Strategies for Developing an Institutional Repository: A Case Study of ScholarWorks@UMass Amherst

Yuan Li, Syracuse University
Marilyn S Billings, University of Massachusetts - Amherst

Available at: https://works.bepress.com/yuan_li2/7/
Strategies for Developing an Institutional Repository: A Case Study of ScholarWorks@UMass Amherst

發展機構典藏之策略：以麻省大學安城校區機構典藏實例為例

Yuan Li
Digital Repository Resident Librarian,
University of Massachusetts Amherst, U.S.A.
E-mail: yuanli80@gmail.com

Marilyn Billings
Scholarly Communication and Special Initiatives Librarian,
University of Massachusetts Amherst, U.S.A.
Email: mbillings@library.umass.edu

Keywords (關鍵詞): Digital Repository (數位典藏); Institutional Repository (機構典藏); Development Strategies (發展策略); Content Recruitment (內容收集); Scholarly Communications (學術傳播); Academic Libraries (大學圖書館)

【Abstract】
ScholarWorks@UMass Amherst is an institutional repository that collects, manages, and disseminates intellectual output of UMass Amherst faculty, researchers, and students. In less than five years, it has become one of the top five Digital Commons repositories with more than 23,000 items and over half-a-million full-text downloads. ScholarWorks content recruitment strategies are examined as a case study for the development of an institutional repository. The authors aim to provide best practices for developing a digital repository in academic settings and inspire colleagues to explore and adapt new strategies.

【摘要】
ScholarWorks@UMass Amherst 是麻省大學安城校區的機構典藏，主要用於收集、管理和發佈本校教師、研究人員和學生的學術成果。自建立至今不到五年的時間，已超過 23,000 件的學術成果收錄量和超過五十萬份的文件下載數，躋身為 Digital Commons 排名前五的機構典藏。本文以 ScholarWorks 的發展策略為實例來探究機構典藏的發展，旨在提供一實施範例作為同行參考，並期望能以此啟發業內同行開發與新策略採用。
Introduction

An institutional repository, also referred to as a digital repository or digital commons, is “a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members” (Lynch, 2003). “An increasing number of academic institutions in the United States and abroad are developing institutional repositories (IRs) in a bid to retain the intellectual output of their scholars and support open access trends in scholarly communication” (Palmer, Teffeau, & Newton, 2008). The best approaches to develop a digital repository and the strategies for success have been popular topics since the adoption of the IR concept and librarians working in scholarly communications have been striving to establish best practices for their development. A few successful strategies for IR development have been reported in the literature, such as collaborating with academic departments and faculty, involving subject librarians and liaisons, and expanding IR services. (Palmer et al, 2008; Gaffney, 2008)

Unfortunately, institutional variance is seldom considered when presenting IR development strategies. This paper takes ScholarWorks as a case study to examine the content recruitment strategies that have been used for its successful development. The local organizational context and the development environment are also provided.

ScholarWorks, the institutional repository at University of Massachusetts (UMass) Amherst, was established in July 2006. ScholarWorks has grown to be one of the top five Digital Commons[1] repositories. This paper presents a successful IR development model for other institutions and suggests good strategies for IR practitioners and academic librarians.

