6-30-2008

How to Fill Your Institutional Repository: Or, Practical Lessons I Learned by Doing

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

How to Fill Your Institutional Repository

Or,

Practical Lessons I Learned by Doing

American Library Association Convention
Anaheim, California
June 30, 2008

Paul Royster

Coordinator for Scholarly Communications

University of Nebraska-Lincoln Libraries

proyster@unl.edu

UNIVERSITY OF
Nebraska
Lincoln

And manager of ….

University of Nebraska-Lincoln's Institutional Repository

http://digitalcommons.unl.edu
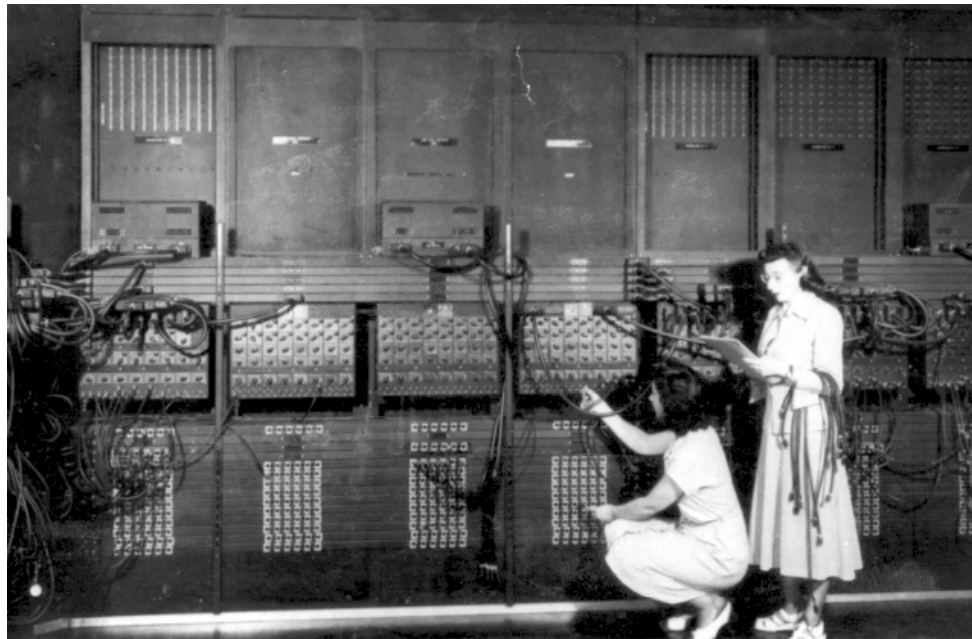
Where I started: Self-Archiving, or ...

"The articles will add themselves."

Self-Archiving: The Ideal View
(fishing metaphor)

Plan A: "Fish Jump into Boat"

- I started out armed with a slick PowerPoint presentation showing 1) the advantages of having your articles online in an open-access repository, and 2) how easy it was to do.

- I went to dozens of department meetings and gave my spiel seeking to recruit participants who would upload their own articles.

- Response rate: <10%

- Maybe I need a new metaphor?

"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):
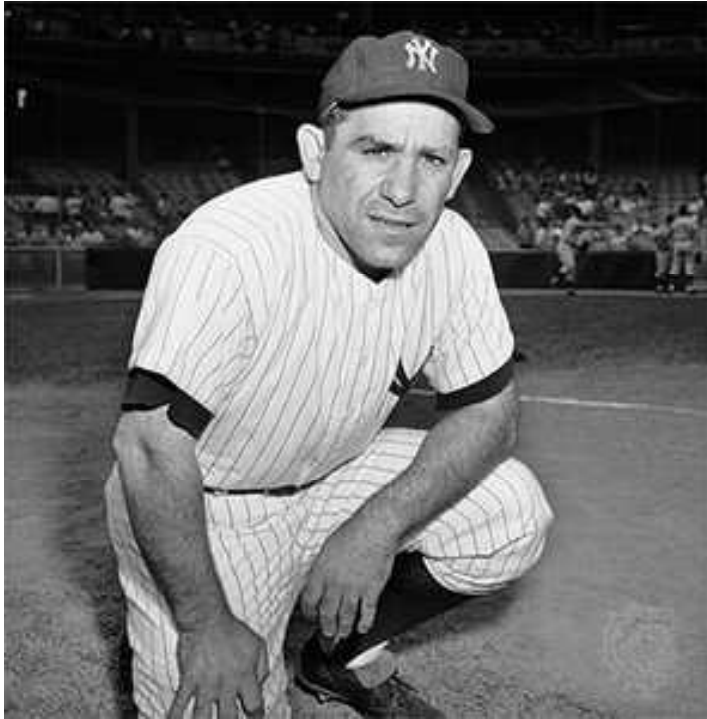
"If people wanna stay away, nobody can make 'em."

