How to Implement an Institutional Repository

A NASIG 2006 Pre-Conference
May 4, 2006

by

Carol Hixson
Head, Metadata and Digital Library Services
http://libweb.uoregon.edu/catdept/home/
University of Oregon Libraries

https://scholarsbank.uoregon.edu/dspace/handle/1794/2484
Contact information

Carol Hixson
Head, Metadata and Digital Library Services
University of Oregon Libraries
chixson@uoregon.edu
(541) 346-3064
How to Implement an Institutional Repository: Part I

Rationale and Planning
Rationale and planning

- Why do it?
- What’s your timeline?
- Who will be involved?
- Can you afford it?
- How will you measure success?
- How flexible is your vision?
Definition

- Digital collections capturing and preserving the intellectual output of a single or multi-group community
- Set of services for the management and dissemination of digital materials
- Not just for universities!
Scholarly communication crisis

- Prices rising faster than inflation
- Movement from paper to electronic
- New pricing and access models for electronic content
- Scholarly output increasing
- Libraries able to provide access to smaller percentage of total scholarly output
Scholarly communication crisis

- Serials cancellations
- Campus discussions on scholarly communication
- Consortial purchases
- Broader sharing of collections
- Cataloging of e-journals
- Promotion of open-access journals
- New management tools (SFX, ERM, etc.)
Change scholarly communication

- How to do this?
- IRs and Open Access
  - 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
Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities

Preface

The Internet has fundamentally changed the practical and economic realities of distributing scientific knowledge and cultural heritage. For the first time ever, the Internet now offers the chance to constitute a global and interactive representation of human knowledge, including cultural heritage and the guarantee of worldwide access.

We, the undersigned, feel obliged to address the challenges of the Internet as an emerging functional medium for distributing knowledge. Obviously, these developments will be able to significantly modify the nature of scientific publishing as well as the existing system of quality assurance.
What is SPARC?

SPARC®, the Scholarly Publishing and Academic Resources Coalition, is an alliance of universities, research libraries, and organizations built as a constructive response to market dysfunctions in the scholarly communication system. These dysfunctions have reduced dissemination of scholarship and crippled libraries. SPARC serves as a catalyst for action, helping to create systems that expand information dissemination and use in a networked digital environment while responding to the needs of academe. Leading academic organizations have endorsed SPARC.

What Does SPARC Do?
The Case for Institutional Repositories
http://www.arl.org/sparc/IR/ir.html

The Case for Institutional Repositories:
A SPARC Position Paper

Prepared by Raym Crow, SPARC Senior Consultant

The Scholarly Publishing & Academic Resources Coalition
21 Dupont Circle · Washington, DC 20036
www.arl.org/sparc
ACRL Scholarly Communication Initiatives

Principles and Strategies for the Reform of Scholarly Communication

Scholarly Communication Defined

Scholarly communication is the system through which research and other scholarly writings are created, evaluated for quality, disseminated to the scholarly community, and preserved for future use. The system includes both formal means of communication, such as publication in peer-reviewed journals, and informal channels, such as electronic listservs. This document addresses issues related primarily to the formal system of scholarly communication.

One of the fundamental characteristics of scholarly research is that it is created as a public good to facilitate inquiry and...
Why do it?

- Change scholarly communication
- Increase institutional visibility
- Highlight individual achievement
- Improve access
- Make connections to other resources
- Preserve materials
- Increase collaboration
What’s your timeline?

➢ Are there external pressures?
➢ When do you want to go public?
➢ How long do you have to plan?
➢ Are you in this for the long haul?
Can you afford it?

➤ Will you absorb the work?
➤ Can you hire new people?
➤ Do you have the needed hardware and software to support it?
Who will be involved?

- Library staff?
- Computing center staff?
- Faculty? Students?
- Campus administrators?
- Community partners?
- Sister institutions?
- Professional marketers?
- Academic or commercial publishers?
How will you measure success?