Literature Review

At the beginning stage of IR development in the 1990s, many repository advocates believed in a philosophy of “if you build it, they will come”. This strategy has proved to be unsuccessful (Foster & Gibbons, 2005). Many IRs have experienced difficulties in content recruitment after their establishment. A survey in early 2006 conducted by the Association of Libraries (ARL) found that member institutions held an average of only 3,884 digital objects in their IRs (Quint, 2006). Since then, developing content recruitment strategies has become a top priority for all IR practitioners. As Bankier et al. (2009) explain, “one must understand that campus ‘awareness’ does not equal campus participation.” Several papers have been published which studied possible marketing and content recruitment strategies. The literature has focused on two general threads: collaborating with faculty and working with other library staff. Foster et al. (2005) report that an understanding of faculty work and a commitment to build the IR to meet faculty needs could improve content recruitment. To take full advantage of IR services, IR administrators also explored new and expanded roles for reference librarians and liaisons in bringing new services and new forms of support to faculty members. Responsibilities to academic departments and knowledge of disciplinary research needs and scholarly communication patterns position reference librarians and liaisons to inform IR growth. Additionally, reference librarians and liaisons should be prepared with repository knowledge such as features, benefits, and services in order to play vital roles in helping to educate and recruit content from faculty (Bell, Foster, & Gibbons, 2005; Jenkins, Breakstone, & Hixson, 2005; Malenfant, 2010; Phillips, Carr, & Teal, 2005). From a broader angle, Gaffney (2008) pointed out the importance of involving the campus community, especially faculty, and also emphasized benefiting from library staff expertise, including cataloging, systems, and reference.

In addition to collaborating with different groups of people in developing an IR, other new strategies include providing new services. “Faculty want clerical
and consultative services. These services could include scanning, mediated deposits, copyright advising and rights-checking” (Bankier, Foster, & Wiley, 2009). Repository- based digital publishing services are an additional benefit for the campus community. By providing scholars assistance with starting a born-digital journal or digitizing back issues, new interests and engagement will be fostered.

Despite the aforementioned efforts, the development of an IR is still in the early phase of establishing guideline principles and best practices as an evolving part of the profession of librarianship (Palmer, Teffeau, & Newton, 2008). The case of ScholarWorks@ UMass Amherst is presented to examine the strategies that have been used to successfully develop the UMass Amherst institutional repository, providing the field with new strategies that have been successful at a research institution.

**Background**

The University of Massachusetts Amherst, the flagship campus of the University of Massachusetts system, has 1,174 full-time instructional faculty and an enrollment of nearly 27,000 undergraduate and graduate students. With 88 major subject areas leading to bachelor's degrees, 73 masters and 51 doctoral programs, UMass Amherst offers a vast array of educational opportunities for full and part-time students, both on and off-campus. Interdisciplinary majors, online learning, the Five College (MA) exchange, internships and international study add to the opportunities. UMass Amherst is ranked as one of the nation's top public very high research activity universities by the Carnegie Foundation. The campus attracts over $140 million in externally sponsored research each year, demonstrating its contribution to Massachusetts’ position as a national technological and economic leader. Research funding supports the creation of new knowledge and its translation into the technical innovations and scholarly works that create opportunity for students, faculty and the public.

The UMass Amherst Libraries consist of the W.E.B. Du Bois Library, the Sciences and Engineering Library, the Image Collection Library and the Music Reserve Lab. The Libraries hold over 3.5 million volumes and 43,906 journal subscriptions. Over 40 librarians, 80 paraprofessionals, and hundreds of student assistants serve both remote and campus users. The libraries are a member of the Association of Research Libraries and participate in many consortia to broaden our collections and extend or enhance our services to UMass Amherst students, faculty, and staff. Community members are invited to visit the two library buildings, borrow materials, and enjoy many of the extensive services.

**Strategies for Developing ScholarWorks**

When the University Libraries first envisioned the creation of an IR in 2005, it was recognized that campus engagement and education would be critical to its success. IR oversight was institutionalized by the creation of the Scholarly Communication Librarian position in July 2006. Begun as a pilot project, ScholarWorks was officially launched at a high profile campus event in March 2007 with keynote speaker David Shulenberger, Vice President for Academic Affairs at the National Association of State Universities and Land Grant Colleges. Ongoing campus education efforts included several scholarly communication colloquia and presentations at Faculty Senate meetings, and librarian education by attendance at national conferences and workshops, such as the ACRL / SPARC Scholarly Communication Institute and SPARC IR workshops. Throughout this period, the Faculty Senate Research Library Council was an active advisory body for ScholarWorks development efforts. From its beginning, ScholarWorks has also been strongly supported by the Office of the Provost, the Graduate School, and the Office of Research. What follows is an in-depth look at these partnerships and what has made them successful strategies for IR development.
Building Partnerships across Campus

