

The What, Where, Why, and How of Institutional Repositories

Marcy E. Rosenkrantz Director of Library Systems





What

- A place to store the intellectual output of an institution or community in digital form, on-line.
 - Predominantly Open Access
 - Usually "born digital" material
 - Data sets, working papers, informal publications, preprints, reprints, course materials, electronic theses and dissertations, etc.
- A source of journal articles for (formal and informal) publications
 - Overlay journals
- The vehicle or software system is not the repository
- · Sometimes called an archive, but...
- Is not the same as the underlying software

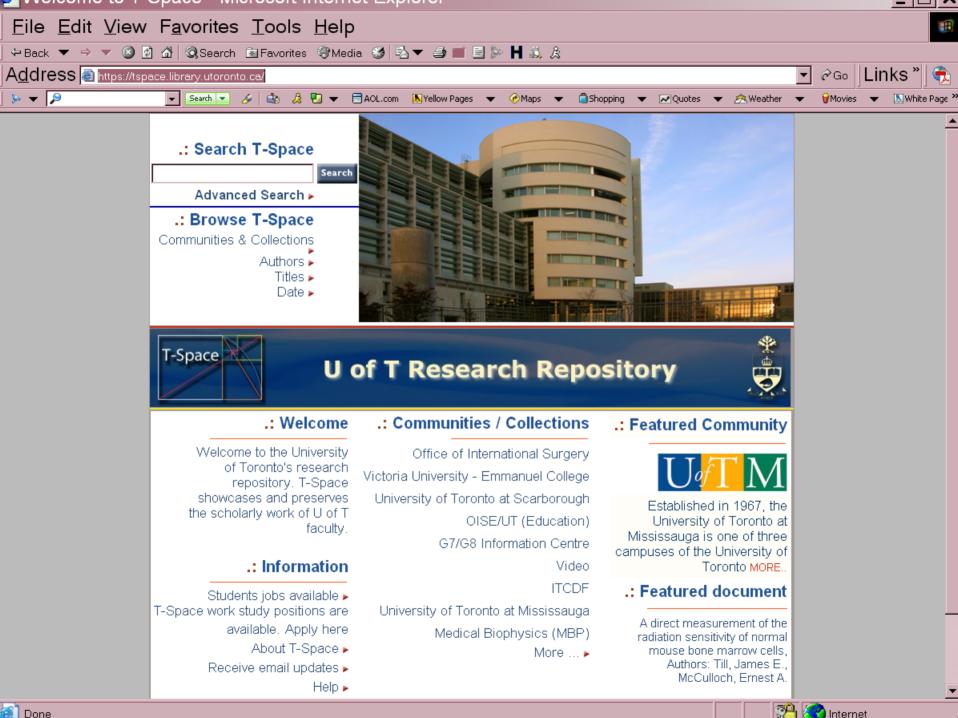


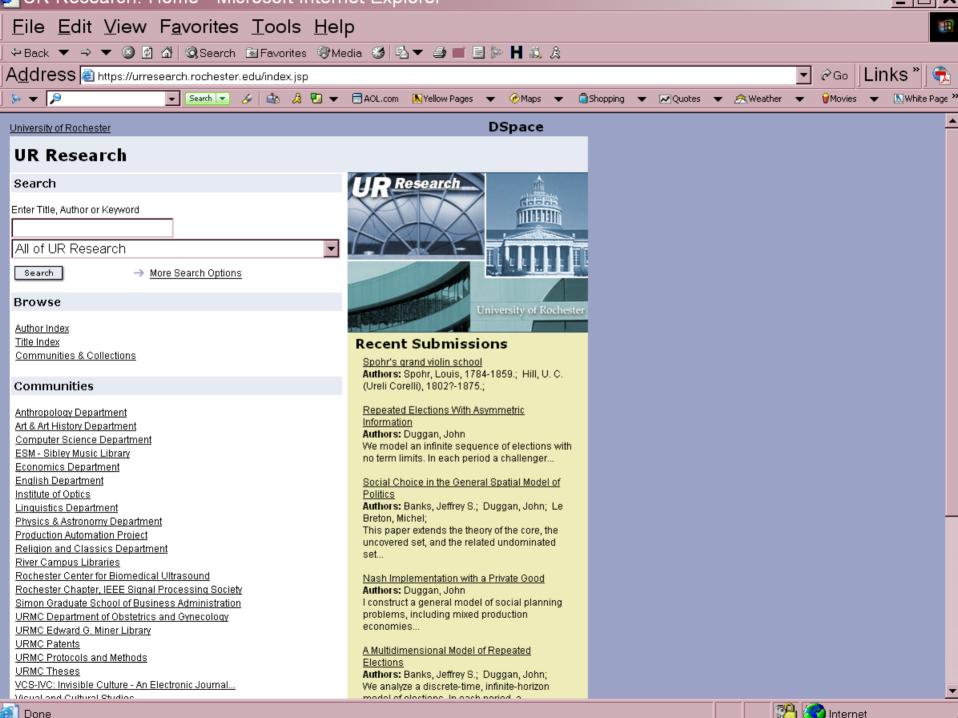


Where: Examples of Institutional Repositories

- · CUL'S OAR
 - http://dspace.library.comell.edu
- · University of Rochester
 - https://urresearch.rochester.edu
- · University of Toronto
 - https://tspace.library_utoronto.ca/
- · Computer Science Technical Reports
 - http://techreports.library.comell.edu









techreports.library.cornell.edu

Uses

- Computer Science Department Technical Reports
- Computing and Information Science Technical Reports
- Cornell Theory Center Technical Reports
