7-12-2009

Institutional Repositories

Paul Royster
University of Nebraska-Lincoln, proyster@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/library_talks

Part of the Library and Information Science Commons

http://digitalcommons.unl.edu/library_talks/54

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Library Conference Presentations and Speeches by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Collecting for Digital Repositories: New Ways to Disseminate and Share Information

Institutional Repositories

American Library Association Annual Convention
Chicago, Illinois
July 12, 2009

Co-sponsored by ACRL EBSS E-Resources in Communication Studies Committee; ACRL Scholarly Communications Committee
Paul Royster
Coordinator of Scholarly Communications
UNL Libraries
University of Nebraska-Lincoln

Manager of the institutional repository:

DigitalCommons@UNL

http://digitalcommons.unl.edu
The State of IR's

"The reports of my death are greatly exaggerated."

Mark Twain, 1897
IR's: How many?

>500 worldwide, >100 in USA, including:

- Michigan
- Ohio State
- Nebraska
- MIT
- California
- Georgia Tech
- Texas A&M
- Johns Hopkins
- Brigham Young
- Rice
- Case Western
- Cal Tech
- UMass Amherst
- Cornell
- Columbia
- Colorado State
- Oklahoma State
- U Texas-El Paso
- Illinois
- Trinity
- Middlebury
- Pennsylvania
- Rochester Inst. Technology
- NYU
- Florida Atlantic
- Oregon
- Kansas
- Brandeis
- New Mexico
- Rochester
- U Conn
- Cal Poly
- Delaware
- Wayne State
- Indiana
- Boston College
- Washington
- Texas Tech
- Missouri
How large?

IR's in USA
≈ 1 million documents

IR's Worldwide
≈ 3.5 million

By contrast, Science Direct (Elsevier) lists 9.6 million articles, and claims to have 25% of the world's total.
Success?

It depends on what you measure
What we measure

• Contents
• Usage
• Participation

And there are intangibles that cannot be measured directly
4 Challenges for IRs

1. Software and implementation
2. The permissions patchwork
3. Faculty apathy
4. The "Roach Motel" issue
## Challenge # 1: Software and Implementation

### Free
1. DSpace
2. E-Prints
3. Fedora
4. Zentity (Microsoft)

### Commercial
5. Digital Commons (BEPress)
6. Content DM (OCLC)
7. Open Archive (Sun)
8. Open Repository (BMC/Springer)
9. DigiTool (ExLibris)
10. EQUELLA (Learning Edge)
11. intraLibrary (Intrallect)
12. VITAL (VTLS Inc.)

See: Repository Software Survey, March 2009 [http://www.rsp.ac.uk/software/surveyresults](http://www.rsp.ac.uk/software/surveyresults)
Open Source: "Free lunch" or "Free puppy"?

- Free software
- Your server, your IT staff
- You install, customize, host, maintain, troubleshoot, de-bug, patch, update, ...
Outsourced solution:

• Vendor installs, customizes, maintains, upgrades, & hosts
• We pay annual license fee (~ $1.50 per FTE)
• No toll on library computer resources or staff
• IR staff focuses 100% on content acquisition
Budget (at UNL)

- Salaries (1.5 FT) + software license + student workers wages $\approx$ $125,000 /year

- 4-year expenditure $\approx$ $500,000

- Yield: collection of 35,000 documents delivery of 2.5 million downloads

By way of comparison, we cut $300,000 in Elsevier publications this year (to offset their price increases).
Challenge #2: The Permissions Patchwork

Authors (and IR managers) are confused by labyrinth of publisher permissions policies
Some publishers allow use of the published version of an article:

American Physical Society
Company of Biologists
University of Chicago Press
IEEE
American Astronomical Society
American Library Association
American Mathematical Society
Am. Soc. Agricultural & Biological Eng.