So, we went to **Plan B ...**

After 6 months of very poor fishing, I improved the offer to "We will do it for you; just send me your vita or publication list."

We call this the DIFM model ("Do it for me").

Also known as "mediated deposit."

"Mediated deposit" entails

- Finding and gathering articles
- Permissioning
- Scanning, re-typesetting, & file preparation as needed
- Metadata preparation
- Uploading or posting

Advantages (apart from greater uptake) include better copyright/permissions compliance, consistency of metadata, & file quality control.

What happened ?

- Response rate improved to ~ 15% - 20%

- Then I met with the Physics department, and their response rate was closer to 70% --
  (I received vitae totaling almost 2,000 articles)

"Be careful what you wish for …"

End of Year 1 (June 2006)

- About 2,400 articles added
- Usage rate ~ 10,000 downloads/month
- Large back-log of postable articles

- Fortunately, our Dean came through with a small budget ($4,000) for student help

Student Workers

- With federal work-study funds paying 80% of their wages, this $4,000 allows me to use 2 students 8-12 hours/week for the academic year.

- One does uploading: her average output is 30-40 articles/day—about 15 articles/hour.
  Avg cost = 10 cents/article

- One does scanning: his average output is about 40-50 pages/hour, from bound book to postable text-OCR'ed PDF file.
  Avg cost = 3 cents/page
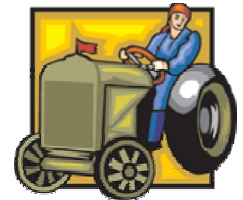
Now I had a different problem:

How to line up enough work to

keep these students busy ?

"Necessity, the mother of invention"
— Plato, The Republic

My New Approach – **Plan C**

1. Find some postable articles first

2. Then approach the author for permission

Success rate ~ 90%

Where can I find postable articles?

1. Journals/publishers who allow their pdf's to be used

Check SHERPA/RoMEO for publishers' policies

http://www.sherpa.ac.uk/romeo.php?all=yes

Some examples:
OK to use publisher's PDFs from

Cambridge University Press

Duke University Press

University of Chicago Press

Rockefeller University Press

American Physical Society

American Astronomical Society

American Society of Microbiologists

Company of Biologists

Public Library of Science

BioMed Central

EDP Sciences

American Fisheries Society

Crop Science Society

American Society for Agronomy

Soil Science Society

IEEE

SPIE

American Entomological Society

American Psychonomic Society

Haworth Press

American Mathematical Society

Society for Animal Science

Dairy Science Association

American Library Association

How do you find the articles?

- Go to the online journal or publisher's site

- Use "Advanced Search"

- Select "Affiliation" and plug in your institution

- If "Affiliation" is not offered, use "full text search"— less efficient, but still effective

Step 2: Email the author

Eileen A Hebets
School of Biological Sciences, University of Nebraska-Lincoln

Dear Dr. Hebets,

I recently came across one of your articles--"Cross-modal effects on learning: a seismic stimulus improves color discrimination learning in a jumping spider" in *The Journal of Experimental Biology* 210 (2007), 3689-3695 --- that is eligible for posting in UNL's online institutional repository--http://digitalcommons.unl.edu. The Company of Biologists (publisher of *JEB*) allows their versions to be used in a university's repository, so it's a simple matter to archive it there.

And I note from perusing **your online vita** < http://www.biosci.unl.edu/faculty/Hebets/publications.html > that you have written many others that are eligible as well.

Unless you have an objection, we would very much like to add your eligible articles to the repository, which is a digital archive for all scholarship, research, and creative activity from people associated with UNL. We have more than 23,000 documents from more than 500 faculty authors. It is the nation's 4th largest institutional repository (after Michigan, Ohio State, & MIT).

Archiving will preserve them permanently as part of the UNL Libraries' digital collections and also make them open-access for all Internet users worldwide. We would post them in one of the "Papers in Biological Sciences" series < http://digitalcommons.unl.edu/bioscipapers/ > in the School of Biological Sciences area.

Each full-text article would have a permanent, separate (and reasonable-length) URL that can be pasted into any email, vita, or hypertext document, or linked from any web page.