Graduate School

The first partnership conversation that the University Libraries conducted in regard to the IR was with the Graduate School. The Graduate School and the University Libraries had already established a successful working relationship for the collection and management of print PhD dissertations. The Graduate School had recently adopted the electronic submission process provided by ProQuest for their Dissertations and Theses database so they were intrigued by the University Libraries selection of Digital Commons IR software, at that time licensed from ProQuest. By 2007 the Scholarly Communication Office was engaged in a pilot project where graduate students deposited their masters theses into ScholarWorks using the same workflow that doctoral students used to deposit their dissertations. This dramatically lessened the number of paper theses the Graduate School had to handle, streamlined their operations, and gave the students the advantage of adding supplemental materials in multiple formats. We have worked together to provide education to the Faculty Senate Graduate Council about open access to electronic theses and dissertations (ETDs) and created successful workshops for graduate students to discuss open access and campus access, embargoes, fair use, author rights, and similar topics. Another advantage of collecting dissertations and theses through ScholarWorks was the dramatic increase in usage they have received. Our usage statistic show that the Masters theses in ScholarWorks have been downloaded an average of 284 times in the past four years and that the dissertations have been downloaded an average of 144 times in the past year. By comparison, our dissertations in ProQuest have only been downloaded an average of seven times.

Office of Research

The second partnership developed quickly between the University Libraries and the Office of Research, evolving out of jointly sponsored colloquia and mutual interest in new scholarly communication models. In FY07, the Vice Provost of Research immediately adopted ScholarWorks by creating his author Selected Works page and by creating a new digital journal in his research area. Based on this experience, he was so convinced of the value-add that ScholarWorks provides for showcasing University research and scholarship that he mandated that faculty who received “Research Leadership in Action” (RLA) grants from the Office of Research deposit the results of that research in ScholarWorks. Generally the grant monies were used to sponsor a conference on campus which seeded the capture of several scholarly UMass Amherst conferences since FY08 (see Figure 1).

The University Press, part of the Office of Research, joined in pilot phase ScholarWorks efforts in 2006, contributing six full text monographs written by UMass Amherst authors (see Figure 2). This early effort was facilitated by the University Press Director’s attendance at the Scholarly Communication Institute mentioned above and his active participation on the Faculty Senate Research Library Council. More details about the Press Library collaboration are described in the digital publishing services section of the article.
2nd Annual Conference and Opportunity Fair, 2009

Clean Energy Connections is a career and business development information and networking event. It is a forum for the individuals and organizations that will accelerate the growth of our clean energy economy.

- ACCESS resources to start your clean energy career or build your business.
- MEET entrepreneurs and community groups mobilizing around clean energy.
- TOUR exhibit hall displays about green technologies, services, practices and products.
- LEARN about clean energy trends. What’s hot and what’s hype.

Browse the contents of 2nd Annual Conference and Opportunity Fair, 2009:
Conference Program & Notebook
2009 Conference Proceedings

Figure 1  RLA grant sponsored conference example
(available at http://scholarworks.umass.edu/clean_energy/)
Office of Outreach

The University of Massachusetts Amherst is a public land-grant institution, meaning that a critical part of the university’s mission is to provide agricultural and technical education and to promote higher education especially to rural and underserved parts of the population. The Faculty Senate Outreach Council monitors and reviews campus-level activities and develops policy recommendations relating to the definition, development, promotion, delivery and evaluation of public service and outreach activities. Under the leadership of the Office of Outreach and the Outreach Council, the University chose to apply to the Carnegie Foundation for the elective Curricular Engagement and Outreach and Partnerships classification in 2008. This classification describes the collaboration between institutions of higher education and their larger communities (local, regional/state, national, global) for the mutually beneficial exchange of knowledge and resources in a context of partnership and reciprocity. This effort led to a new strategic partnership with the Office of Outreach. While the application was under review, the ScholarWorks team created a new “Communication Engagement” section

