



Date : 27/06/2006

**Institutional Repositories: A Review of Content Recruitment Strategies**

**Timothy Mark and Kathleen Shearer**  
 Canadian Association of Research Libraries  
 Ottawa, Ontario  
 Canada

<b>Meeting:</b>	<b>155 Information Technology with National Libraries with Academic and Research Libraries and Knowledge Management (part 2)</b>
-----------------	--

<b>Simultaneous Interpretation:</b>	<a href="#">Yes</a>
-------------------------------------	---------------------

*WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL*

**20-24 August 2006, Seoul, Korea**

<http://www.ifla.org/IV/ifla72/index.htm>

***Abstract***

*The Institutional Repositories movement has attracted great attention from academic libraries worldwide over the past several years, often as a way of furthering the cause of Open Access to scholarship.*

*The appeal is immediate and apparent. An Institutional Repository is a way for every academic institution to ‘showcase’ its intellectual prowess through the systematic collection, organization, making accessible and preservation of its intellectual output. Moreover, the linking of individual repositories has the potential of developing a true knowledge network where the researcher can “harvest” relevant material from any number of repositories across the world. Such an initiative is in the finest traditions of international scholarly collaboration – and libraries would play a central part.*

*As part of its Institutional Repositories initiative the Canadian Association of Research Libraries (CARL) has conducted an international review of content recruitment strategies for populating Institutional Repositories.*

*This paper presents the results of the review and discusses some conclusions to be drawn. The findings will be generally applicable and should be of interest to any library that has, or is contemplating, implementing an Institutional Repository.*

## **I. Introduction**

One of the major barriers faced by scholars and researchers is the lack of access to the current literature in their field. In Canada, for example, from 1986 to 2002, serials expenditures in Canadian research libraries have increased by 233%, while the number of subscriptions has increased only 22% (CARL, 2003). As a result, CARL libraries are spending over three times as much on serials as they did in 1986. This is a worldwide problem. However the situation is even more critical in developing countries, where library budgets can be extremely small. As a consequence the teaching and research in some countries is being performed without the essential input and benefit of research being conducted internationally. The grossly uneven availability of information resources around the world is a matter of record and an issue that a growing number of initiatives seek to remedy, such as the open access movement.

Open access calls for the free availability of scholarly literature on the Internet. The open access movement has gained significant momentum over the past several years. In 2001, the Budapest Open Access Initiative, a strong statement of principal supporting open access was drafted. The initiative has been signed by a growing number of individuals and organizations from around the world. Since then, there has been numerous of activities in support of open access including conferences, discussion lists and open access policy statements. For a full discussion of the Open Access movement the reader is directed to the Open Access News, which is maintained by Peter Suber, a Policy Strategist with Public Knowledge and SPARC. [<http://www.earlham.edu/~peters/fos/fosblog.html>]

Institutional repositories (IRs) are proposed as one of the major strategies for achieving open access. Essentially, they collect and provide free access to the research output of a given institution. In the Canadian context, IRs can contribute to the 'repatriation' of some at least of the research output of Canadian scholars, much of which is published outside of Canada.

## **II. The CARL Institutional Repositories Project**

In 2003, the Canadian Association of Research Libraries (CARL) launched the CARL Institutional Repository Project. Jean-Pierre Coté, Directeur general, bibliothèques de l'Université de Montréal and head of the CARL IR Project presented the rationale for the project at an Open Access conference hosted by the CARL in 2003:

"It is the responsibility of each academic institution to preserve, organize and distribute the intellectual output of their faculty". Furthermore, institutional repositories are "a way that institutions can 'get back' some of the output from the researchers and accelerate the movement toward open sharing of knowledge."

The project, spearheaded by the Canadian Association of Research Libraries, is a Canadian initiative to implement institutional repositories at several Canadian research libraries.

[[http://www.carl-abrc.ca/projects/institutional\\_repositories/institutional\\_repositories-e.html](http://www.carl-abrc.ca/projects/institutional_repositories/institutional_repositories-e.html)]

At the time of writing this paper thirteen CARL members had IRs up and running and five more are in the pilot/planning stages. CARL's role is to monitor the development of individual projects, facilitate exchange of between institutions, and demonstrating the interoperability of IRs. The Association has also supported the ongoing development of the harvester and has formed a working group to contributed identify a metadata profile for Canadian repositories.