American Society of Microbiologists
Cambridge University Press
Duke University Press
BioMed Central
Research Council of Canada
Animal Science Association
Society of Mammalogists
Entomological Society of America
Less than perfect, but better than some, these publishers have given authors permission to post an “author’s version,” but not their exact publisher’s version:

- Elsevier
- Springer Verlag
- Institute of Physics
- Oxford University Press
- Lippincott
- Nature Publishing Group
- John Wiley & Sons
- Taylor & Francis
- Sage Publications
- American Psychological Society
- National Academy of Sciences
- American Society of Civil Engineers
Evil only

These publishers do not allow full-text posting of any versions:

- American Chemical Society
- American Sociological Association
- American Society of Mechanical Engineers
- Karger Publishers
- Geological Society of America
- American School Psychology Association
- Mary Ann Liebert
OA content by permissions status (at UNL)

- Publisher's version: 30%
- Public domain: 33%
- UNL copyright: 25%
- Author version: 11%
- Original content: 1%
Content types

• UNL faculty articles
• University publications
• Technical reports
• Journal backfiles
• Original materials
• Works of relevance to Nebraska community
Some UNL Publications we post:

- Nebraska Swine Reports
- Nebraska Beef Cattle Reports
- Great Plains Research
- Nebraska Studies in Language, Literature, and Criticism
- Cornhusker Economics
- Manure Matters
Journals we host or archive:

- Library Philosophy and Practice
- Journal of Parasitology
- Insecta Mundi
- Court Review
- RURALS
What is not in copyright?

- pre-1923: everything ("public domain")
- 1923-1963: maybe/maybe not
  - Most © were not renewed
- 1963-1976: probably in ©
  - if published with notice
Copyright Renewal: 1923-1963

- Works published 1923-1963 have passed into public domain if they were not renewed in their 28th & 56th years of coverage.

- These can be checked at the website: http://www.scils.rutgers.edu/~lesk/copyrenew.html
The Federal Employee Loop-hole

§ 105. Subject matter of copyright: United States Government works
“Copyright protection under this title is not available for any work of the United
States Government,* ...”

*A “work of the United States Government” is a work prepared by an officer or
employee of the United States Government as part of that person's official
duties.

– Copyright Law of the United States of America and Related Laws
  Contained in Title 17 of the United States Code

This means articles by federal employees
can be posted regardless of the
publisher’s policy.
A work is **Public Domain** if any co-author is a US government employee:

- National Institutes of Health
- Department of Agriculture
- Fish & Wildlife Service
- Geological Service
- NASA
- NOAA
- Centers for Disease Control
- Department of Energy
- Department of Defense
- Veterans Administration
- National Parks Service
- *et al.*

Tip: Searching on your institution + “USDA” (etc.) can produce lots of postable articles.
State Sovereign Immunity

If you mistakenly post a work that is in copyright, your (state) institution cannot be sued for damages, because of the principle of "state sovereign immunity."

The Judicial power of the United States shall not be construed to extend to any suit in law or equity, commenced or prosecuted against one of the United States by Citizens of another State, or by Citizens or Subjects of any Foreign State.

— 11th Amendment (1793), formalizing understanding that the States had not surrendered their immunity from suit in ratifying the Constitution.
Challenge #3: Faculty Apathy

Despite the proliferation of IRs, most faculty are not motivated to self-archive or deposit their works.
4 Models for Content Acquisition:

1. "If you build it, they will come" [The articles will add themselves]
2. Make it seem fun/cool/attractive [Tom Sawyer's fence-painting]
3. Mandates: make it compulsory
4. Provide services
Content Acquisition Model #1:

“If you build it, they will come.”

W. P. Kinsella, Field of Dreams (a baseball fantasy) ....
Yogi Berra (looking at the empty seats in Cleveland’s Municipal Stadium):

“If people want to stay away, nobody can make ‘em.”
Content Acquisition Model #2

Tom Sawyer paints a fence (by persuading others it's fun)
What you may get:
Issues with self-archived materials

• permission violations

• incomplete metadata

• nasty files: poor scans, non-OCR'ed text, huge file sizes
Content Acquisition Model #3: Mandates

• Get faculty to require themselves to deposit research articles in the repository

• Follows Harvard example, passed in early 2007

• Sometimes accompanied by institution's assertion of part-ownership interest in the publication rights
Why we are not pursuing this path at UNL

1. Conflicts with our intellectual property policy
2. Would put Library in a rule-enforcement role
3. Not worth the cost in political capital and good will
4. Would not necessarily produce more deposits
5. We already have more business than we can handle

(Adds neither carrot nor stick to our repertoire.)
Collecting 101

Honey

Vinegar
Content Acquisition Model #4:

Provide Services

“Opportunity is missed by most people because it is dressed in overalls and it looks like work.”

— Thomas Edison
Services UNL provides:

- permissioning
- hunting and gathering
- scanning
- typesetting
- metadata-ing
- uploading & posting
- usage reporting
- promoting
- POD publication
Useful Tools & Skills

- Adobe Acrobat
- Adobe Photoshop
- Adobe InDesign (or Quark Xpress)
- MS Word
- scanning
- graphic design
- proofreading
- copy-editing
Some fields are easier than others.

