April 2006

Back Talk -- Institutional Repositories: Wars and Dream Fields to Which Too Few Are Coming

Anthony (Tony) W. Ferguson

University of Hong Kong, ferguson@hkucc.hku.hk

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Ferguson, Anthony (Tony) W. (2006) "Back Talk -- Institutional Repositories: Wars and Dream Fields to Which Too Few Are Coming," *Against the Grain*: Vol. 18: Iss. 2, Article 14.

DOI: <https://doi.org/10.7771/2380-176X.4643>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

I Hear the Train A Comin' from page 84

choose to join the **LOCKSS Alliance**, a strategy group that discusses and helps shape **LOCKSS** policies and direction. **Alliance** membership currently ranges from \$1,080 to \$10,800 figures annually, tied to **Carnegie** classifications. **LOCKSS** expects these fees to drop as more members join the **Alliance**.

Note that before the **LOCKSS Box** can preserve a journal, the publisher must give permission for the **LOCKSS** system to collect and preserve the journal. Permission is granted through a Web page called a **LOCKSS** publisher manifest. A publisher manifest permits the **LOCKSS** software to collect, and preserve the content. The manifest allows the publisher to control, by content unit (e.g., volume), the timing of the content preservation. In other words, the ability for the library to preserve an electronic journal is tied to the timeliness with which the publisher updates the manifest. Once in the manifest, the material may be preserved by any **LOCKSS** library participant that subscribes to the publication.

The material preserved on the **LOCKSS Box** is meant to serve first and foremost as a dark archive. Content will be served to an institution's readers when a publisher's Website is not available under terms of original license agreement. To return to the Megaquake scenario in my introduction, imagine that a catastrophic event wipes out **The Berkeley Electronic Press**. Because our materials are cached in the **LOCKSS Box**, a subscribing library could simply redirect requests for our content to the local **LOCKSS** version until our own servers returned to operational status.

The **LOCKSS** model is predicated on a diffuse set of redundancies (presumably not simply because the acronym for My One Copy Keeps Stuff Safe isn't as cool as the one they chose). Whereas **LOCKSS** bestows upon its participating libraries the responsibility to maintain local backups, **Portico** has opted for a centralized solution. Formally launched in 2005 with support from **JSTOR**, the **Library of Congress**, the **Mellon Foundation**, and non-profit **Ithaka**, **Portico's** stated goal is to ensure enduring access to scholarly e-journals via a single, secure archive.

Content for the **Portico** archive comes directly from members of the scholarly publishing community who have agreed to contribute to the archiving service. In its first months of operation, **Portico** has secured the participation of close to a dozen publishers. Participating publishers deliver their content to **Portico**. The data are then normalized by **Portico** in a format based on the open standard **Journal Archiving and Interchange DTD** developed by the **National Library of Medicine**. This normalization facilitates both preservation and future migration. **Portico** has pledged to migrate publisher data to new archival formats that emerge over time, and to supply the migrated data back to the publishers upon request.

Once the content has been ingested and normalized, it sits in secure, redundant dark storage. It is periodically checked for decay, and select librarians at participating libraries are granted password-controlled access for verification purposes, but otherwise not accessed. **Portico** provides all libraries supporting the archive with campus-wide access to archived content only if specific trigger events occur. Examples include a publisher ceasing operations or discontinuing a title, a publisher no longer offering back issues, or a catastrophic and sustained failure of a publisher's delivery platform. In these instances, participating libraries will be granted IP access to the archived literature. Note that, like **LOCKSS**, the completeness of the **Portico** archive is dependent on the timely and accurate supply of data from the publishers.

The logistical aspects of a library's participation in **Portico** are minimal. Keeping an updated list of IP ranges on file at **Portico** is the primary activity. Annual participation fees are tied to a library's materials expenditures. At the low end, this translates to \$1,500 annually, and graduated along a 16 step scale to \$24,000 (although discounts apply to early adopters in 2006 and 2007). Publishers also pay an annual participation fee ranging from \$250 to \$75,000 depending on journal revenue.

The main difference between **LOCKSS** and **Portico** is the degree of centralization each brings to preservation. Other distinctions occur as well, of course. The focus of **Portico's** preservation effort is the intellectual content and not the look and feel of the journal. **LOCKSS** enables libraries to cache Web-delivered pages and images to capture both appearance and content. Whereas a trigger event within **LOCKSS** will unleash access only to the subset of **LOCKSS** library participants who subscribed to the affected journal, within the **Portico** model all library participants will have access to the affected journal, regardless of subscription history. (Note that **Portico** does offer a "perpetual access" feature via which a publisher may designate **Portico** as the mechanism they will use to provide ongoing access to library subscribers who have terminated a license.)

