

Slide 1 - Title

I'm here to talk with you about institutional repositories , scholarly communication and using the University of Oregon's experiences as the focal point for this review.

Couple caveats:

- Selected content for this audience
- I'm not an expert on IRs or scholarly communication – just on the UO experience
 - John Cox's presentation on scholarly communication was excellent and I encourage you to read his paper when it becomes available if you missed his presentation

Slide 2 - Background

- **journal prices rising faster than inflation for the past 35 years.**
four times faster than inflation in the last 10 years
- libraries dramatically increased expenditures on electronic resources
 - Electronic journals account for the about 92% of this
 - Electronic journals now account for 26% of a library's overall serials expenditures (ARL 2004)
- New pricing models are taking their toll.
 - bundling access to electronic content
 - forcing libraries to maintain print subscriptions in order to subscribe to the electronic
 - temporary embargoes on full-text access to specific issues
- Scholarly output also increasing, both monographs and serials

Slide 3 – Result?

Formally-published scholarly output is more expensive to purchase or license, there is an increasing amount of it, and libraries are able to provide access to an ever smaller percentage of the total. We're all familiar with this scenario.

Even barring price increases we would be unable to provide access to increasing amount of content.

Slide 4 - Responses

The university of Oregon's responses will probably look very familiar to you.

[First bullet] In 1992/93, cut \$350,000. In 1996 targeted \$500,000 In 2000, another large, serials cancellation project again in 2003 and 2004. acknowledged that the process is continuous and renamed it "*serials review*"

[2bd bullet] March of 2001, discussion with departmental library representatives on e-only access for some materials. Ongoing discussions with different faculty groups outlining a variety of scholarly communication issues and challenges.

May 2004, a presentation by a Professor at Oregon State University and Editor of an academic society journal to discuss the scholarly communication crisis from the publisher's standpoint

[3rd bullet] Consortial purchases of electronic resources (Orbis/Cascade Alliance)

[4th bullet] UO/OSUO Shared Collections Initiative

[5th bullet] Cataloging of all e-journals and other licensed e-resources done routinely

[6th bullet] Cataloging DOAJ titles

[7th bullet] Brought up SFX August 2003; implementing ERM

Slide 5 – IR and open access

IR: digital collections capturing and preserving the intellectual output of a single or multi-university community

IR definition from SPARC position paper entitled “The Case for Institutional Repositories,”

Open access: allows all members of society to freely access relevant cultural and scientific achievements, in particular by encouraging the free (online) availability of such information

Open access definition taken from the Wikipedia (May 2005)

So, that’s the context.

Slide 6 - SPARC

One important player in this area is SPARC.

- Scholarly Publishing and Academic Resources Coalition (SPARC) begun by Association of Research Libraries (ARL) in 1997
- as a “**constructive response to market dysfunctions in the scholarly communication system.**”

Slide 7 – Case for Institutional Repositories

In particular, we found SPARC's Case for Institutional Repositories Position Paper to be **an effective statement** in support of such endeavors,

Slide 8 - Rationale

postulated that **“Institutional repositories can provide an immediate and valuable complement to the existing scholarly publishing model, while stimulating innovation in a new disaggregated publishing structure that will evolve and improve over time.”**

That was the rather naive hope.

IRs could increase institutional visibility and prestige, “by capturing, preserving, and disseminating a university's collective intellectual capital,

archiving output not otherwise captured – the grey literature.

Slide 9 - ACRL

We found the work that ACRL was doing on scholarly communication particularly helpful.

One of the strategies that they identify in this report is the development of institutional repositories

Slide 10 – Reforming Scholarly Communication

Some of the Principles Supported in the ACRL report include:

- the **broadest possible access** to published research and other scholarly writings
- **increased control by scholars and the academy** over the system of scholarly publishing
- **competitive markets** for scholarly information
- a **diversified publishing industry** open access to scholarship
- quality assurance in publishing through peer review
- fair use of copyrighted information for educational and research purposes
- extension of public domain information
- preservation of scholarly information for long-term future use

Slide 11 – Issues in Scholarly Communication

We continue to track on scholarly communication developments so that we can keep the campus informed and plan how to market our IR and the service it provides effectively. **This is a growing focus for our IR.**

Slide 12 – Getting started

began as a task force or library initiative in January 2003.

small group: me, the Director of the Center for Educational Technologies, the Head of Reference, the Map Librarian, and the University Archivist,

charged to:

- Gather background information on emerging institutional repositories;
- survey campus for potential pilot project;
- move forward on any opportunity for implementation; and

We began by researching the issues and the IR movement that I've just outlined for you.