- History and Theory of Machines and Mechanisms Technical Reports
- Library Papers and Preprints
- Watershed Management Papers
- · Over 2000 articles (majority in Computer Science)
 - Self selected, self managed, self contributed, sys admin by CUL
- Software system is an early version of DPubS





dspace.library.cornell.edu OAR

· Uses

- Bio and Environmental Engineering Student Projects
- Cornell University East Asia Papers
- Cornell University Graduate School
- Cornell University Library
- Cornell University Professional Degree Programs
- Internet-First University Press
- Electronic Theses and Dissertations
- Cornell Plantations
- East-Central Europe Sources
- · Self selected, self managed, sys admin by CUL
- Based on DSpace software





www.ar $oldsymbol{X}$ iv.org a disciplinary repository

· Uses

- Preprints in Physics
- Preprints in Computer Science
- Preprints in Mathematics
- Preprints in Quantitative Biology
- Over 320,000 objects (increasing by ~200 objects/day)
 - Self selected, self contributed, administered by CUL
 - 70% of articles ultimately appear in peer reviewed journals
- Based on arXiv software

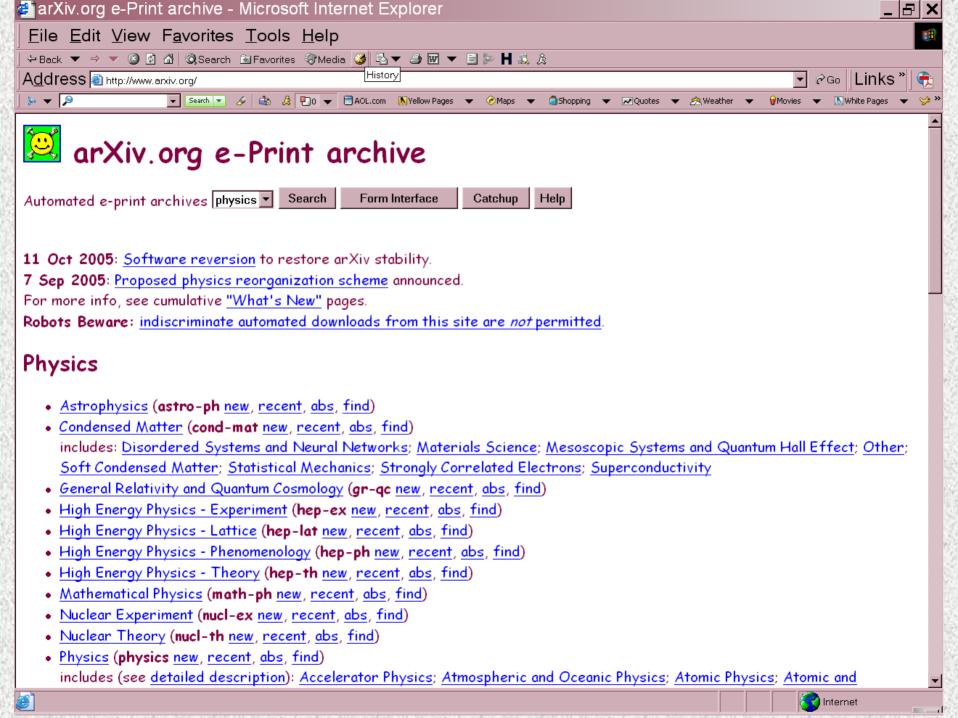


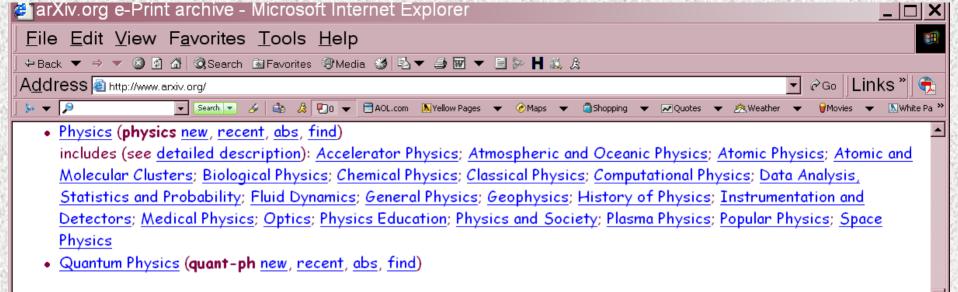


Why is arXiv so successful?