Figure 2 University of Massachusetts Press book publishing
(available at http://scholarworks.umass.edu/umpress_tne/)
to provide a place to showcase the outreach and engaged scholarship that had been highlighted in the application. This has proven to be a large “hook” for pulling in interest from faculty. This partnership proved so successful that it resulted in a presentation at the National Outreach Scholarship Conference held Sept 2009 at the University of Georgia, followed by the submission of a journal article to The Journal of Higher Education Outreach and Engagement, the premier peer-reviewed journal in this field (see Figure 3).

Figure 3  Community engagement project example
(available at http://scholarworks.umass.edu/cranberry/)

**Commonwealth Honors College**

In 2009, the University Libraries began conversations with the Dean of the Commonwealth Honors College to discuss the deposit of their undergraduate honor capstone papers and undergraduate research conference. We agreed that, based on the success with the masters theses project, we should begin capturing the top undergraduate research capstone projects and showcase their conference using the events functionality in ScholarWorks mentioned previously.

By early 2008, it was clear that the small ScholarWorks Team[4] needed additional full time professional help so when the Graduate School, Office of Research, and Office of Outreach were approached to co-fund a new two year term entry level Digital Repository Resident Librarian
Collaborating within the Library

As it is frequently reported in the literature, collaboration in the library is strategically planned to ensure success in promoting IR initiatives. The roles of reference librarians, liaisons, and other regular staff need to be reshaped and expanded; their potential values need to be realized; and their efforts need to be involved and incorporated. Collaborating with reference librarians, liaisons and other regular staff has proved one of the most successful strategies for ScholarWorks. In summer 2010, the ScholarWorks team participated in the University Libraries Summer Project[5] in which the ScholarWorks team brought together reference librarians, liaisons, and staff from Acquisitions, Cataloging, E-reserves, ILL, and the Business Office to create personal researcher pages (Selected Works pages) on ScholarWorks. This project entailed generating pages for UMass Amherst faculty, and obtaining permission from faculty members to build these pages, collecting publications and other research output, and populating their online personal researcher page to showcase their research. The project included four phases: 1) creating Selected Works pages (performed by staff volunteers); 2) requesting permission from faculty and obtaining related materials (performed by liaisons); 3) exporting citations from targeted databases (performed by staff volunteers); and 4) populating faculty personal researcher pages by importing citation data (performed by staff volunteers). Based on the collaboration with librarians and staff, the project was very successful with 1,183 personal researcher pages[6] created by the end of summer (see Figure 4).

Figure 4  Faculty personal researcher page example
(available at http://works.bepress.com/michael_ash/)
**Harvesting from Existing Repositories and Databases**

Building partnerships and creating collaborations are good long term strategies. However, these strategies are time-consuming and it is challenging to demonstrate immediate results. Therefore, the ScholarWorks team developed another strategy to populate the IR: harvesting from existing repositories and databases. The team collected contents from identified subject repositories and databases in order to quickly build the collections.

“Subject repositories which collect and provide access to the literature of a single subject or a set of related subjects are often cited as highly successful scholarly communication initiatives, especially in relation to institutional repositories” (Adamick & Reznik-Zellen, 2010). Subject repositories such as PubMed, arXiv, and RePEc continue to grow in size and are widely used and recognized within their respective disciplines. Therefore, subject repositories are very rich sources from which institutional repositories can recruit content, including pre-prints or post-prints contributed by an institution’s faculty. The ScholarWorks team identified three subject repositories, arXiv, RePEc, and CiteSeerx, as the targeted sources from which to harvest content. The criteria used for identifying them included the number of items contributed by UMass Amherst faculty, the best version of the item deposited into the repository (pre-print, post-print, or publisher version), and the quality of metadata provided. The whole workflow includes the steps of harvesting, permission requests from faculty, uploading, and publishing (see Figure 5).