Online open-access articles are viewed and cited much more frequently than those available only in paper or only by limited subscription. It is one of the best ways to maximize the impact of your research work. If you agree to allow us to do this, you will get an automatic email each time we upload one of your articles, and you will receive monthly reports (by email) on how often each document is downloaded.

May we have your permission to proceed ? If we encounter others of your articles, may we have permission to include those as well (always considering the publisher's policies and guidelines) ? If you have updates to your online list of publications, we would be happy to work on gathering and posting them, too.

Thank you for your consideration. Sincerely,

Paul Royster
Scholarly Communication
306 Love Library
University of Nebraska-Lincoln
402 472-3628
proyster@unl.edu
http://digitalcommons.unl.edu
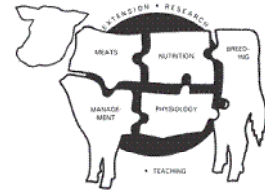
Other types of postable articles

Publications of your university

- newsletters

- annual reports

- extension publications

- journals, monograph series, technical reports

You should frequently browse & search
your institution's web pages and library catalogue.

UNL Publications we post:

- Nebraska Swine Reports

- Nebraska Beef Cattle Reports

- Mid-West Quarterly

- Nebraska Studies in Language, Literature, and Criticism

- Cornhusker Economics

Other types of postable articles (2)

Public domain documents

- Published pre-1923

- Copyright 1923–1963 <u>and not renewed</u> *(more than 93% of books were <u>not</u> renewed)*

- Published pre-1976 without a copyright notice

- U.S. Patents

A work is Public Domain if <u>any co-author</u> is a US government employee:

- National Institutes of Health
- US Department of Agriculture
- US Fish & Wildlife Service
- US Geological Service
- NASA
- NOAA
- Centers for Disease Control
- Department of Energy
- Department of Defense
- Veterans Administration
- National Parks Service
- Livermore National Laboratory (through 2007)

Tip: Searching on your institution + "USDA" (etc.) can produce lots of postable articles.

Find and form partnerships

Groups already engaged in digitizing or publishing
who may need a hosting platform

Some examples from our site include:

UN State Museum Parasitology Lab

Journal of Parasitology backfile

1915 - 1974

UN Tractor Test and Power Museum

1500+ tractor test reports 1919 - 2008

Internet Center for Wildlife Damage Management

2000+ conference proceedings and symposia articles

National Wildlife Research Center (USDA)

Staff publications

~ 800 articles

Florida State Collection of Arthropods

Center for Systematic Entomology

Insecta Mundi backfile 1985 – current

500+ articles

Insecta Mundi

Don't overlook the emeritus faculty

1. They have the time to get involved

2. They have the articles -- a career's worth

3. They are concerned about what happens to their legacy when everything becomes online
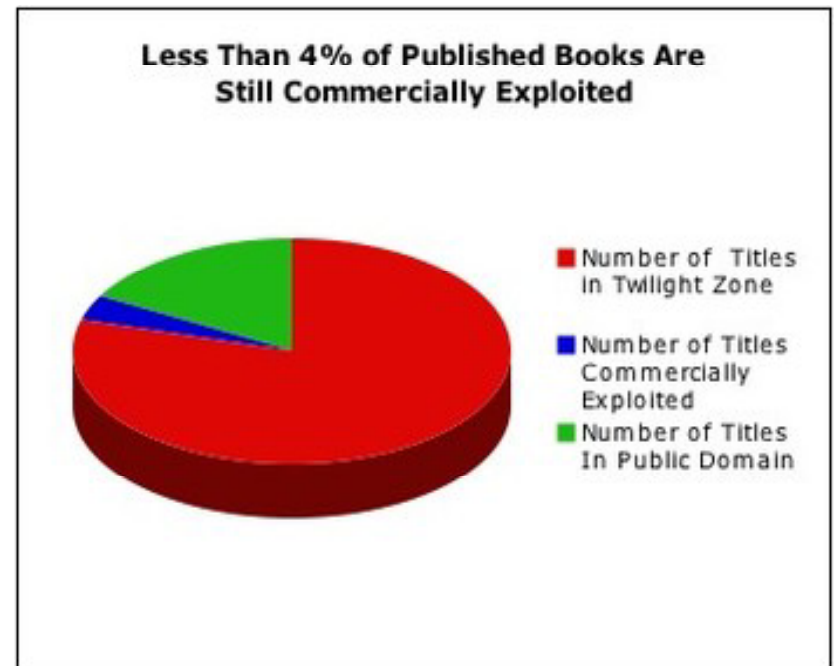
"Longevity"