We work all across the board, but do find some areas are easier pickings:

• **Physics:** professors publish a lot and have many co-authors. Most major journals allow their PDFs to be used.

• **Electrical engineering**

• **Biology & microbiology**

• **Natural resources**

• **Agronomy**

• **Animal science**
And some fields are harder, *e.g.*

- mechanical engineering
- chemistry
- geology
- medicine

But even in these, there are postable articles to be found.
Which professors should I pick on?

My advice: Go for the big names, the senior chaired profs with the long *vitae*.

Junior faculty (who would benefit much more)
  a.) have fewer articles, and
  b.) have more reservations about online publication.
(And I realize this is counter-intuitive.)
Most successful recruiting strategy:

1. Find postable articles
2. Email the authors ("I have recently seen your article ....)
3. Request permission and additional publications list
How do I find postable articles?

• Use SHERPA/RoMEO publisher site (or OAKList) to find publishers who allow posting

• Search those publishers' sites for your institution name

Challenge #4: The "Roach Motel"

or, the belief that items archived in an institutional repository will remain there unfound and unused—"They don't check out!"

With a tip o' the cap to Dorothea Salo
Our Experience at UNL
We furnished 137,072 downloads in May 2009.
77% of Open-Access content was downloaded in May 2009

- Downloaded: 17460 (77%)
- Not downloaded: 5320 (23%)
Faculty Publications, Department of Psychology

http://digitalcommons.unl.edu/psychfacpub/

367 articles → 5,008 downloads

avg. = 13.6
Robert Katz* Publications

http://digitalcommons.unl.edu/physicskatz/

190 articles → 1,357 downloads

avg. = 7.1

* retired in 1987
UNL Larsen Tractor Museum Archives

http://digitalcommons.unl.edu/tractormuseumlit/

2,274 articles → 16,648 downloads

avg. = 7.3
Dissertations: Department of History

http://digitalcommons.unl.edu/historydiss/

19 documents → 994 downloads

avg. = 52.3

---------------------

Dissertations: Modern Languages and Literatures

http://digitalcommons.unl.edu/modlangdiss/

6 documents → 984 downloads

avg. = 164.0 !!
Most Downloaded Work:

Online Dictionary of Invertebrate Zoology

http://digitalcommons.unl.edu/onlinedictinvertzoology/

1 document → 824 downloads

24 documents (including separate letters) → 1,903 downloads
36,000 downloads (26%) went to international users

3,999 United Kingdom
3,856 Canada
3,109 India
2,261 Australia
1,363 Germany
1,148 France
1,126 China
  878 Brazil
  848 Spain
  773 Mexico
  743 South Africa
  723 Italy
  645 Pakistan
  629 Turkey
  619 Poland

147 countries in all (plus the USA)
10% of our traffic comes from within the state of Nebraska (pop. 1.7 million).

About 7% of site traffic comes from Lincoln, NE.
## Traffic Sources

<table>
<thead>
<tr>
<th>Source Type</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search engines</strong></td>
<td>63.3%</td>
</tr>
<tr>
<td>Google</td>
<td>56.0%</td>
</tr>
<tr>
<td>Yahoo</td>
<td>4.2%</td>
</tr>
<tr>
<td>other search</td>
<td>3.1%</td>
</tr>
<tr>
<td><strong>Referring sites</strong></td>
<td>26.4%</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>9.5%</td>
</tr>
<tr>
<td>UNL websites</td>
<td>6.0%</td>
</tr>
<tr>
<td>Online Books Page</td>
<td>1.2%</td>
</tr>
<tr>
<td>other</td>
<td>9.7%</td>
</tr>
<tr>
<td><strong>Direct traffic</strong></td>
<td>10.3%</td>
</tr>
</tbody>
</table>

**Total:** 100.0%
Scholarly Communication

We are entering an era of competition between:

• The restricted-access, for-profit, scholarship-as-property publishers, and

• The open-access, for-knowledge, scholarship-as-shared-resource publishers and re-publishers

And that is what repositories essentially are — publishers and re-publishers. Our clientele is the world, not just our local campus.
Asymmetrical Competition:

The Non-Level Playing Field
Goal: Maximize revenues

Means: Control access

Holdings: 40 million articles

Strategies: Conventional

User universe: 20 million

Author feedback: no
Repositories

Goal: Maximize distribution

Means: Open access

Holdings: 14 million articles

Strategies: Innovative

User universe: 1 billion

Author feedback: yes
Documents in OA Repositories (worldwide)

Source: Registry of Open-Access Repositories
Collection strategies @ UNL

1. Be inclusive, not exclusive
2. Be proactive, even aggressively so
3. Think of the global audience
4. Everything open access
5. Everything full-text
6. Ample metadata—especially abstracts
7. Utilize work-study students
8. Link back to your site
9. Give depositors feedback — publishers don't
10. Measure, measure, measure, . . .
Thank You!
Good afternoon. I was asked today to speak about Institutional Repositories, and thank you for asking me, Adrian. I do appreciate the soap-box.

First of all, let me warn you, this will be a presentation of the view from the trenches, or from the front lines; not an angel’s-eye picture, but more like a worm’s eye view. I have today a collection of anecdotes and statistics and advice and adages based on 4 years of IR management. I am not a systems guru, or an open-access philosopher, or even a librarian. I am a publishing "has-been" with 63 Powerpoint slides, so I better get going ...

Let’s start by asking “What is the state of the institutional repositories?” They were introduced 6 or more years ago with great expectations; a number of institutions have adopted them—but in the years since, the rumor has circulated that they have not lived up to these early promises. Some adopters have been disappointed at the lack of action and results, and I will address some of their challenges in a minute, but let me first say a word in defense of what the IRs in general have already achieved, and refute the idea of their early and untimely demise—as Mark Twain so pungently said “Reports of my death are greatly exaggerated.”

How many institutional repositories are there? According to the Registry of Open Access Repositories, there are more than 500 worldwide, and more than 100 in the United States. I show here 40 or so, and I apologize if I have left off those of anyone in this audience.

How large are these IRs, and what do they hold? Well, those in the US hold almost a million documents. Worldwide there are about 3.5 million documents in institutional repositories. By contrast, Elsevier’s Science Direct claims 9 and half million documents, and contends that it holds about one quarter of the worlds scholarly articles, which would put the whole number available somewhere around 40 million. By comparison, does this make the institutional repositories a success or a failure?
It depends on how and what you measure.

What do we measure, at Nebraska?

1. Contents -- how many documents we have. Currently the number is 35,000, which puts us second among the traditionally configured IRs, behind Michigan’s Deep Blue.

2. Usage -- as measured by the number of downloads. And we regard downloads as a better measure than simple traffic. More on this later.

3. Participation -- we think right now that we have about 30% of our campus faculty participating in one way or another.

Intangibles or non-quantifiables include things like reputation, degree of acceptance, viral spread, etc.

What we are really interested in in all of these cases is seeing progress and improvement in the statistics; that’s the whole reason for measuring—to tell us if we’re doing something fruitful, or not.

Here are what I call the 4 challenges for institutional repositories, and I will work through them in order.

**Challenge #1: Software and implementation** — So you want to start a repository ... Where do you turn? There’s the free open-source systems D-SPace, Fedora, E-Prints; and a new Microsoft freeware. And there are now at least 7 commercial systems, all of which promise to create, manage, and optimize your repository experience. I won’t go into them all, but I will point out a recent comparison survey at the URL shown that compares their features.

Now the open-source systems — D-SPace, Fedora, E-Prints — as someone said at the SPARC Repositories meeting in Baltimore last November: “Free lunch or free puppy.” There’s a Latin proverb, “Prandium gratis non est”—or in English, there is no free lunch. Now we do not run any one of these systems at Nebraska, but what I seem to observe at those sites who do, is that there is a considerable commitment of time, energy, and resources
devoted to set-up, maintenance, operability, and engineering of the system. So the free puppy requires a good deal of care and investment.

[CLICK]
At Nebraska, we use an outsourced system—the Digital Commons from Berkeley Electronic Press. We pay an annual license fee, and we don’t worry about the system. We spend all our time recruiting and developing the content.

[CLICK]
To show you what it costs us: we have 1-½ full-time staff, our license fee, and a $5,000 budget for work-study students — that’s all. In total, about $125,000 annually, or about $500,000 in the four years since inception. Our yield on that half million dollar investment has been roughly 35,000 documents archived and 2.5 million downloads furnished.