### **III. Content recruitment: the biggest challenge**

Given the vision and the potential, it is surprising and disappointing that Institutional Repository collections have generally grown more slowly than proponents had anticipated. The phenomenon is worldwide. Even with a variety of creative ideas and promotional activities, faculty uptake has been reluctant where voluntary compliance is needed. Certainly implementers have found that 'recruiting content' is the biggest challenge and frustration.

The literature cites a number of reasons why faculty participation rates are so low. At the most basic level, faculty members lack awareness of the existence of institutional repositories. Several surveys have found that many academic authors are not familiar with the concept of any institutional repositories on campus (Swan, 2004; Swan and Brown, 2005).

Copyright and intellectual property issues are also a concern for faculty. Despite the fact that most publishers allow authors to make their articles accessible via their university's IR, authors are concerned that they may be violating the copyright agreements they have signed with their publishers by depositing their papers into an IR.

As well authors also express concern that posting to an institutional repository will be considered prior publication.

Another barrier is the lack of perceived incentives by researchers for self-archiving their work. Self-archiving adds a small, but additional workload to already busy researchers, without offering any obvious rewards. They "just do not see the point"—all of the existing reward-mechanisms are based on the current publishing system, not self archiving (Pinfield, 2004).

Clearly these hurdles must be crossed if we are to succeed in populating the growing number of institutional repositories around the world.

#### **IV. Review of content recruitment strategies**

In the winter of 2006, CARL conducted a review of content recruitment strategies being undertaken internationally to support its members in populating their IRs. The review involved a scan of the current literature, private communication with IR implementers, and a public call for input on several of the listservs.

A range of content recruitment strategies was identified through the review and placed into six categories: general promotional activities, mediated depositing services, content harvesting, researcher bibliographies, usage information, and self-archiving policies.

##### **1. General promotional activities**

Most institutions begin their content recruitment activities through a variety of promotional activities on campus. Most commonly, such activities include passing out brochures, conducting presentations to faculty committees, publishing articles in the library or campus newsletters/newspapers, and formally launching the repository. “Academics have to hear about your institutional repository service many times, over a period of time, and from several sources (print, online, in person). A good rule of thumb is that someone needs to have been exposed to your service seven times before they are fully aware of your service. Be sure to outline explicitly the benefits of your service to academics.” (Barton, 2004)

Also typically communicated through these activities are the benefits of depositing in an institutional repository for faculty members. “As you begin to build a service, it is critical to communicate how the service benefits the university community – in other words, to do some marketing to advertise the service on campus.” (Barton, 2004)

A study conducted by University of Rochester found that “faculty members do not speak the same language as librarians. Moreover, the features of an IR that are most exciting to librarians, such as persistent URLs and metadata schemas, rarely register the same enthusiasm for faculty. The resulting “Tower of Babel” is a significant hindrance to the increased use of IRs by faculty. To address this, they have established a group of out-reach librarians specially-trained in the features, benefits, and mechanics of their IR. They have also developed a “crib sheet” for librarians of responses to faculty questions and concerns about the IR.

[<http://docushare.lib.rochester.edu/docushare/dsweb/Get/Document-0808/IR%20Crib%20sheet.pdf>]

These types of promotional activities are important because they raise awareness of the repository. However, Canadian implementers have found, that used alone, they do not populate repositories. And they must be supplemented with other more targeted content recruitment strategies. Others have found this to be the case as well:

“Within the first year of the project a University-wide event on open access and institutional repositories was held. The event was publicized by email and in the University Newsletter. Subsequent to the event each of the attendees was contacted individually, and follow-up meetings were arranged. It was hoped that this would be a good method of generating content. However,

although it helped to open a dialogue with academics, this did not always translate into content.” (Mackie, 2004)

## **2. Depositing services**

Self-archiving does not require a lot of time once an author is familiar with the process, however, an important barrier to self-archiving is “the perceived time required and possible technical difficulties”. (Swan, 2005). The review found a variety of services aimed at reducing the workload for authors when depositing content. The idea being that “the easier it is for academics or departments to add content, the more likely they are to do so.” (Barton, 2004) The most common of these types of services are copyright checking activities, metadata assignment, and library depositing of content.