Slide 13 – Investigate software

looking for software that would meet the following criteria:

- it would be **easy to implement**
- It would be **OAI-compliant** (Open Archives Initiative) so that we would be able to join up with other repositories
- It would **not need a major investment of hardware** to support it
- The **software would be cheap**
- it would have been **tested by other institutions**

We chose **DSpace** because we felt it best met all of these **criteria** at the time.

As an **open source application**, it depends upon its users to **continue developing it**.

The community of users is active and helpful, sharing technical, policy, and procedural solutions freely.

Slide 14 – Scholars Bank

- Scholars' Bank is the name of the UO's institutional repository
- **brought up in test in May of 2003**
- The largest spurt of growth has taken place in the last six months. We've had a 255% increase in content and a 215% increase in use in the past six months.
- *As of May 15, 2005, it had 782 submissions in 23 different communities and 42 different collections.*

Our institutional repository is:

- digital archive of the output of the **faculty, researchers, staff, and students**
- accessible both within and outside of the institution
- **Our definition has expanded so that materials deposited in the archive are either themselves scholarly or else support the University's scholarly mission.** That has enabled us to go after a broader range of content, much of which would not otherwise be archived – or not effectively and reliably archived.

Our vision for Scholars' Bank has changed over time

- *we came to realize that it is not a panacea for the scholarly communication crisis but could become a part of a new model.*
- Our initial vision was to have a digital archive for **faculty where they would submit their own work directly**, without any mediation or intervention by the Libraries.
- The vision of self-submission has not yet been borne out, however, as well **over 95% of the current content was added by library staff** with the authors' permission.

Slide 15 - local policy framework

In setting up an IR, it's necessary to establish local policies. Policies evolved over time and continue to evolve based on what we encounter as we work with different academic departments or campus groups.

- what our policies on submission of material to the archive would be (individuals, communities, authorizations, review)
- metadata standards (mapped to Dublin Core – input standards for different fields, controlled vocabulary, quality review)
- what our institutional commitment to the archive needed to be (long-term, short-term, in perpetuity, preservation of bitstream or full content, etc.)
- what would be our stand on copyright and licensing
- who within the library would be responsible for setting up and coordinating the archive and what role would campus groups play

Slide 16 – Develop local context

We've tried to develop our local context.

Part of what we did at the beginning was to take a look at what was going on across campus – which departments or individual scholars were already archiving their scholarly output in some fashion and how were they doing that.

As we spread out and talked to individuals and groups, we began to get a sense of what they considered the significant issues. We have developed a series of web pages that provide more information about the service.

We are continually adding to and refining the local context.

Slide 17 – Local contextual wrapper

In our local context, we try to:

- Educate on issues and link to broader movement
- Provide overview of services using meaningful language
- Personalize the information
- Answer questions of local interest
- Provide enough detail
- Build in redundancy (so the same information is provided in multiple places – to account for different entry points into the archive)

Slide 18 – Educate on issues of concern

- We have utilized the ACRL, ARL, and other groups' positions on scholarly communication
- tried to condense it down to a few essential points
- give people the option of clicking to another page for more detailed information.

Slide 19 - Scholarly Communication Crisis : Background

Short, simple explanations backed up with more complete explanations and extensive lists of resources that individuals may consult for further information. *On this page, we tried to find statements from organizations that would have meaning for faculty, as well as library organizations;*

such as the AAU, AAUP, NIH. I just added the Cornell Faculty Senate Resolution of May 11 calling for faculty to make changes in the way they publish.

Slide 20 - Link to Broader Movement

- present information to persuade faculty and campus administrators to make use of the repository.
- UO News Items are articles that have appeared about Scholars' Bank within the campus publications that we have archived
 - incestuous relationship.