- Built on a solid model
 - Scientists COMMUNICATE
 - Physicists 'always' circulated their preprints
- Success breeds success
- · It's my research not my institution's
- Publisher acceptance vs. publisher hostility







Mathematics

Mathematics (math new, recent, abs, find)
 includes (see detailed description): Algebraic Geometry; Algebraic Topology; Analysis of PDEs; Category Theory;
 Classical Analysis and ODEs; Combinatorics; Commutative Algebra; Complex Variables; Differential Geometry; Dynamical Systems; Functional Analysis; General Mathematics; General Topology; Geometric Topology; Group Theory; History and Overview; K-Theory and Homology; Logic; Mathematical Physics; Metric Geometry; Number Theory; Numerical Analysis; Operator Algebras; Optimization and Control; Probability; Quantum Algebra; Representation Theory; Rings and Algebras; Spectral Theory; Statistics; Symplectic Geometry

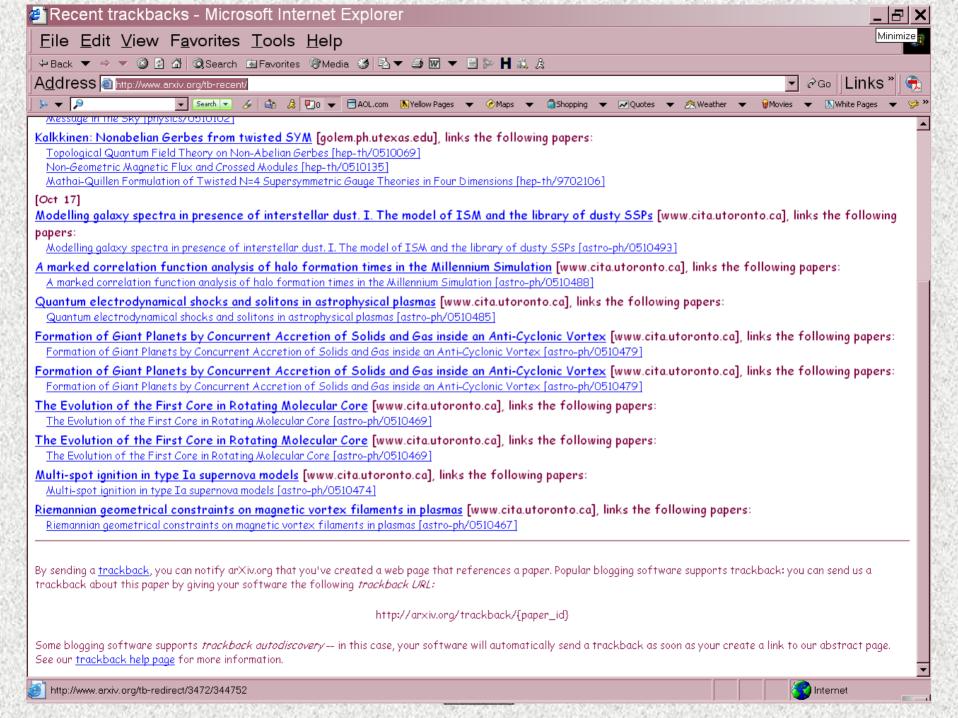
Nonlinear Sciences

Nonlinear Sciences (nlin new, recent, abs, find)
includes (see detailed description): Adaptation and Self-Organizing Systems; Cellular Automata and Lattice Gases;
Chaotic Dynamics; Exactly Solvable and Integrable Systems; Pattern Formation and Solitons