One example is the services offered by Edinburgh University Library, for example. They provide a ‘Mediated Deposit Service’ to help authors deposit their work. The service allows authors to simply e-mail their content to the library and library staff will then deposit the material on their behalf. [<http://www.era.lib.ed.ac.uk/help/deposit-guide.jsp>]

Copyright checking is another service offered by many IRs. Publisher copyright agreements can be complex and authors need assistance in understanding their rights. This type service is offered by Cornell University, for example. The Cornell IR has developed a protocol for addressing the entire copyright and checking process. The service includes “everything from checking copyright permissions, negotiating with publishers, requesting final manuscript versions from faculty, and tracking all of this information in a database to eventually uploading the document with associated metadata”. [<http://www.library.cornell.edu/insidecul/200605/digitalcommons.html>] The SHERPA/ROMEO [[www.sherpa.ac.uk/romeo.php](http://www.sherpa.ac.uk/romeo.php)] list is an invaluable tool for institutional repositories that plan to offer copyright checking services. The list uses a simple color-coated system to identify the policies of publishers in regards to depositing pre- and post-prints into IRs.

## **3. Content harvesting**

A few institutions are adding content to their repository by collecting the material themselves. This can be a good way of initially seeding the repository. As part of a multi-pronged content recruitment strategy, the DAEDALUS Project at University of Glasgow, ‘mined’ faculty and departmental websites for content. They found that most authors who already had content on the web were “happy for (the IR) to establish which of their publications could be added to the repository.” (Mackie, 2004)

At Glasgow, they took this a step further by directly depositing articles from open access journals and other publishers that allow self-archiving. They identified the journals and publishers that permitted archiving in repositories, and then searched article indexes to establish which of their authors had published in these journals. After establishing who the target authors were, they contacted them and used an opt-out strategy. Authors were told that their articles would be added to repository unless they chose to opt-out. “So far no members of staff have ever got back to us and asked

us not to add their article, but it is unclear whether this is because they support the project or because they have no strong feelings about the issue.” (Mackie, 2004)

#### **4. Researcher bibliographies**

Researcher participation is key to populating institutional repositories. Even with services in place that assist the depositing process, author support is a prerequisite for the populating IRs over the long term. To attract researchers, some IRs have built researcher bibliographies on top of IR platform, as an alternative access point. One of the largest projects of this kind is the “Cream of Science” project in the Netherlands. [[www.creamofscience.org](http://www.creamofscience.org)] Cream of Science is part of the DARE initiative [[www.darenet.nl](http://www.darenet.nl)], a joint program by all the Dutch universities and the National Library of the Netherlands, the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organization for Scientific Research (NWO). For the project, all DARE partners selected ten of their most prominent scientists and made their complete publication list available through their IRs and the centralized harvesting service. A personal page was set up for each author containing basic information: photo, affiliation, research field(s), etc. as a links to the publications available through the repositories. In total, about 24,000 full text publications were made accessible in this way.

Others are doing similar things. The University of Rochester has developed “Researcher Pages”, personal webpages that resides within the University of Rochester IR. [<http://wiki.dspace.org/DspaceProjects>] The pages include information about each research, along with a photo, links to publications. They have developed this as a patch to the DSpace platform and the software is freely available to anyone for download. And, l’Université Libre de Bruxelles is planning to build the entire academic bibliography of University researchers into their institutional repository. [<http://bib3.ulb.ac.be/RDIB/DISpace/>]

#### **5. Usage/citation information**

Researchers want their papers to be read, and usually as widely as possible. One of the greatest benefits of the institutional repository is that it provides open access to the content within. It is not surprising, then, that many IRs are providing usage information for the articles contained in the IR as a way of attracting authors.