OA dissertations are valuable and popular content

- ProQuest/UMI deposit agreement is non-exclusive-- the PhD has the right to publish anywhere else

- Our open-access dissertations get hundreds of downloads

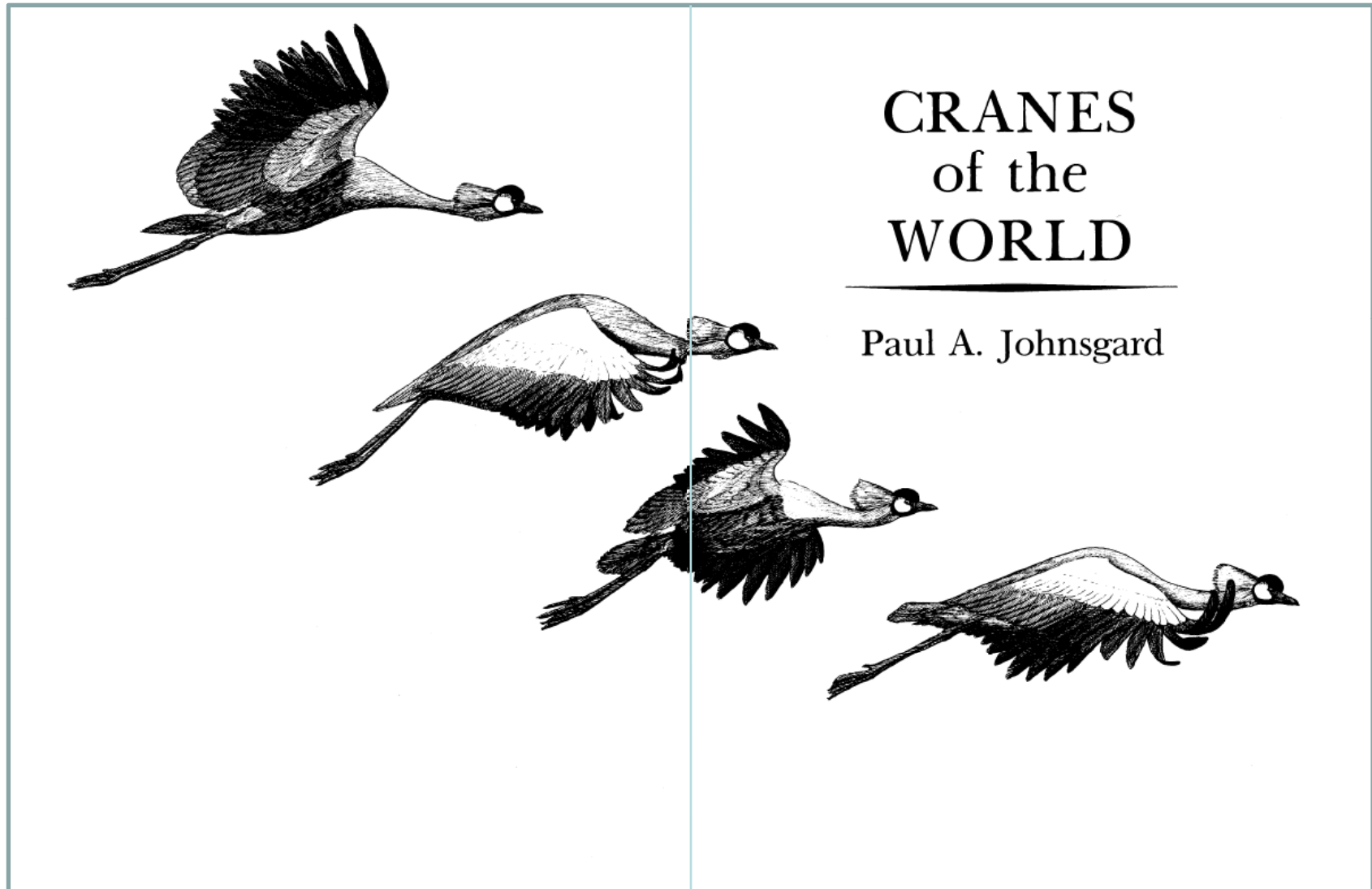- Currently gathering only about 20% of new PhD's

Out-of-print books

- Most publishing contracts include a clause that gives the rights back to the author if the book goes out-of-print.