Computer Science



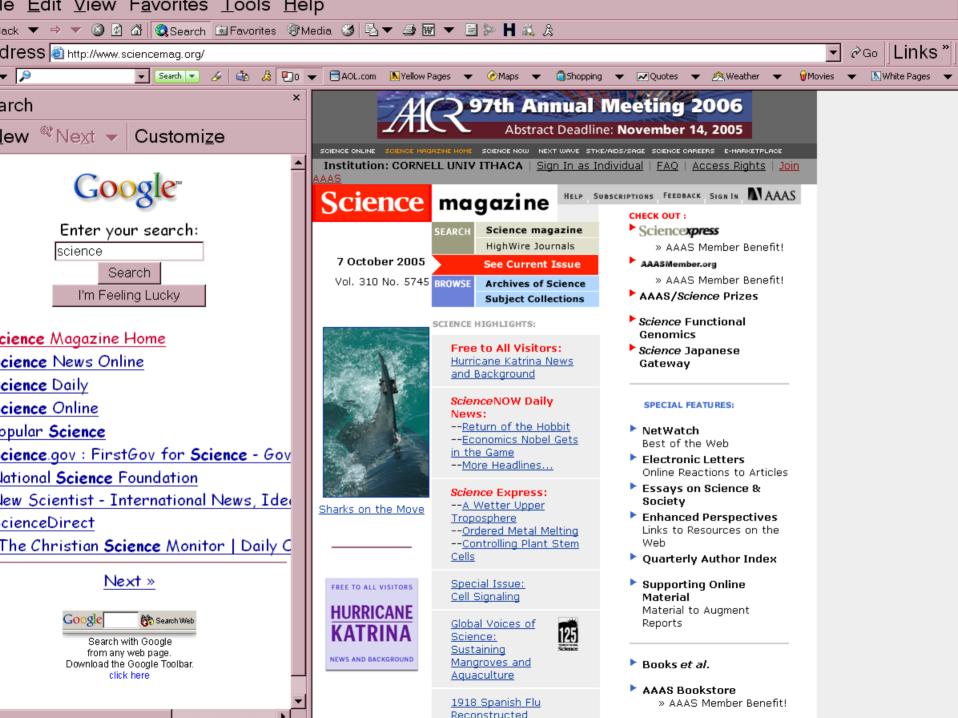


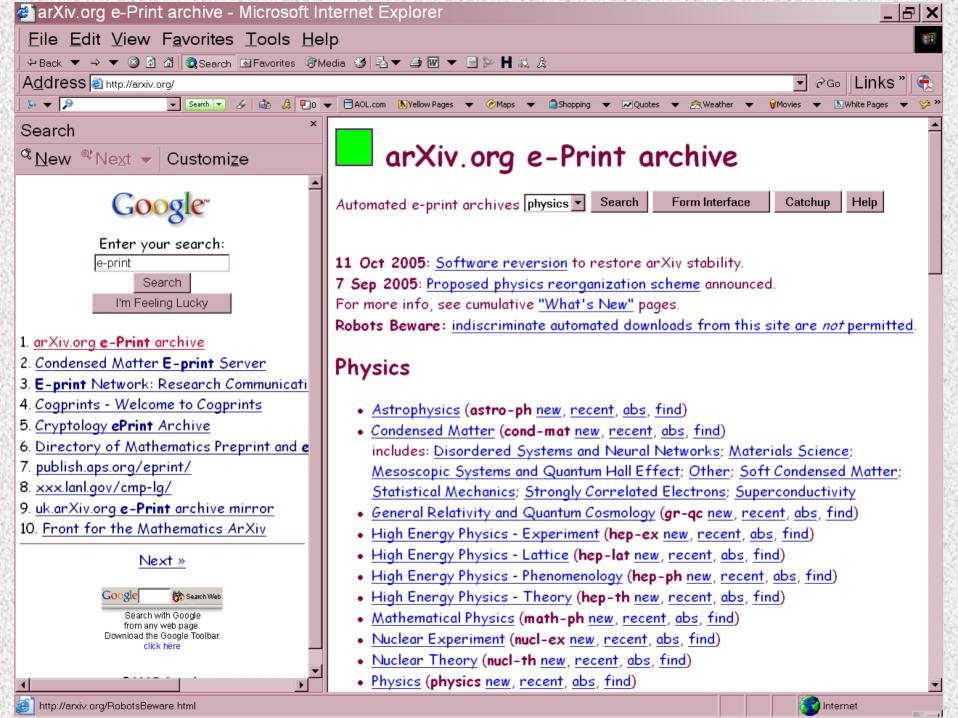


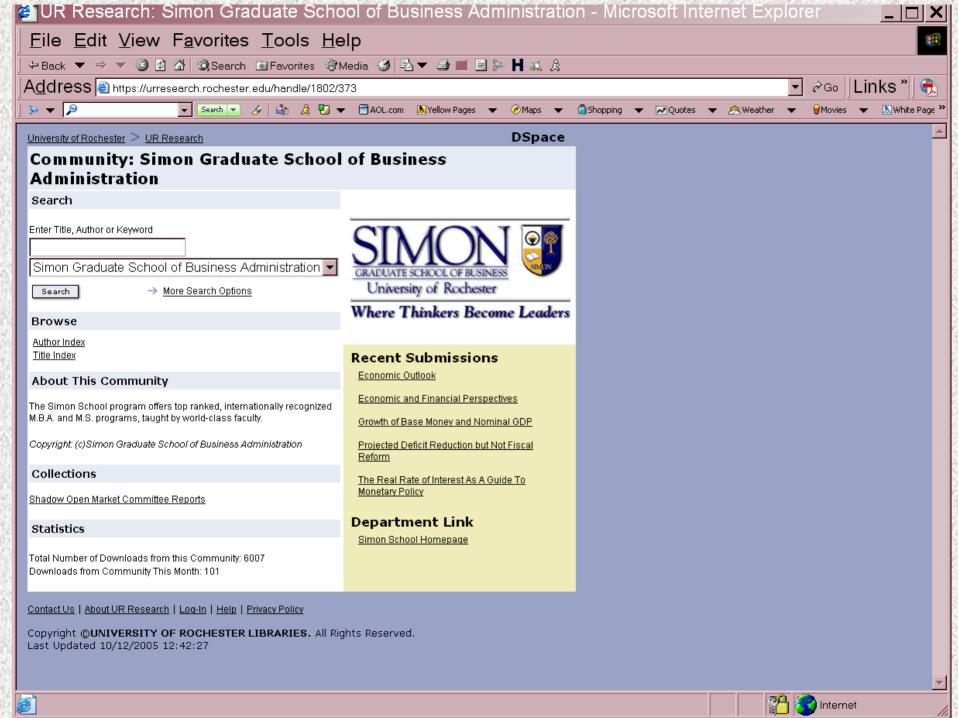


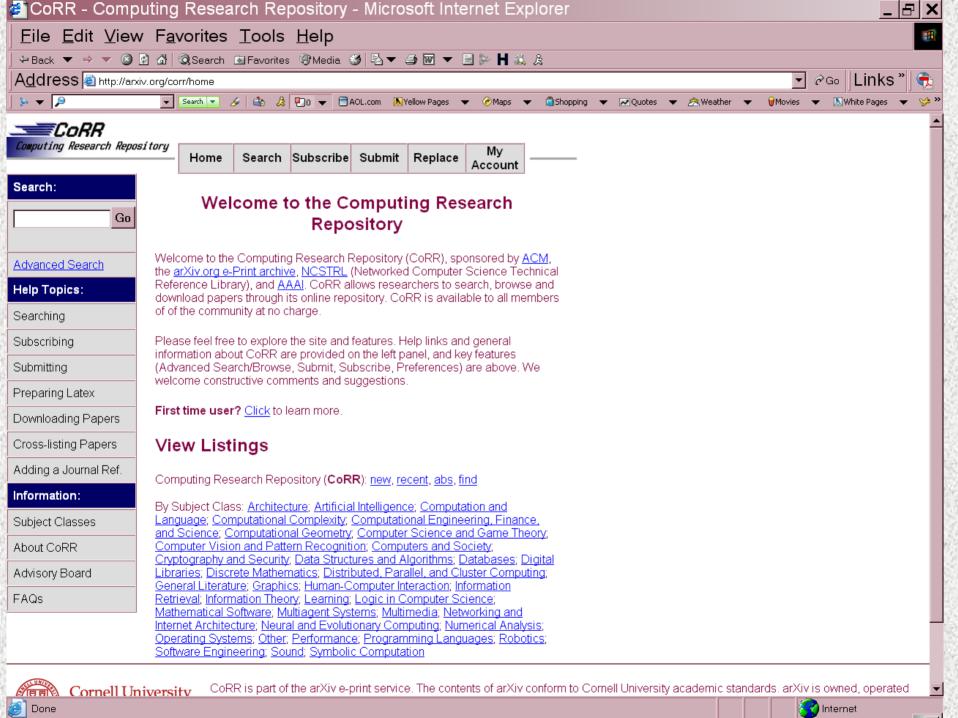
Success is...













Why: A Response to the Crisis in Scholarly Communication

- · The model
 - The institution pays the faculty and researchers
 - The faculty and researchers perform the work and write papers describing their results
 - They give their papers to the publishers
 - Sign over their copyrights
 - The institution's library buys the publication from the publisher, pays staff to process it, buys new shelves, binds, preserves, catalogs, ...
- · Journal costs (especially STEM) are increasing
- · Library budgets are decreasing

What's wrong with this picture?