The ability to track download statistics is being built into many of the IR software packages. A good example is the University of California eScholarship Repository. The repository, which uses the Digital Commons software, has built in functionalities for monitoring the usage of content. It tracks how many times each paper has been downloaded. The repository also keeps a running list of top ten downloads of all time and a ‘Paper of the Day’. Moreover, it tracks the number of total downloads for the whole repository, and number of total weekly downloads.

[<http://repositories.cdlib.org/escholarship/>]

#### **6. University policies**

A few universities have implemented ‘self-archiving’ or ‘open access’ policies that support the use of IRs on campus [<http://www.eprints.org/openaccess/policysignup/>].

The University of Kansas University Council passed a resolution on access to scholarly information in March 2005. Among other things, the resolution “endorses the contributions of the University of Kansas and its faculty to (national and international efforts to shape a more diverse and sustainable system of scholarly communication), including the KU ScholarWorks repository, a digital archive that can provide access and long-term preservation for the scholarly works of KU faculty and staff”.

[[http://www.provost.ku.edu/policy/scholarly\\_information/scholarly\\_resolution.htm](http://www.provost.ku.edu/policy/scholarly_information/scholarly_resolution.htm)]

The policy encourages but does not require the deposit of research publications in the university IR.

The Queensland University of Technology (QUT) has had a mandatory policy in place since January 2004. The policy states “Material which represents the total publicly available research and scholarly output of the University is to be located in the University's digital or " E print " repository, subject to the exclusions noted.” [http://www.mopp.qut.edu.au/F/F\_01\_03.html] Although the compulsory aspect of the policy is downplayed it has been helpful in populating the repository because it adds an extra legitimacy and authority when contacting authors for their papers.

Mandatory policies can be a highly effective and sustainable content recruitment strategies. A 2005 survey of authors conducted by the Joint Information Systems Committee and the Open Society Institute found that the vast majority of authors would self-archive if obliged to do so by their employer or funding body. (Swan and Brown, 2005) That being said, these policies are not always easy to implement. Both Queensland University of Technology and University of Kansas were able to implement their policies because they had an administrative support in the university that promoted the policy amongst colleagues and faculty.

## **V. Conclusions**

Because Institutional Repositories are still comparatively new, the review did not uncover a lot of direct evidence of the long-term efficacy of particular content recruitment strategies. However, a number of preliminary conclusions can be drawn. For one thing, populating an institutional repository involves a multi-pronged content recruitment strategy. Promoting the repository on campus is important, because it raises awareness of the existence of the repository.

One good way of promoting an IR is through faculty recommendations. In this way, early adopter, or demonstrator communities, can act as effective marketing agents for the repository if their initial experience is a positive one. Also, “One reason faculty have not rushed to put their work into IRs, therefore, is that they do not recognize its benefits to them in their own terms.” (Foster and Gibbons, 2005). Faculty speak a different language than librarians. Librarians tend to talk about metadata and open source software, while faculty respond to concepts such as visibility and about impact. IR implementers will be more successful in promoting their IRs if they speak the language of the researcher. Promotional activities such as these are not sufficient in themselves and must be complemented with other types of support and incentives for depositing. The benefits of depositing in an IR must be clear and demonstrable to each faculty member. Implementers must be prepared to “sell the sizzle.”

Bring the workload down for faculty as much as possible and certainly initially IR staff should be prepared to deposit, assign metadata and check copyright on behalf of faculty. The time and work required for faculty to deposit should be as minimal as possible. There are a number of automated services being developed to assist in populating IRs. These include: automated metadata extraction, a service currently being assessed the JISC (Joint Information Systems Committee); or automated importing of metadata. For example, the University of Glasgow has developed a script to import the bibliographic details of publications from Reference Manager databases maintained at the departmental level for Research Assessment Exercise.

Researchers are interested in improving the visibility and impact of their work. If IRs can demonstrate that the papers in the repository are downloaded often, this will encourage them to deposit more content. Faculty should be aware that IRs are indexed by Google and other general search engines, as well as specialized search engines such as Scirus, Google scholar and OAIster. Publicizing download statistics, which many repositories are doing is also extremely important in demonstrating that IRs are providing wide access to content.

