> URLs in this document have been updated. Links enclosed in **{curly brackets}** have been changed. If a replacement link was located, the new URL was added and the link is active; if a new site could not be identified, the broken link was removed.

Viewpoints

Let's Get it Started!

George S. Porter
Engineering Librarian
Sherman Fairchild Library of Engineering & Applied Science
California Institute of Technology
Pasadena, California
george@library.caltech.edu

Most academic librarians have reached the conclusion that institutional repositories (IR) are a good idea. SPARC has hosted national meetings and workshops on the subject, as well as hosting the SPARC Institutional Repositories Discussion List (SPARC-IR). On the academic side of the house, Stevan Harnad is into the second decade of his jeremiad on the subject of self-archiving. A number of platforms have been created to support institutional repositories, including Dienst, e-prints.org, D-Space, the Open Knowledge Project, and FEDORA, in addition to commercial hosting services from BE Press, BioMed Central and ProQuest. If librarians and academicians agree on the desirability of institutional repositories, and software platforms and services are available to make repositories technically feasible, one is left to ponder a few questions. Why are there so few institutional repositories up and running? Why are the existing institutional repositories generally not well filled with the intellectual output of their respective institutions?

The difficulties involved in establishing an IR are not economic or technological in nature. Rather, they are sociological and strategic, with organizational inertia being a large obstacle to this early phase of implementation. Here are a few suggestions for focusing initial efforts to get an IR off the ground.

Focus on Late-Career Faculty

In order to create a campus culture where faculty routinely deposit material in the IR, it is necessary to get some individual faculty to become early adopters. Of course, this is an occurrence of the classic "chicken or the egg" problem. The surest way to convince faculty that an IR is worth their while is to demonstrate that the content in the IR is highly visible and read in the rest of the world. Without content, such a demonstration is undeniably difficult.

There are several strategies circulating regarding how best to recruit faculty to contribute their content into a nascent IR. Targeting tech-savvy young faculty is an oft-espoused idea. Broadcast appeals to begin filling an empty repository have been tried. My response to both of these strategies is to recall *Raiders of the Lost Ark.* As Indiana Jones and Sallah realized on the outskirts of Tanis, "They're digging in the wrong place!"

What's wrong with the tech-savvy Gen-Xers or Millennials approach? It has never been clear to me how this would spread throughout an institution to arrive at full participation in a time span measured in units smaller than decades. The youngest faculty are not necessarily the primary opinion leaders on a campus, in part because they have not been around

long enough to have made sufficient connections to exert that kind of informal influence on a large scale. They are generally under tremendous pressure to establish their labs and research groups so that they can begin producing the research results which will advance them along the tenure track.

Why not pursue the "Build it and they will come" avenue? Broadcast requests to begin filling an empty IR are not likely to generate the groundswell of faculty support required to integrate IR participation into the DNA of the campus culture. The material doesn't spawn within the digital environment. Someone has to identify, describe (metadata), and deposit the material. The investment of time and effort, whether their own or that of their staff, is being requested as an article of faith.

The way to win over scientists and engineers is with data. Data will only be generated by having content in the IR. Pump priming or a pilot project, using library staff to get the initial critical mass of material into the IR, is necessary to demonstrate that content distributed through an IR will be read. With data in hand, a case could be argued on more than mere supposition.

At Caltech, the most enthusiastic early adopter was a late career professor who wanted to document his oeuvre. He compiled a complete bibliography and pursued clearances from the publishers. He approached the library about digitization and participation in the IR. The project has been so successful, he has moved on to documenting the career of one of his early mentors at Caltech.

I pursued permission to digitize a technical report series from an emeritus professor for almost four years. After he finally relented, the library had a better feel for the agreement language which would satisfy the concerns of the faculty. Additionally, we were able to produce better image PDFs than we had been at the beginning of the process. His acquiescence led to permission for the library to digitize two technical report series. He has also become an advocate for the IR as an enduring repository for individual faculty.

Perhaps a practical solution to the "chicken or the egg" issue is to try to actively engage late-career faculty. This may seem counter-intuitive to many librarians, since so many new library initiatives are pitched to and eagerly adopted by the newest generation of academicians. The senior faculty may view the proposition as a capstone/culmination/collected works project for their career. They are also more likely to have a large enough publishing portfolio -- many articles, technical reports or working papers, books -- that a critical mass of content could be collected from publishers with enlightened policies regarding author rights retention. Such an approach postpones the day when permissions must be actively sought to make further progress.

Focusing on content which is easy to incorporate into an IR (low-hanging fruit)

While recruiting content is a major obstacle to the success of an IR, another critical component is the ability to distribute the material which an institution has produced and re-acquired. SHERPA/ROMEO has been considered the preeminent compilation of journal publishers' policies with respect to preprint and postprint distribution, but it suffers a number of shortcomings as a tool to assist with an IR at an operational level. The mergers and acquisitions in the publishing industry muddy the waters, since the copyright holder today may not be the publisher of record at the time of release nor of the imprint today. Ascertaining a publisher's policy with respect to the version of a paper which an author contributes to an archive is a tedious and expensive process [{John Ober, SPARC/ACRL Forum, ALA Midwinter 2006}]. An effort is underway [{ASEE ELD Scholarly Communication Committee}] to address more specifically the policies and conditions of science and engineering journal publishers.

A limited number of publishers/publications permit the use of the as-published PDF to be harvested and uploaded to an IR. The peer-reviewed material which an institution has produced and published within these enlightened journals is the low hanging fruit. Begin harvesting the intellectual heritage of your institution from the material which presents the least difficulties with respect to publisher permissions. From an authoritative standpoint, it is hard to argue that the publisher's PDF is in some way inferior to a preprint or whatever ill-defined concept of post-print one cares to contemplate. The imprimatur of the journal, valued by both the publisher and the author, is (usually) quite clear in such

cases. There cannot be any question of version control or pagination errors.

Other rich sources of readily available content include the institution's gray literature: technical report series, working paper collections, theses, and dissertations. The wrangling for permission for these resources is an in-house exercise providing librarians practice in negotiating for certain rights in a friendly environment. The clarity gained from negotiating with one's campus colleagues may well be of benefit when it comes time to pursue permissions from publishers who do not already cede the necessary rights back to the authors.

Following Willie Sutton's sage wisdom

Willie Sutton observed that he robbed banks because "...that's where the money is." To get an institutional repository up and running, librarians need to go where the content is, preferably content which will entail little effort to clear rights and permissions for distribution. Where are the mother lodes of content, lying near the surface?

- Theses and dissertations
- Technical reports
- Prolific authors
- Conspicuously cooperative publishers

By tapping into these rich areas requiring minimal negotiation, a repository can be established and stocked with a substantial amount of content. Start with the easy stuff. The hard stuff will still be waiting, after you've had a chance to establish how worthwhile and practical an institutional repository effort is in your locale.