Addressing the Crisis

- Encourage the use of digital and institutional repositories by outreach to Faculty
 - Departmental talks
 - Symposia and
- Policies enacted at Cornell
 - Archive policy for the Cornell Computer Science
 Department
 - Cornell University Faculty Senate Endorses Resolution on Open Access and Scholarly Communication
- · Education regarding 'green publishers'
 - http://www.sherpa.ac.uk/romeo.php
- Education regarding copyright
 - http://www.arl.org/spara/author/addendum.html





11/7/2005

Response

- Use Open access repositories as sources of articles for peer-reviewed journals (OA and not)
- Repositories can be part of overlay journals
 - http://www.lmcs-enfine.org/
- Will OA repositories replace peer-reviewed journals?
 - What about tenure review?
 - http://www.arxiv.org/tb-recent/





Encourage Use, but How?

- It takes a community...
 - Visit departments
 - Find a champion (or several)
- · Give Rewards
 - Tenure review?
- Discuss the advantages of Open Access Publishing (peer review and IR) with a few examples
 - Run the numbers (arXiv cost/article vs Elsevier)
- · Walk the Talk--Librarians should use the repository themselves. (see this paper and others at a local declaration and local declar

HLA

©
SOME RICHING RESERVED



What's stopping them?

- I forgot
- I'm too busy
- · It's too hard, takes too long, ...
- · "Publishers won't publish my
 - Book
 - Dissertation
 - Article, etc.
 - if it has already appeared on line. "
 - balderdash