[CLICK]
**Challenge #2: The Permissions Patchwork** — the crazy quilt of publishers’ permissions policies for archiving in an institutional repository. This is, I believe, one of the largest obstacles to faculty participation. Everyone is confused by what you can and cannot post, and frozen by the uncertainty.

[CLICK]
Well, let me simplify it. You have the good guys, who are the true scholarly publishers, who do not stand in the way of research dissemination. You can post their content; you can use their files or page images. These folks deserve a round of applause or a free drink at the bar. Librarians: don’t ever cancel your subscriptions with these publishers!

[CLICK]
Next, we have those where good and evil are mixed: you can post the content, but you cannot use the publisher’s version. And these are mostly the commercial publishers, but also some so-called learned societies who believe in the dissemination of knowledge, but only so far as authors can be discouraged from encroaching on the publisher’s property rights.

[CLICK]
And finally, we have the unalloyed and unmitigated “evil only”—no posting, no versions, no way, no how.
Here is an approximate pie chart of our content mix by permissions status. Roughly 30% publishers versions (with permission), roughly 33% public domain, 25% copyrighted by our own institution, 10% author versions as allowed by publishers, and about 2% original materials--copyright by the authors.

What types of content do we have?
- UNL faculty articles
- University publications
- Technical reports
- Journal backfiles
- Original materials
- Works of relevance to the Nebraska community

File types are probably 99% PDF, with the remainder being spreadsheet, PowerPoint, and MP3 or MPG files.

I spoke of university publications—here are some examples. You see we are big into livestock and their by-products.

We also host backfiles of journals from outside the university; and we have 1—and only 1—open access journal that we started and run.

Public domain: There is a lot more than you might think that is not in copyright. First: everything from before 1923, everything before 1963 that was not renewed, and everything before 1976 that did not carry the copyright declaration.

There is a website to check copyright renewals for things published 1923-1963. The vast majority were not renewed and are now public domain.

Works by federal government employees cannot be copyrighted. Even though publishers put the copyright symbol on them, and you would assume
since it says “copyright” that it is, but legally it is not. Do not be intimidated by the publishers’ disinformation, or what one author calls “copyfraud.”

[CLICK]
If any co-author of a research paper is an employee of one of these agencies, the work is in the public domain. Take advantage of this. Assert the public’s right of access. Post these articles. With glee. No permissions, no embargoes, no conditions.

[CLICK]
If you accidentally go too far, and mistakenly violate someone’s copyright, the state schools among us have a get-out-of-jail-free card in the legal doctrine of “state sovereign immunity.” A state cannot be sued for damages for violations of federal law (such as copyright). It can be made to “cease and desist,” in which case you just pull the article down; but you are not liable for damages, and you have essentially nothing to lose.

[CLICK]  
**Challenge #3 — Faculty apathy.** Despite the proliferation of IRs, most faculty are not motivated to self-archive or deposit their works. No other IR challenge has received so much blogging space, conversation, or research funding as this one. The repository is open for business, but the faculty is apparently failing to do their part—they are not coming through with the content.

[CLICK]  
I want to suggest that we have seen 4 models for content acquisition

1. "If you build it, they will come" [The articles will add themselves]
2. Make it seem fun/cool/attractive [Tom Sawyer’s fence-painting]
3. Mandates: that is, make it compulsory
4. Provide services

Now bear in mind that models are miniaturized and exaggerated versions of the real thing, and I recount these not to critique the efforts any of my colleagues, but rather to describe or explain our own success or lack of it with them.

[CLICK]  
**Model #1:** "If you build it, they will come" a phrase from W. P. Kinsella’s novel (and later movie) *Field of Dreams*. In your dreams is right. Build the baseball diamond and the legends will walk right out of the cornfield. Well, if
there’s one thing we have in Nebraska, it’s cornfields. So we built the IR, told the faculty about it, but very very few came to play.

[CLICK]
It’s more like what Yogi Berra once said: “If people want to stay away, nobody can make ‘em.”

[CLICK]
**Model #2:** Tom Sawyer’s fence-painting; Make it seem fun/cool/attractive or somehow desirable (which, in fact it is). And so I put up some stuff, and went around going, “Look, see how cool ! Don’t you want to, too …” But the fence-painting we got looked more like this:

[CLICK]
[graffiti fences] The few self-archived articles we received have been almost without exception problematic.