Slide 21 - Wellcome Trust and Open Access

- attempt to show our connection to a broader movement, as well as show the local context.
- we point them to positions such as that articulated by the Wellcome Trust.

Slide 22 – Berlin Declaration

Or the Berlin Declaration on Open Access

Slide 23 - Provide Overview of Services

overview of what Scholars' Bank is and what they can expect from it

Slide 24 - Personalize the Information

We personalize the information for specific communities, trying to make them understand the benefits of the archive for them personally.

Slide 25 – What’s in it for them?

We attempt to personalize the information
what is in it for them, personally.

Slide 26 – Answer questions of local interest

As we talk to individual faculty members, departments, the University Library Committee, and the campus Deans, we hear many of the same questions repeated. Addressing some of these questions up front and making the answers available in an FAQ has allowed us to offer a consistent face to the different groups we work with.

Slide 27 - Copyright Concerns

One of the biggest concerns for our faculty is that of copyright and whether posting their research in our repository will jeopardize their relationships with their publishers. We try to address their concern directly and point them to resources that will help them feel on firmer ground about this.

Slide 28 - Sherpa

SHERPA, part of the UK's Joint Information Systems Committee's work, has provided a site which lists the copyright policies of many commercial publishers.

Authors can use the site to find a summary of permissions that are normally given as part of each publisher's copyright transfer agreement.

By pointing people to this site, we are trying to inform them that they may already have the right to post pre-publication and even post-publication digital versions of their work, even without negotiating a special arrangement with their publishers. *Every time I've made a presentation to a group of faculty, they always write down this information.*

Slide 29 – Sample Policy on copyright

This is an example of a sample copyright policy for a major publisher listed in Sherpa.

This one for Elsevier gives authors the right to post pre and post-prints of their work, with certain conditions.

Slide 30 – Provide Enough Detail

We try to provide detail for those who want it, but without making everyone dig through it if they don't want it.

even though only a very small number of individuals contribute their own work (most opt to have the library do it for them), we address the nitty gritty details for submitting to SB.

Slide 31 – License agreement

One of the things that sometimes frightens an author is the license.

- Our license agreement is designed to grant us the right to make the submission publicly available over the web and to allow us to migrate files to another format if needed in order to preserve it.
- helpful for people to be able to see the full text of the license agreement
- We also explain in plain English what it is we are asking for

Slide 32 - **Structure of Scholars' Bank**

Structure specific to DSpace. Other packages also handle this type of content but this is the one we chose.

The structure of SB is this:

- top level: communities
- Sometimes sub-communities
- within communities there are collections
- within collections there are individual titles
- titles can be comprised of single or multiple files

Communities, collections, and items all have a unique and stable handle which can be used like a URL and cited.

Slide 33 – Sample community

The Economics Department is one of our more active communities with 123 titles.

We discuss with a potential community:

- types of materials they want to submit,

- average size of individual files,
- whether or not they want restrictions on who can access the materials within it
- who can submit to their collections
- whether there are multiple collections needed for the community,
- who will be handling the actual submissions (usually us)

A process that supports peer review is possible within DSpace.

- Even though they express concern about the quality, no one yet willing to undertake the review process – they rely on us to doublecheck with them

Slide 34 – Collections within communities

- community can have multiple separate collections with different guidelines and standards for each one.
- Each collection has its own page, and its own parameters, workflow, and authorizations.
- Each collection can have its own branding, as you can see here with the logo for Renaissance Editions. **(first animation)**
- Recent submissions to a collection appear off to the right. **(2nd animation)**
- Email notifications from particular collections, communities or the entire archive are possible. **(3rd animation)**

In some cases, the community and collection name may be the same. The structure reflects the assumptions that the original developers had about how such an archive would be used.

Slide 35 - **Individual Titles within Collections**

Individual titles can be searched from within a community, a collection, or across the entire repository, as can authors, keywords, or other metadata.

We run an indexing program nightly so that Items and the content of them are keyword searchable, if the files were properly formatted, even PDFs. Items scanned from hard copy are not keyword searchable, unless OCR has been done as part of the scanning process.