![Figure 5](image-url)  
*The workflow of harvesting from subject repositories*
In addition to harvesting from subject repositories, harvesting from research databases is a highly efficient way to improve the speed of content recruitment. Research databases are primary resources that libraries subscribe to for their communities. These databases generally contain very rich subject descriptions in the form of keywords, subject classification terms, abstracts and full-text of the indexed contents. Harvesting full-text content from databases can be very complicated because of involved copyright issues and licensing agreements. However, metadata, such as author(s), article title, journal title, publication date, keywords, and abstracts, are often publicly accessible and can be harvested into an IR. The ScholarWorks team developed a new workflow to batch export citation records in XML from targeted databases, such as Web of Science and JSTOR, by using web services, [7] to transform the XML to the IR format, and batch import the records into the IR. Data entry is the most time-consuming part of the traditional workflow of ingesting content. Every record, including all metadata and full-text content, has to be entered and published manually into the IR. According to our records, the average speed of data entry is 20 records per hour per person. The new workflow improved the efficiency of content ingesting by 25 times, or 500 records per hour per person (See Figure 6).

Figure 6  The workflow of harvesting from databases
Providing New Services

The Association of Research Libraries’ Office of Scholarly Communications published a report entitled “Research Library Publishing Services: New Options for University Publishing” in 2008, which showed that 65% of responding libraries offer or plan to offer some form of publishing support using editorial management and publishing systems including Open Journal Systems (OJS), DPubs, Digital Commons and homegrown platforms (Hahn, 2008). Considering the library as a digital publishing services provider via IRs is another excellent strategy to promote IRs regardless of the marketing or content recruitment advantages. IR-based digital publishing services and personal researcher page creation are two of most attractive services provided by ScholarWorks. These services have gained strong interest from a wide variety of constituents including faculty, graduate students, and other affiliated scholars and researchers, and, as demonstrated above, they have also fostered partnerships and campus-wide engagement. Library publishing services are focused in three areas: 1) e-journal publishing, 2) conference proceedings publishing, and 3) monograph publishing.

E-journal Publishing Services

Scholarly journals have been playing an important role in scholarly communication for hundreds of years. The internet has had a profound effect on traditional scholarly publishing with e-journal publishing having a major impact on this traditional venue. However, the high cost of e-journal publishing has delayed a complete transition from print to digital, particularly for many high quality scholarly journals from small societies and research groups. IRs, like ScholarWorks, provide faculty editors with the resources to transition their journals from print to digital or to create a born-digital journal[8] (see Figures 7 and 8).
Conference Proceedings Publishing Services

Conferences have long been a mechanism for faculty to share their research in a more timely manner than formal publishing. Providing the ability to capture the structure and content of scholarly conferences, this is another type of publishing service that libraries can contribute to the campus community. The UMass Amherst Libraries collaborate with the Office of Research and University Conference Services on initiatives of publishing and archiving conference proceedings on ScholarWorks (see Figure 9). The collaboration with University Conference Services keeps us informed of all conferences hosted by the university well in advance, which enables us to offer ScholarWorks services to conference planners. University Conference Services also promotes this service by displaying ScholarWorks brochures in their office, including them in their campus promotion packages, and by putting ScholarWorks on their Facebook page. One of the unanticipated gains from this service is a dramatic increased interest by faculty in using ScholarWorks for capturing other scholarship and research data. Examples of this are included in the Academic Department Services section.
Monograph Publishing Services