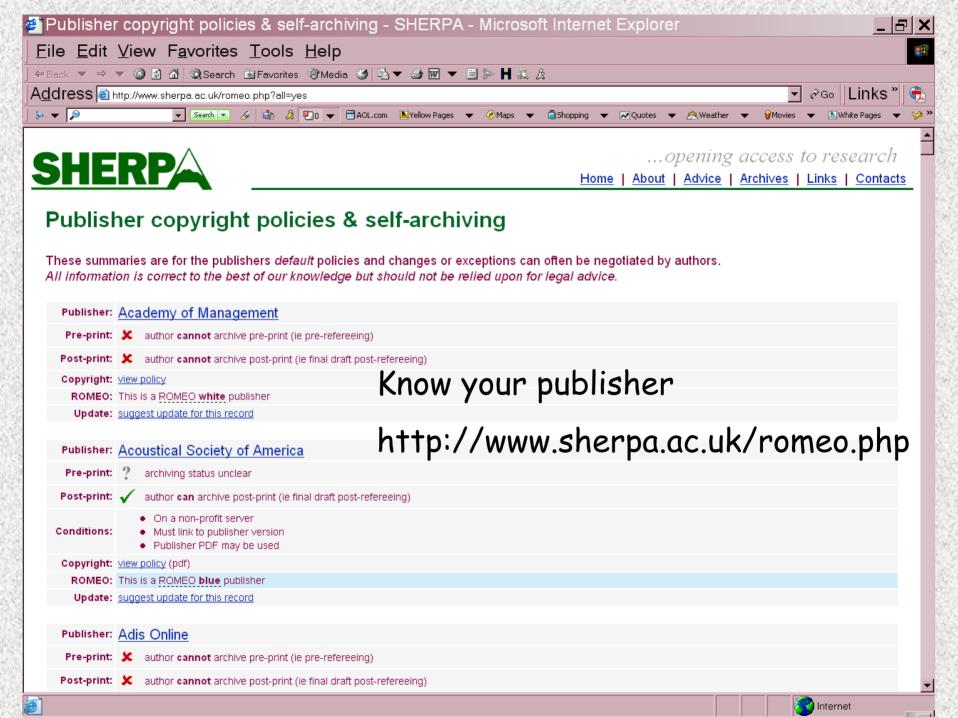
Campaign for use

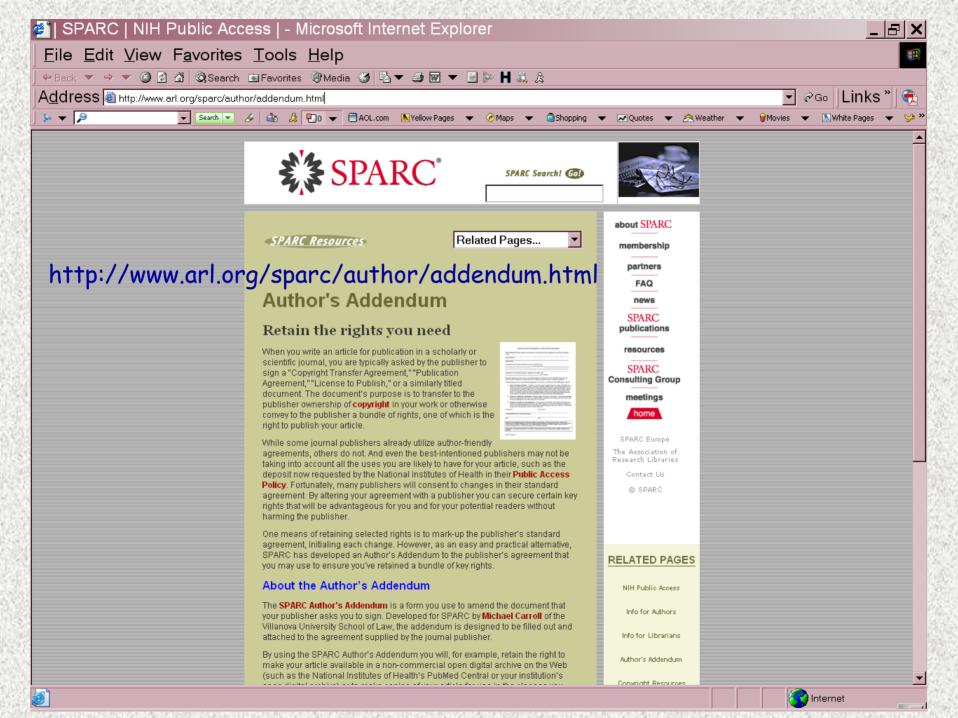
- Know Your Publisher
- Keep Your Copyright
- Save Your Stuff

Put it on a bookmark
Use it in your e-mail signature block
Talk it up

 Encourage university policies supporting OA (see handouts)









Costs and Sustainability

- DSpace—grant supported through 6/2006
 - Server: ~\$19,000
 - Server maintenance, housing, etc. ~\$3300/yr
 - Personnel: 0.2 FTE + ...
- arXiv—supported initially by CU Provost, now from library operating expenses
 - Server: ~\$10,600 (gift)
 - Server maintenance, housing, etc. ~\$3500/yr
 - Personnel: 2.75 FTE (2 administrators + .75 programmer
 + 2 student employees @ 10 hrs/wk each
- Sustainability
 - Move from project to program, but how?
 - What is the business model?
 - What are the measures of success?





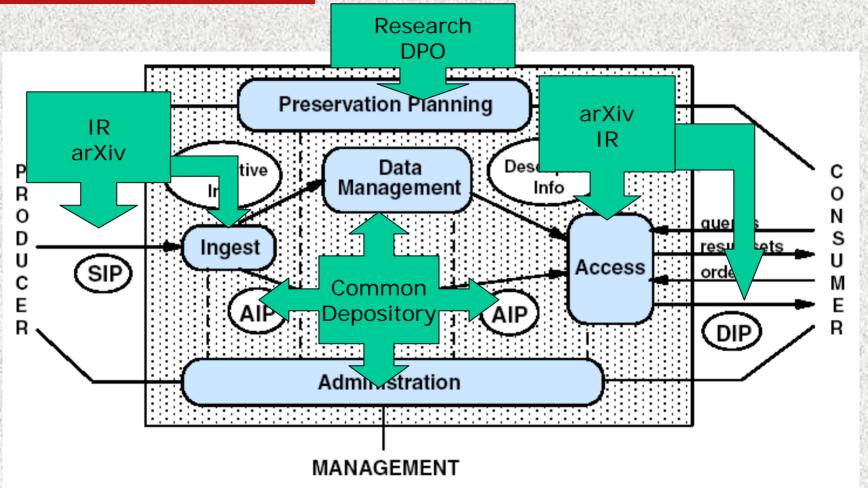
Digital Repositories and Digital Preservation

- What's the difference between storing in an IR vs digital preservation?
 - Faculty and researchers view digital repositories as "archives."
 - They use "archive" to mean "deposit in a repository"
 - The CS policy states: "All papers emanating from the department will be saved in a publicly accessible archive like
 - (Some) Librarians and (all?) archivists see things differently
 - Format changes, file type changes, bits rot, software obsolescence...





OAIS at Cornell



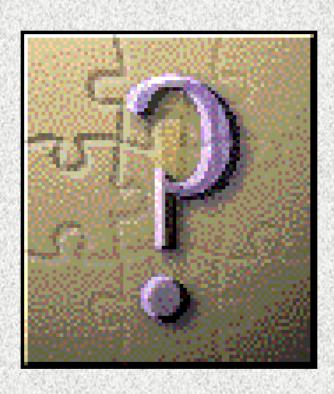


Things we've learned

- Learn patience
- Disciplinary Repositories more successful than Institutional ones
 - Faculty have greater allegiance to their disciplines than to to their institutions or departments
- Focus initial efforts in one or two areas
 - Avoid the shotgun approach
 - Start small
- Set administrative/library policies, roles and responsibilities
 - Collection curators, technical curators
- Match level of commitment with resources available
 - Start small
- Have communities set and maintain their own policies and develop selection criteria
- Multiple repositories increase exposure
- Exposure encourages contribution
- · Bubble up is better than trickle down
- Learn patience (the Zen of Institutional Repositories)