Be prepared to invest time and money: content recruitment needs adequate funding and adequate staffing, especially in the initial stages. The review did not uncover a lot of information about costs of strategies. One example provided is the costs of the Cream of Science project in the Netherlands. The project reported it cost about 10,000 Euros per scientist, which translates into an average cost of 50 Euros per document. Once processes and infrastructure become more standardized, costs may drop to 10 Euros per document. Even the lower figure represents a significant amount of funds for any IR. Adequate staffing levels are also a major issue for IRs. A large contributing factor to the slow growth in content in IRs is that they are run as demonstration projects, and do not have sufficient budgets for staff to spend time on content recruitment. In Canada, for example, most of the IRs are being run by ½ full time employee or less. This is definitely insufficient to both manage the repository, and recruit content. The success of IRs in the long term rests on them being phased into the regular operational budgets of libraries, and assigned to dedicated staff members. Moreover, content recruitment must be an ongoing activity. Short-term strategies work in the short term only. Thus, continually collecting new content means an ongoing and long-term commitment to content recruitment.

Certainly the most effective strategy for content recruitment is to implement an institutional policy requiring the archiving of research publications into IRs. Such a mandatory policy is infinitely preferable to voluntary compliance (provided that the library is prepared to take on the duties required) because of course it solves the riddle of successful content recruitment. A comparison of IR content policies in Australia in 2005 found that voluntary policies were not nearly as effective as self-archiving mandates. (Sale, 2006) This is consistent with the results of a recent assessment by the National Institutes of Health of their voluntary Public Access Policy. The policy “requests that investigators funded by the National Institutes of Health (NIH) submit an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication to the NIH National Library of Medicine's PubMed Central (PMC).” (NIH, 2005) The assessment of this policy found that only 3.8% of the research articles that are governed by this policy were added to the repository since the policy was implemented. (NIH, 2006).

There is no doubt that successful content recruitment for institutional repositories is a tough nut to crack. However we hope that we have demonstrated through these international examples that IR many stakeholders are working towards ways and means of accomplishing this. It is critical to do so because successful content recruitment is essential if the dream of a worldwide network of institutional repositories is to be realized.

## VI. References

Hughes, Carol Ann. Escholarship at the University of California: a case study in sustainable innovation for open access. *New Library World*, Volume 105, Number 1198/1199, 2004, pp. 118-124.

Mackie, Morag. 'Filling Institutional Repositories: Practical strategies from the DAEDALUS Project'. *Ariadne*, Issue 39, April 2004. Available at: <http://www.ariadne.ac.uk/issue39/mackie/>

National Institutes of Health, *Report on the NIB Public Access Policy*, January 2006. Available at: [http://publicaccess.nih.gov/Final\\_Report\\_20060201.pdf](http://publicaccess.nih.gov/Final_Report_20060201.pdf)

Pinfield, Stephen. Self-archiving publications. Published in Gorman, G.E. and Rowland, Fytton (ed.s). *International Yearbook of Library and Information Management 2004-2005: Scholarly publishing in an electronic era*. London: Facet, 2004, pp 118-145.

Sale, Arthur. Comparison of IR content policies in Australia. *First Monday*, volume 11, number 4 (April 2006), Available at: [http://firstmonday.org/issues/issue11\\_4/sale/index.html](http://firstmonday.org/issues/issue11_4/sale/index.html)

Swan, Alma and Sheridan Brown. *Open access self-archiving: An author study*. May 2005. Available at: [http://www.keyperspectives.co.uk/openaccessarchive/reports/Open%20Access%20II%20\(author%20survey%20on%20self%20archiving\)%202005.pdf](http://www.keyperspectives.co.uk/openaccessarchive/reports/Open%20Access%20II%20(author%20survey%20on%20self%20archiving)%202005.pdf)

Swan, Alma and Sheridan Brown. 'Authors and open access publishing'. *Learned Publishing*, Volume 17, Number 3, 1 July 2004, pp. 219-224(6) Available at: <http://www.ingentaconnect.com/content/alpsp/lp/2004/00000017/00000003/art00007>