- Listing it with the Online Books Page draws substantial activity immediately: http://onlinebooks.library.upenn.edu/

- Some samples:

Less Than 4% of Published Books Are Still Commercially Exploited

■ Number of Titles in Twilight Zone

■ Number of Titles Commercially Exploited

■ Number of Titles In Public Domain

279 pages, 8½ x 11", 26 Mbytes
2,038 downloads in 4 months

CRANES
of the
WORLD

Paul A. Johnsgard

NEBRASKA WILD FLOWERS

Robert C. Lommasson

Electronic edition

University of Nebraska-Lincoln Libraries

2008

Plate 63   Western wallflower
Erysimum asperum

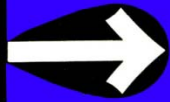Plate 64   Dame's rocket
Hesperis matronalis

Plate 65   Silvery bladderpod
Lesquerella ludoviciana

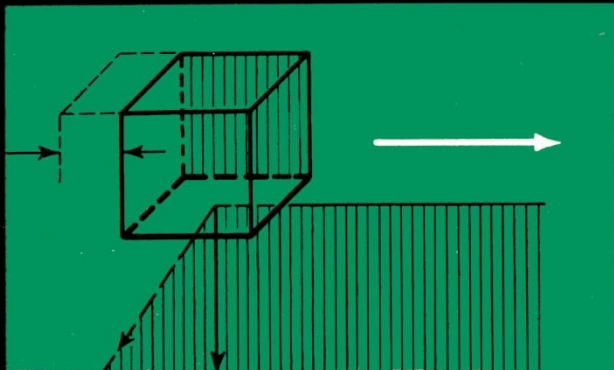Plate 66   Water cress
Rorippa nasturtium-aquaticum

Plate 67   Tumble mustard
Sisymbrium loeselii

COMMISSION ON COLLEGE PHYSICS

An Introduction to the Special Theory of Relativity

ROBERT KATZ

WATERFOWL
OF NORTH DAKOTA

Paul A. Johnsgard

NORTH DAKOTA INSTITUTE FOR REGIONAL STUDIES
NORTH DAKOTA AGRICULTURAL COLLEGE
Fargo

End of Year 2 (June 2007)

- Added 6,000 articles

- Usage increased 3-fold, to 30,000 downloads/month

End of Year 3 (June 2008)

- 7,000+ articles added
- Usage > 120,000 downloads/month

- 75% of downloads come via Google searches
- 1/3 of Google searches (= 25% of total downloads) come from outside USA
- In a normal month, we furnish downloads to users in over 120 countries

Measures of acceptance

Now, our most prolific source of content recruitment is → **word of mouth**: faculty telling other faculty that this is something they should want to do.

Lately, I've been getting about 2 "fish jumping into the boat" per week. Last week I got an entire college.

Most effective "resale" device for creating faculty loyalty & satisfaction: monthly download reports.

Some fields are easier than others.

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

- **Physics:** professors publish a lot and have many co-authors. And most major journals allow the use of their PDFs

- 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 named-chair profs with the extensive *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.)

Objection #1

Q: In many cases, you are only putting up something that is online already; so what have you achieved?

A: At least 4 things
1. Open access, in some cases
2. Author (and co-author) participation
3. Direct association with your institution
4. Preservation, regardless of what happens at original site
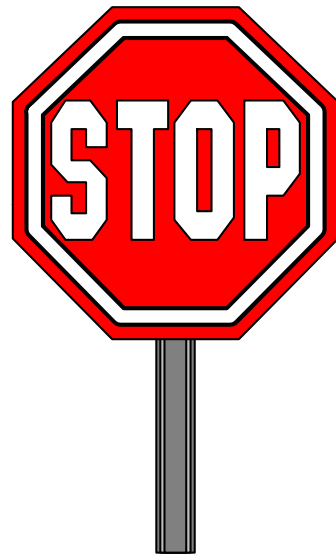
Objection #2

Q. Some of what you suggest will yield old and out of date articles; what is the point?

A. Recency has very little to do with usage. Many of our most-downloaded documents are dozens or even hundreds of years old. "Most relevant" is far more important than "most recent."

Objection #3

Q. **Isn't the IR just a "roach motel" for articles ?**

A. **I don't think of ours that way because:**

- Each month more than 75% of available documents get downloaded at least once, and the average document is downloaded more than 7 times.

- Many obscure or neglected articles have a revived new life online.

- We have established the **Library** as the best place on campus for electronic publication.

- We are furnishing UNL-branded content to hundreds of thousands of users worldwide

for questions