Questions





Cornell Faculty Senate resolution on scholarly publishing, passed 11 May 2005

Resolution from the University Faculty Library Board Concerning Scholarly Publishing

WHEREAS Cornell's longstanding commitment to the free and open publication, presentation and discussion of research advances the interests of the scholarly community, the faculty individually, and the public, and

WHEREAS certain publishers of scholarly journals continually raise their prices far above the level that could be reasonably justified by their costs, and

WHEREAS the activities of these publishers directly depend upon the continued participation of faculty at Cornell and similar institutions acting as editors, reviewers, and authors, and

WHEREAS a lasting solution to this problem requires not only interim measures but also a long range plan, and

WHEREAS publication in open access journals and repositories is an increasingly effective option for scholarly communication.

THEREFORE BE IT RESOLVED THAT

The Senate calls upon all faculty to become familiar with the pricing policies of journals in their specialty.1

The Senate strongly urges tenured faculty to cease supporting publishers who engage in exorbitant pricing, by not submitting papers to, or refereeing for, the journals sold by those publishers, and by resigning from their editorial boards if more reasonable pricing policies are not forthcoming.2

Reaffirming and broadening the proposals discussed during its meeting of December 17, 2003, the Senate strongly urges the University Library to negotiate vigorously with publishers who engage in exorbitant pricing and to reduce serial acquisitions from these publishers based on a reasonable measure of those subscriptions' relative importance to the collection, taking into account any particular needs of scholars in certain disciplinary areas.

The Senate strongly encourages all faculty, and especially tenured faculty, to consider publishing in open access, rather than restricted access, journals or in reasonably priced journals that make their contents openly accessible shortly after publication.3

The Senate strongly urges all faculty to negotiate with the journals in which they publish either to retain copyright rights and transfer only the right of first print and electronic publication, or to retain at a minimum the right of postprint archiving.4

The Senate strongly urges all faculty to deposit preprint or postprint copies of articles in an open access repository such as the Cornell University DSpace Repository or discipline-specific repositories such as arXiv.org.5

DISCUSSION

This matter has been before the Senate previously. On December 17, 2003, the Senators present unanimously supported the Cornell University Library's efforts to control spiraling acquisition costs by tough negotiations with certain journal publishers who were exploiting their market power.





Archive policy for the Cornell Computer Science Department

On December 1, 2004, the Cornell Computer Science Faculty by consensus adopted the following policy, which was proposed by Bill Arms, Joe Halpern and Steve Vavasis.

All papers emanating from the department will be saved in a publicly accessible archive like <u>arxiv.org</u>.

This document will attempt to explain the rationale and implementation of the policy.

Rationale for the policy

A crisis has been evolving in the past few years in the realm of scholarly publishing because commercial journals have raised their prices substantially without a proportional benefit to the community of authors or readers. For example, the EMPS (Engineering, Math and Physical Sciences) library at Cornell has seen a 9% subscription increase in just the past year. The worst offender seems to be Elsevier, which publishes many CS journals.

A second looming concern with scholarly publishing is that commercial publishers are using pricing policies to push libraries into switching to all-electronic subscription. All-electronic subscription gives the commercial publisher unprecedented control over who can read articles and for what purposes those articles are used. Furthermore, an electronic subscription means that the publisher expands its role to become also the archivist of the material. There is no reason to believe that a company like Elsevier is qualified to usurp the role traditionally filled by libraries as the archivist of scholarly work over a period of decades or centuries. For more information about the problems faced by university libraries, please visit the home.page of the SPARC project of the Association of Research Libraries.

An obvious solution to these problems is for the academic community as a whole to create its own archive under the control of scholars rather than a corporate board of directors. This is the goal behind arxiv.org. We believe that all academics ought to include their publications in this kind of archive. Therefore, we are establishing this as a departmental policy. We would like to establish it as a policy for the whole world, but we have to start somewhere!

Naturally, a member of the department could easily follow this policy on his or her own initiative without the existence of a departmentwide policy. Indeed, several of us already archive our papers as a matter of course because archiving brings several benefits to the author including enhanced visibility of the result and proof of precedence of discovery. But we believe there are three reasons why it is useful to make archiving an official policy of the department.

- 1. By making it a policy, we are making a public statement in favor of open archiving.
- 2. There is clearly a snowball effect at work: the more computer scientists who are the more useful the archive becomes, and hence more people will