Slide 36 - **Individual Files Make Up Titles**

Each title may consist of one or multiple items, various file formats

Slide 37 - **Items Can Be in Multiple Collections**

The software allows the same item to appear in multiple collections. The URL remains the same but the item can appear and be searched within multiple collections.

helpful for faculty who want to appear as part of a larger community but who also want to highlight their own individual work. (**animation**)

Slide 38 – **Mediated Submissions**

One of the features of the DSpace software is that it attaches a license file with each submission.

- software was designed for self-submission
 - When we handle the submission for an author, we get **email or paper agreement to the license which then convert to a text file that we link to the submission in the archive**

- view from within the administrative structure for a particular submission.
- **(1st animation)** top license file automatically generated by the system when the item was submitted and carries the email identification of the person who submitted it
- **(2nd animation)** second license file contains the converted text file,
- Neither license file is displayed to the public but it records the permission.
 - labor-intensive, but we consider this important step until the software develops to be more accommodating for different submission models
- In the case of a journal, we archive the editor's agreement with the first issue we archive.

Slide 39 – Serials in Scholars' Bank

There are a number of issues relating to serial publications within an institutional repository. These are the issues that we have encountered so far at the University of Oregon.

I'll be coming back to these and the other points in this slide as we look at specific serial issues.

Slide 40 – Locating appropriate content

with serials, as with everything else, our first step was locating appropriate content on the campus.

In the serials realm, campuses have a proliferation of content that could benefit from being organized, archived, and disseminated through an institutional repository

Slide 41 - CultureWork

This is our first serial publication

- In many respects it was the ideal publication.
- editors contacted us first.
- already published electronically
- presented consistently like a normal serial (it had distinct issues with consistent numbering.)

Slide 42 - **Getting permission to archive**

Once appropriate content is located, you must then get permission to make it available electronically and to convince the editors or authors that you are trustworthy.

This is usually not difficult, because the idea of the library assuming responsibility for long-term archiving is appealing.

- Offer to do whatever they want: if they want the archive but have no time, be prepared to do all the work for them. We have considered this necessary to get a critical mass of content. *Later on, we will be encouraging publications to do their own submission, after they've really had the value of the archive proven to them.*

Slide 43 - **Conversion of Existing Files**

- Some of the content is only distributed in paper, although usually someone on campus has the electronic files that were used to generate the print.
- Sometimes, publications have been made available electronically as Web pages.
 - we convert the Web pages to PDF. We do this for several reasons: 1) it stabilizes the content; 2) it presents better within the archive.

This publication existed in html and is archived on the web that way. We have harvested and saved files in PDF. All of the links work internally within the document if we have harvested them.

Slide 44 – Complexities of harvesting

Harvesting html-based publications presents a number of unique challenges. **One of the challenges we have dealt with is how deeply to follow links and harvest the content of the linked pages.**

One of the campus serial publications we have archived in our institutional repository is the biweekly campus newsletter, Inside Oregon.

Slide 45 - HTML Archives and Multiple Pages How Deep do You Go?

Inside Oregon is published as a series of linked web pages.

- We do not follow every possible link from every page.
- we harvest the links to individual articles within the publication and all links one level down from each article page.
- The other links remain in the final document and are active, as long as the html address continues to exist.
- normally works reasonably well, allowing us to capture the primary content of an issue without ending up like the Internet Archive and trying to archive the web.

Slide 46 – Capturing links

However, A recent issue of the newsletter Inside Oregon that focused on blogs and their growing use on campus stretched this model. On this issue, we harvested 240 pages. A normal issue is closer to 50 pages of harvested content.

What constitutes the publication? How deep do you go? we have made these decisions on a case-by-case basis in consultation with the publishers or editors.

Slide 47 - **Broken or Inaccurate Links**

Another challenge in defining the nature of a publication has arisen when we are harvesting html-based publications that first appeared years ago.

- **If the issue is not archived and stabilized when it's published, links may break or contain different content.** The content of back issues of html-based online serials often has not remained static.
- Harvesting years after the fact produces an inaccurate picture of what the issue was at the time of publication.
 - In this example, we were harvesting a web-based publication that first appeared in 1995 issue
 - one of the links was still active but went to a page that had been revised in 2005.
 - We have drawn the line at researching the earlier content in the Internet Archive's Wayback Machine and trying to capture what it was at the time. We've compromised on this and might revise this approach if we get negative feedback from our users.

