholarlyCommons has been opened up to the entire campus for over a year, it's time for us to get feedback from our depositors, the participating units, and librarian liaisons. We have discussed a number of possible methods: a survey, pulling together a focus group, one-on-one meetings with unit administrators, and these initiatives will be forthcoming soon.

The "personal researcher page" has been popular with some of our more prolific faculty authors who are interested in disseminating their research more widely. Here is one example– the personal page for Ralph Rosen, Professor in Classical Studies.   Templates we use for these pages, though, are fairly spare in design, and revisions to the pages must be done by repository administrators rather than the faculty themselves. We're interested in finding ways to make the researcher pages more attractive and to place more control over the design and content in the hands of the individual. We hope that this would lead faculty to see a personal repository page as more of a personal homepage and then inspire them to upload more publications.

Like most ventures involving new services based on a new technology, our experience with ScholarlyCommons@Penn has required us to remain flexible and change our plans and approaches as we work with faculty and staff on building the repository and deal with unexpected challenges. Based on Penn's experience, we offer this advice to colleagues who are planning their own implementation:

- Understand your campus environment. Look at where the decision making occurs for the units that are likely participants. Is it the dean, the department chair, the faculty? Focus your energy on gaining the support of the people who make the decisions and set the direction for the unit. It may vary from department to department and from discipline to discipline. Also, identify people on campus who might be strong supporters of such a project and who could serve as advocates for the repository in conversations with faculty and administrators.

- Be proactive in securing participation and content. Instead of assuming that if you build it they will come, expect the opposite. Publicize, ask for time at departmental meetings to promote the

initiative, and seek out prospective participants.

- Start small. Accept that resources are limited and that building a solid foundation with a small group of contributors will be more productive in the long run than starting with dozens of units covered haphazardly.

- Be willing to increase support as the repository grows. As more units are added, there will be a corresponding increase in the workload. Be prepared either to increase staffing and resources or to shift some of the workload off onto unit staff.

- Let the repository grow naturally. Target units that will be enthusiastic contributors likely to spread interest through your campus, and try to accommodate any unit that makes an effort to approach you about participation.

And Finally, be patient.

Thank you