One of the digital publishing services that ScholarWorks provides is hosting and dissemination support for monographs and supplemental materials. In addition to the collaborative work with the University Press that was described above, the IR can provide a permanent location for additional or supplemental materials that the author wants to disseminate. An excellent example of this type of collaboration is the team’s work with the chair of the Department of Languages, Literatures, and Cultures who was publishing a monograph with the Stanford University Press. Dr. Julie Hayes had extensive primary materials and footnotes that the Press could not include in her publication due to their publishing costs. Instead, the Press recommended that she create a website, a suggestion that she didn’t want to take the time to explore. She remembered the presentation about ScholarWorks at her academic department meeting and contacted the ScholarWorks Team. ScholarWorks provided a perfect solution for the author by creating a site with the book cover image (permission given by Stanford University Press), a brief book description, a link to the Stanford University Press website to facilitate the ordering of the book, as well as all the supplemental files (see Figure 10). The Stanford University Press website includes a link back to the ScholarWorks page where all the supplemental files are located and includes that persistent URL in its printed monograph.
Figure 10  Book supplemental files example, written by UMass Amherst author
(available at http://scholarworks.umass.edu/french_translators/)
**Academic Department Services**

With many different types of content in the IR, working with Academic Departments has proven to be another lucrative way to demonstrate to the faculty the value-add that ScholarWorks can provide. Not only can virtual communities specific to the department be created to showcase selected scholarship, such as dissertations and theses, but specific academic programs, such as the European Field Studies in Anthropology, can have their unique materials hosted in the Academic Department section (see Figure 11). In addition, the department may have print working papers or research papers series that they have hired a staff member or graduate student to manage. These are ready candidates for inclusion in this section, especially since these types of materials generally do not have copyright restrictions. The department community is a place to gather individual faculty Selected Works pages and create a faculty publications series to highlight works from those pages. As seen by this example, academic departments can deeply benefit from working with the ScholarWorks team to populate a meaningful showcase of their research and scholarly materials.

Figure 11  *Academic department services example* (available at http://scholarworks.umass.edu/anthro/ )
Conclusion

As stated in the introduction, it has been well recognized that content recruitment has been one of the major challenges faced by libraries and IR administrators to date. The experience that the ScholarWorks team has gained by exploring a wide variety of possible services using the features and functionality of the Digital Commons software platform have provided extensive “hooks” to use when recruiting content and participation by faculty and research groups. It has also been a critical component in being able to engage both the library staff and the campus communities in meaningful dialogue about new scholarly communication and publishing approaches. Projects such as the summer project 2010 gave librarians and staff the opportunity to actively participate in this new library endeavor. Services such as those articulated in this paper have enticed the faculty and the broader campus. Keeping our eyes open and ears peeled to really hear what the faculty and researchers need has been a key approach used to build those services. The ScholarWorks team encourages other libraries and IR administrators to build some of these services and partnerships in their ongoing efforts to populate their own IRs.

Acknowledgements

During the process of the survey and preparation of this paper, we received a lot of support from our colleagues and friends. Here we would like to thank our colleagues, Meghan Banach and Jessica Adamick from the W.E.B Du Bois Library at the University of Massachusetts Amherst and Dr. Marta Deyrup from Seton Hall University Library for their insightful comments on the paper.

Notes

[1] Digital Commons is a hosted repository platform provided by the Berkeley Electronic Press.

[2] Selected Works is a component of our open access digital repository and is a web-based individual faculty personal researcher page that highlights publications, performances, and other scholarly work. It is a software product that is licensed from the Berkeley Electronic Press.

[3] The Research Leadership in Action (RLA) program was an internal grant competition sponsored by the Vice Provost for Research for full-time UMass Amherst tenure-track and research faculty between 2005 and 2010. The program supported faculty interested in showcasing leadership in their field of research and scholarly activity.

[4] The ScholarWorks team included the Scholarly Communication librarian and part-time assistance from two Metadata librarians and one science librarian involved in creating a nanotechnology subject repository.

[5] UMass Amherst Libraries accepts two or three project proposals for Summer Projects every summer which allows librarians and staff members to develop new skills and expertise by voluntarily participating in a summer project for a few hours per week in lieu of their usual job responsibilities.


[7] A web service is a software system designed to support interoperable machine-to-machine interaction over a network.

[8] Refer to http://scholarworks.umass.edu/peer_review_list.html for a list of journals, including born-digital journals, and graduate student run journals.

References