Portico and **LOCKSS** take different approaches to address a significant problem — the secure, long-term, and independent archiving of scholarly journals in their electronic form. Many publishers and libraries are content in the immediate term to experiment with both. This may be a sensible solution. If one believes in the value of assisted archiving, having a backup to the backup is a perfectly logical extension.

Back Talk from page 86

13. Contact all users of learning platforms like Blackboard, WebCT, etc., what they are doing with the digital files are not now actively using to teach their classes and suggest that they add them to the IR.

14. Ask the department(s) that process professional leave applications to send an IR brochure to all those who get travel support urging them to add their papers to the IR. Point out the value of triple dipping: give paper, put in IR, revise for publication.

15. Get all university administrators to encourage participation: To require that the intellectual output produced using University funds be placed in the repository. You need, however, to be careful about taking this latter step since many faculty members will automatically view with great suspicion anything a university's central administration mandates.

16. Look for departments that are publishing their own journals/periodicals and work with them to use the IR for that purpose.

17. Ask your university press to let you digitize their out-of-print books and add them to your IR.

18. Work with your university's archive to use the IR as a medium for exposing University related digital resources to the world.

Hopefully some of the above 18 are new to you or you have not yet done them. If you have other suggestions, please send them to me (ferguson@hkucc.hku.hk) and I will get **Katina** to list them for everyone's benefit. By the way, you are all invited to come to Hong Kong to help us celebrate the addition of our **millionth eBook**. For details, send an email to <mlyip@hkucc.hku.hk>.

In addition to surfing the Websites of repositories, I culled many of these ideas from the following sources:

Ashworth, S., Mackie, M., & Nixon, W.J. (2004). The DAEDALUS project, developing institutional repositories at **Glasgow University**: the story so far. *Library Review*, 53, 259-264.

Chan, L. and Costa, S. (2005). Participation in the global knowledge commons: Challenges and opportunities for research dissemination in developing countries. *New Library World*, 106, 141-163.

Genoni, P. (2004). Content in institutional repositories: A collection management issue. *Library Management*, 25, 300-306.

Gibbons, S. The Institutional Repository Issue. (July/August, 2004). *Library Technology Reports*, 40.

Horwood, L., Sullivan, S., Young, E., & Garner, J. (2004). OAI compliant institutional repositories and the role of library staff. *Library Management*, 25, 170-176.

Rogers, S.A. (2003). Developing an institutional Knowledge Bank at Ohio State University: From concept to action plan. *Portal: Libraries and the Academy*, 3, 125-136.

Back Talk — Institutional Repositories: Wars and Dream Fields to Which Too Few Are Coming

Column Editor: **Anthony (Tony) W. Ferguson** (Library Director, University of Hong Kong; Phone: 852 2859 2200; Fax: 852 2858 9420) <ferguson@hkucc.hku.hk>

Bertolt Brecht is supposed to have said "what if they held a war and no one came?" I am not sure this is true but that line has occurred to me as we are embarking upon the establishment of our institutional repository (IR). The same goes for the line from the 1989 film **Field of Dreams**: "if you build it they will come."

We have been thinking about jumping on the IR bandwagon for a couple of years but have continued to just say no. We recently decided it was time and so last year when I was asked to talk about IRs at a records management conference in Singapore, I said that I would talk about how to solve the problem of getting members of the university community to contribute. I had originally thought that by the time of the meeting I would have a lot of first person experiences to share. But as usual things take longer than they should.

Nancy F. Foster and **Susan Gibbons** in their "Understanding Faculty to Improve Content Recruitment for Institutional Repositories" in *D-Lib Magazine* (January 2005 identify the number one problem facing institutional repositories today:

An April 2004 survey of 45 IRs found the average number of documents to be only 1,250 per repository, with a median of 290. This is a small number when considering the hundreds of thousands of dollars and staff hours that go into establishing and maintaining an IR. For example, **MIT Libraries** estimate that their IR will cost \$285,000 annually in staffing, operating expenses, and equipment escrow. With approximately 4,000 items currently in their IR, that is over \$71 spent per item, per year.