Slide 48 - **Logical or Useful Presentation**

Another challenge will resonate especially with this audience – the need to educate campus publishers about how to create effective serial publications

The concept that having volume or issue numbering or dates to facilitate an orderly display and consistent access is revolutionary to many of these publishers.

Other challenges relate to the software. DSpace software was developed to handle articles very well, but it was not designed expressly to handle entire serial titles. The inconsistencies of the original publications in designating issues, coupled with the limitations of the software have presented interesting challenges.

- **(animation)** We experimented with a couple of ways of dealing with the metadata to get a useful display.
- **Unlike an online catalog, where issue numbers are captured in holdings or check-in records**, in this software each issue has a separate record and it's necessary to capture issue numbers somewhere in that record.
 - **We chose to include it in the title field because it helps with orderly sorting.**

Slide 49 - Numbering or the Lack Thereof

This publication was particularly problematic.

- Earlier issues were numbered. but the numbering jumped from Vol 8 number 1 to Volume 11 number 1.
- Then all numbering stopped and they went to seasonal designations without any month of publication being obvious.
- **(animation)** problem was exacerbated by the software's inability to sort properly by issue number, putting 11 before 1 in the listing.

Clearly, this was not a useful display.

Slide 50 - Chronological Displays of Issues

After consulting with my Head of Serials Cataloging, Mary Greci, I arrived at a way to get issues to display in chronological order by inserting the year.

(animation) There is still the problem of how to handle seasonal designations, which will require supplying numbers to help sort out the display.

Slide 51 - Chronological Displays of Issues

This newsletter, *From the Center*, lacked any numbering and used only seasonal designations.

- **(animation)** we supplied a number for each issue in that calendar year it represented.

Slide 52 – Actual issue of newsletter

Slide 53 – Appropriate level of metadata

Each issue has its own URL and its own metadata record.

- **How far down do we want to go in providing descriptive metadata** for each issue of a serial publication?
- Do we want to provide indexing for individual issue titles and for authors in each issue?
- **(animation)** we did provide access to individual titles within an issue-specific record.
- However, the issue full-text searchable. No real need for deeper indexing.
- **We're experimenting** and are letting ourselves be guided by what each community wants for their publication.

Slide 54 – Links from the catalog

We have decided to catalog all of the campus **serial** publications that we are capturing and to provide the link to the Scholars' Bank version, (**animation**) as well as the original site archive.

Slide 55 – Cataloging Issues

we expect the catalogers to reference the original publication or our capture of the content for actual descriptive information, rather than rely on the descriptive information that we supply about each issue in the SB record.

- we provide notes on the home page for the title within Scholars' Bank that will alert catalogers to look elsewhere.

I'm going to show you some examples from some of our other collections.

Slide 56 – Current efforts

- Departments, programs, institutes
- Hosting ejournals, newsletters, web sites
- Electronic theses
- Individual class archives
- Undergraduate Research Award program
- Individual faculty sites
- Graduate student community

Slide 57 – Electronic theses

We have been working for some time with our Graduate School to get their support for ETDs. That process is ongoing.

- We have managed to acquire some individual dissertations, without the active support of the Graduate School.
- **individual academic departments or programs are eager** to have their students archive their theses and dissertations.

Slide 58 – Individual class archives

archiving the student research for a particular class.

As we capture more and more student work, we are **training up a new generation of faculty** to expect the open access model to be in place.

Slide 59 – Undergrad Library Research Award

establishment of an award program for high-quality undergraduate research. Beginning in 2004-5, each student who submitted a work for review **had to agree to have it archived in Scholars' Bank**. The review committee consisted of teaching faculty and librarians. A small scholarship was given to each of the four winning students.

The caliber of the student work in Scholars' Bank is very high and we have found that these items are among the most frequently viewed.

Slide 60 – Individual faculty collections

Establishing collections for individual faculty members is a new endeavor. Only two faculty have taken advantage of it so far but interest in it is growing. It has also spawned requests from individual graduate students

for such a collection. We have decided that graduate students must be willing to submit their own work. For faculty, we will handle all submissions for them, if they prefer it be done that way. But we want to begin moving away from the mediated model and so are expecting students to do the legwork for themselves.