[CLICK]
Sometimes authors self-archive versions from publishers who do not permit posting. They leave out any publication or copyright information, co-authors, and abstracts. They submit “nasty” files, badly scanned, without OCR-ing the text, or crank up the scanner resolution to where an 8-page article becomes a 180 meg file. It takes longer to fix one of these than it would to gather and post it right the first time.

[CLICK]
**Model #3:** Mandates: Get faculty to require themselves to deposit articles in the repository. This model follows the Harvard example, passed in early 2007. Sometimes it is accompanied by the institution’s assertion of part-ownership interest in the publication rights.

[CLICK]
Mandates are very hot, and every campus is supposed to convince their faculty to place this new and additional requirement upon themselves, for the sake of open access and the good of mankind. We have discussed this option at Nebraska and decided not to pursue it. Here’s why:

1. It conflicts with our intellectual property policy, which allows the faculty author complete control and ownership of their own output.
2. It would put Library in a rule-enforcement role; we can’t even get faculty to pay overdue fines.
3. It is not worth the cost in political capital and good will.
4. It would not necessarily produce more deposits; since most mandates have opt-out provisions and none have any means of compelling compliance.
5. We already have more business than we can handle.
In short, we feel that a mandate would add neither a carrot nor a stick to our repertoire. Now I understand that it’s a very different situation for NIH, where they have 250 million carrots to award.

I think it’s like the entomology article that I read which proved that a nectar-derived fructose-sucrose solution of *Apis mellifera* secretions enabled the capture of a significantly larger population of *Musca domestica* than an acidic solution with pH of 2.4 derived from ethanol fermentation.

**Model #4: Provide services.**
Thomas Edison: “Opportunity is missed by most people because it is dressed in overalls and it looks like work.” Well, it looks like work because it is, but the opportunity is there and the work is useful and rewarding.

Services we provide at Nebraska include:
- permissioning
- hunting and gathering
- scanning
- typesetting — we try to typeset original materials to look professional, and we typeset author versions to match the published edition for layout, pagination, footnoting, etc.
- metadata creation
- uploading & posting
- usage reporting
- promoting
- POD publication

These are the skills and tools we use on an everyday basis. It’s interesting—I recently saw a white paper from the JISC listserv that discussed the skill set an IR manager should have, and I didn’t have any of them. Their set was
almost exclusively programming and IT skills, and I think what has set Nebraska apart from many IRs is that we think of ourselves as a publishing operation, not a data or technology operation. I only know about 5 html codes, but I live in Adobe Acrobat. I don’t see how you could manage a set of PDF documents without the ability to massage, manipulate, adjust, and tinker with PDF files. Contents do not come to our doorstep as perfectly finished jewels — they are diamonds in the rough, and very often need some cutting and polishing before they meet the world.

In recruiting content, I’ve discovered that some fields are easier than others, notably physics, electrical engineering, biology & microbiology, natural resources, agronomy, and animal science.

And some fields are harder, e.g. mechanical engineering, chemistry, geology, and medicine. But even in these, there are postable articles to be found.

Who should you start with? Which professors should I pick on? The big names, the senior chaired professors with the long vitae.

My most successful recruiting strategy? Find the postable articles first, and then approach the authors for their permission to upload them.

Where can you find postable articles? Sherpa/Romeo (and bless them for their work) and QUT’s Oaklist show hundreds of publishers who allow their files to be used. Search those publishers’ sites for works from your institution.

**Challenge #4: The "Roach Motel"**

And a shout-out here to Dorothea Salo, who coined the term, or first applied it to IRs. Specifically, the belief that items archived in an institutional repository will remain there unfound and unused—"They don't check out!" Or as it used to say on the subway: "Las cucarachas entran, per no puede salir."
[Note: Dorothea Salo has graciously pointed out that I misinterpreted her use of the term. In fact, her “roach motel” model refers to the situation that (at some repositories) a submitted paper cannot be removed or revised by its author, not to the lack of usage or downloads by the public. “See her Twitter post: "Roach Motel" actually referred to depositors’ inability to change IR content or use the IR for in-progress work, but OK.” #ala20091:54 PM Jul 12th from web; see also her original article "Innkeeper at the Roach Motel." Library Trends 57:2 (Fall 2008), online at http://minds.wisconsin.edu/handle/1793/22088 ]

The best way I can think to address this issue is to describe our experience at UNL: we have them checking out by the thousands every day.

We furnished 137,072 downloads in May 2009.

77% of our Open-Access content was downloaded in May 2009.