You can bet I didn't highlight this problem when securing funding for our repository but it is certainly the most important issue that must be overcome by those chasing the IR dream. Unless our repositories get a lot fatter, all the worries about metadata and preservation, and all the hype about their rescuing scholarship from commercial publishing are sounding brass and tinkling cymbals. **Foster and Gibbons** analyze the problems faced by authors and points out that IRs simply don't address many of these issues. They then offer several good suggestions on what can be done (particularly number 3 below). Here is my cumulative list of what others are doing and some ideas of our own:

1. Look for "low hanging fruit" or digital content that is already in another campus database or otherwise immediately available. To find such material you can:

- Use OAI metadata harvesters like OAIster or subject archives to see if you already have faculty who are contributing to digital archives. In the case of the **University of Hong Kong** I found several thousand content items at other non-Hong Kong archives.
- Surf your own institution's Web pages looking for faculty members who post their own materials
- Determine if your faculty are publishing in open access journals, or serving as editorial board members of the same, they should be easy to convince to add their materials to your IR.
- Look for existing like electronic theses and dissertations collections that you could add to your IR.

- Do a Google search of your institution's Web pages to find policy statements and other valuable research related documents in digital form that need a long-term home.

2. Conduct active campaigns to educate your community about the value and uses of institutional repositories and how they can get their content into the database. This will involve f2f presentations, articles in newsletters, newspapers, etc.

3. Showcase the work of individual faculty members. Make their pages a basic unit of your IR.

4. Periodically share information with faculty members about how many times their IR items have been accessed. Publicize the top 10 IR papers/files accessed each month.

5. Call upon recently retired or soon to be retired faculty members. They are likely to have significant amounts of electronic content that they know is valuable, would like to share it with others, but are unsure what to do with it.

6. Get your library liaison officers involved. Ask them to contact faculty members who are the most active publishers or who have datasets. Once you find a few enthusiastic faculty members ask them to give presentations about IRs and their value.

7. Send out surveys to all members asking questions like who is already contributing to online archives or preprint collections, where and how do they store their digital content, the number of times they send copies of publications to others, and how confident they are that electronic copies of their unpublished materials will still be there in five or 10 years.

8. Create a special Web page for the IR and include specific step-by-step information on how to add material to the IR, how to refer people to specific articles in the IR.

9. Give faculty members information about the policies of journal publishers regarding the inclusion of preprints, articles, etc., for materials they publish in IRs.

10. Provide your scholars with the specific language they can use to request IR rights when they sign contracts/agreements with publishers before their works can be published.

11. Meet with graduate school students and encourage them to add their materials to the IR as a means of getting themselves known and to gather reactions to their work.

12. Check your university's calendar of events for conferences, symposia, etc., and ask the sponsors if they would like your help to add copies of the papers presented, etc., added to the IR.

continued on page 85

ADVERTISERS' INDEX

53	ABC-CLIO	49	ASME INTERNATIONAL	69	ENDOCRINE SOCIETY
39	ACM	5	ATG	73	FRONTIERS IN ECOLOGY AND
9	ACS	75	BASCH SUBSCRIPTIONS, INC.		THE ENVIRONMENT
55	AM. ECONOMIC ASSOCIATION	87	BOOK HOUSE	7	IDEA GROUP
41	AM. INSTITUTE OF PHYSICS	14	BROCKHAUS/GERMAN BOOKS	88	MIDWEST LIBRARY SERVICE
27	AM. PHYSIOLOGICAL SOCIETY	17	CAMBRIDGE SCIENTIFIC ABSTRACTS	31	PORTLAND PRESS
37	AM. PHYSIOLOGICAL SOCIETY	79	THE CHARLESTON ADVISOR	21	PROJECT MUSE
47	AM. PHYSIOLOGICAL SOCIETY	10	THE CHARLESTON REPORT	2	PROQUEST
61	AM. PHYSIOLOGICAL SOCIETY	13	COLD SPRING HARBOR LABORATORY PRESS	67	RITTENHOUSE
71	AM. PHYSIOLOGICAL SOCIETY	51	CROSSREF	15	SAGE
81	AM. PHYSIOLOGICAL SOCIETY	59	EASTERN BOOK COMPANY	35	SERIALS SOLUTIONS
43	ANNUAL REVIEWS	23	EBSCO INFORMATION	33	WILEY
3	APPI	11	EBSCO PUBLISHING	29	YBP
19	APPI	25	ELSEVIER		

Ad Manager: **Edna Laughrey** Internet: elaughrey@aol.com
phone: 734-429-1029; fax: 734-429-1711