Slide 61 – How are we doing?

Since we're an ARL library, we're used to measuring success in terms of numbers.

- one measure of success is by the amount of content that we have in our archives.
- **can't all be about numbers.**
 - As we collectively acquire more content in archives that are built on the same open access standards, these materials can be indexed and harvested – by Google, Google Scholar, OIAster, and other indexers.
 - making a greater volume of scholarly content available to the world at large
 - we must also measure our success by the degree to which we have opened discussions, raised awareness, and captured previously lost or hidden materials on our campuses
 - materials such as student work, presentations, working papers, conference papers, newsletters, electronic archival materials, and the like. **Collectively, these materials form a rich resource for scholarly research. Not just the record of research output**

Slide 62 - MIT

MIT. One of the original DSpace developers, MIT had for a long time had difficulty acquiring a lot of content. With 14332 titles, they seem to have overcome the original problem they were having.

Slide 63 - CDL

The California Digital Library, which uses different software, has 6900 titles from nine UC campuses.

Slide 64

University of Toronto - 3041 titles

Slide 65

University of Glasgow, one of the leaders in revising the software and increasing its flexibility and functionality, has 215 titles in its Dspace archive and another 1200 or so in e-prints archive

Slide 66

The **University of Rochester**, one of the institutions that has strong campus administrative support, has 1278 titles

Slide 67

University of Edinburgh has 462 titles

Slide 68

University of Kansas, which has a well developed site with strong and clear policies has 230 titles

Slide 69

University of Arizona is just starting up and has 49 titles.

Slide 70 – UO’s Scholars’ Bank

Comparatively, by the numbers, the **University of Oregon** isn’t doing badly. We’ve acquired around 782 items by May 15 2005, after roughly a year of extensive marketing and increasing publicity.

without having any staff devoted exclusively to the effort; all work has been absorbed by existing staff.

The numbers in other archives around the world vary considerably,

- we are all learning more about our campuses,
- gaining a better understanding of how faculty and other researchers operate
- what they **value** and **fear**,
- we are raising awareness among our primary clientele of the changes taking place in scholarly communication.

Slide 71 – Are we changing scholarly communication?

- we haven’t ignited a revolution within the faculty
- they are beginning to see potential
- Discussions are starting to occur in some academic departments about allowing works deposited in such archives to count towards tenure and promotion.
- There is still substantial resistance among the old guard campus administrators and change is happening slowly.
- we are expanding access to grey literature and under-collected research and primary source materials.
- Complements rather than supplants the traditional publishing model

- Our archive contains:
 - output from academic departments, programs, or institutes
 - journals, newsletters, or web sites
 - faculty working papers and presentations
 - electronic theses or student papers
 - archives for the output of specific classes
 - graduate student communities or research interest groups
 - campus administrative or planning documents
 - finding aids
 - campus electronic records.

Slide 72 – Next steps at the UO

What are our **next steps**?

- **Continue to acquire content.** This means responding quickly to any expressions of interest from faculty, staff, and students of the University of Oregon.
- **Continue to market,** which means finding ways of keeping the archive in the spotlight through articles in campus newsletters, special programs, and routine library instruction.
- **Align more closely with instructional programs,** including developing more archives for specific classes.
- **Develop self-submission model among some communities,** which is going to be key for our ability to sustain the archive.
- Establish an **advisory group consisting** of faculty, staff, and students who can help us identify likely contributors, understand better what people want and need from the archive and help us be more responsive.
- **Develop searching guides and targeted search interfaces** that can be utilized by a particular class or department.
- **Contribute further to software development,** so that when we identify aspects of the archive's structure that interferes with or impedes use and submission to the archive, we can contribute to the solutions for other DSpace users.
- **Refine use statistics**

Slide 73 – Contact Information for Scholars’ Bank

I am happy to talk with anyone who is trying to set up such an archive or who wants to know more about any aspect of it. If your questions are too technical, I’ll pass you on to other people in the UO libraries who can help.

The presentation is available in its full form at the handle listed on the front of your handout.