Our Psychology Department has 367 articles online, which produced 5,008 downloads in May, an average of 13.6 per article, for the month.

Robert Katz—who retired from the Physics Department in 1987, over 20 years ago — has 190 publications online, which produced 1,357 downloads in May. He called me up when he got his monthly downloads report—“These numbers are astronomical!” He added up all his downloads to date (over 11,000)—“You have resurrected my scholarly work!”

Our biggest series is the UNL Larsen Tractor Museum Archives, with 2,274 articles. It had 16,648 downloads for the month.

Most surprising (to me) is the action we get on PhD dissertations. In the History Department, we have 19 documents online that produced 994 downloads, an average of 52 each for the month. And even more incredible is the Dissertations from our Modern Languages and Literatures department,
where only six documents produced 984 downloads, an average of 164 each — for one month !! And 4 of these 6 dissertations are in Spanish.

[CLICK] Our most downloaded document was an original work, the Online Dictionary of Invertebrate Zoology, a 900-page dictionary, with 824 downloads. If you include the separate-letter files that we split out for easier access, parts of the work were downloaded 1,903 times.

[CLICK] We sent 36,000 downloads (26% of the total) to international users in 147 countries.

[CLICK] 10% of our traffic comes from within the state of Nebraska (population 1.7 million). About 7% of site traffic comes from our hometown and campus in Lincoln, NE. Some of those little red dots you see across the state are not much more than 40 cows and a general store, but they’re finding us and using the resources.

[CLICK] Where does our traffic come from? Well, this report is from Google Analytics and it shows traffic sources for May 2009: 63% came via search engines, with the vast majority of that from Google and Google Scholar. 26% was referrals from other sites, with the leading one being Wikipedia, where we actively place links from relevant articles to appropriate materials—these accounted for almost 10% of traffic. Our own campus website links provided 6%; the Online Books Page out of Penn’s Library sent us 1.2%. Direct traffic—people who have bookmarked the site or who type in the URL—were slightly over 10%.

[CLICK] I believe we are entering an era of competition between:
• The restricted-access, for-profit, scholarship-as-property publishers, and
• The open-access, for-knowledge, scholarship-as-shared-resource publishers and re-publishers.
And that is what repositories essentially are — publishers and re-publishers. Our clientele is the world, not just our local campus.
This is—in modern military terminology—an *asymmetrical* competition. The contestants have totally different aims and strategies, and sometimes the tilted playing field favors one side, sometimes the other. Or, it's like one team is playing football and the other is playing Ultimate Frisbee.

[CLICK]
The Publishers’ goal is to maximize revenues; their means is by controlling access. Their holdings are about 40 million articles. Their strategies are the time-honored and traditional ones. Their user universe is roughly 20 million, although no single publication reaches that many. And do they provide authors with feedback? No.

[CLICK]
The Repositories, on the other hand, are the little fish. Their goal is to maximize distribution; their means is Open access. Their holdings, currently, are about 14 million articles. Their strategies require innovation, change, departures, thinking outside the box. Their universe of potential users is 1 billion and growing. And they can provide what publisher’s don’t: Author feedback.

[CLICK]
Look at the growth over the last 4 years in open-access repository contents, as recorded by the ROAR: from less than 2 million to almost 15 million, an 8- or 9-fold increase. If this continues—and my aim today has been to encourage its continuance—in four years from now, there could be over 100 million open access works available. But we need more little fish; and we need those little fish to work as hard as they can at pumping up their contents.

[CLICK]
Finally, let me recap our strategic 10 commandments for content recruitment at Nebraska:
1. Be inclusive, not exclusive — *the public will decide what it wants*
2. Be proactive, even aggressively so — *push the envelope*
3. Think of the global audience — *you will have users in Namibia and Uzbekistan*
4. Everything open access — *I am disdainful when I hit a link in an supposedly open access repository and it takes me to a toll-access article on a publisher’s site. Hrumph, I say.*
5. Everything full-text — *what good is a citation by itself?*
6. Ample metadata—especially abstracts — help the search engines do their job
7. Utilize work-study students — They are willing, able, & affordable
8. Link back to your site — It helps your traffic and your Google rankings.
9. Give depositors feedback — Publishers don't, and it's our greatest advantage in getting the faculty excited and involved.
10. Measure, measure, measure, . . . — and share your numbers; they tell a most interesting story.

This ... is a note to myself.

So, ... thank you very much for your patience and indulgence.