- Numbers of items collected?
- Use of materials?
- Income generated from it?
- Numbers of participants?
- User studies or surveys?
How flexible is your vision?

- Target group
- Criteria for submission
- Model for submission
- Subsidiary services
- Long-term preservation and access
How to Implement an Institutional Repository: Part II

Technical Issues
Technical Issues

- Technical expertise
- Hardware and software
- Metadata support
- Interoperability
- Version control and revision
- User Interface
- Digital preservation
Technical expertise

- Knowledge of operating systems and servers
- Knowledge of database structure
- Ability to troubleshoot
- Ability to install patches and updates
- Ability to pull together pieces from variety of sources
- Knowledge of standards
Hardware and software

- Open source
- Purchased or licensed
- Locally mounted or hosted externally
- Adequate server space
- Robust backup mechanisms
Software review

- Dspace
- E-prints
- GnuPrints
- ProQuest
- ContentDM
Software requirements

- Permit the easy creation, use, and administration of digital objects distributed over the Internet
- Facilitate the creation of collections of materials in different disciplines or categories
Software requirements

- Support any type of file
- Carry out searches based on standard metadata
- Flexible metadata capture, edit, and display
Software requirements

- Plug into your local authentication system
- Be constructed using components and technologies that are standard and non-proprietary
- Easily integratable
Software requirements

- Customizable user interface
- Modular
- Flexible system administration
- Granular authorizations
Software requirements

- Scalable
- Manage licenses and permissions
- Recoverability
- Ease of managing underlying database
- Statistics and reports
Software requirements

- Flexible egress
- Flexible ingest
- Already implemented in other institutions
- Easy to set up
- Affordable
Metadata support

- Underlying metadata structure
- Ease of modification
- Global change capabilities
- Controlled lists of terms
Interoperability

What’s the objective?

➢ Facilitate sharing based on common standards
  ➢ Link digital archives around the world
  ➢ Provide access to metadata – and files
Interoperability

What’s the solution?

- OAI-PMH
  - Consistent interface
  - Minimal implementation
  - XML representation of Dublin Core metadata set
OAI Registries

Registry of Open Access Repositories (ROAR)

1) //ask23 //archivsystem k23
   Other softwares (various) (OAI)
   Germany
   Other
   ask23 is an electronic, webbased platform for publishing and archiving digital media with a focus on artistic and scholarly work, publication and research.
   ask23 is a project of the Laboratory of Arts and Science, Academy of fine Arts Hamburg, Germany. ask23 is eine Archiv- und Publikationsplattform für die künstlerische und wissenschaftliche Arbeit, zur Veröffentlichung von Textproduktionen und der Abfrage von digitalen Ressourcen sowie des analogen Archivs in Raum k23.
   Total OAI Records: 42
   50% freely accessible fulltext ('estimate')

2) 11th Joint Symposium on Neural Computation
   GNU EPrints (OAI)
   United States
   e-Journal/Publication
   info:other:archives.eprints.org:import
   Total OAI Records: 30
   100% freely accessible fulltext ('estimate')
Version control and revision

- Correct identification
- Security
- Revision
Public user interface

- Submission
- Searching
- Reuse of content
Digital preservation

- Ensuring the long-term maintenance of a bitstream (the zeros and ones):
  - backing up files and keeping a copy at an offsite location
  - running checks to track the deterioration of storage media, files or bitstreams
Digital preservation

- Providing continued accessibility of the contents:
  - *viability* - making sure that information is intact and readable from the storage media
  - *renderability* - making sure that information is viewable by humans and able to be processed by computers
  - *understandability* – making sure that information is able to be interpreted by humans
Digital preservation strategies

- Bitstream copying
- Refreshing
- Durable/persistent media
- Digital archaeology
- Analog backups
- Migration
- Emulation
Digital preservation components

- Metadata registry
- Format registry
- Checksum verification
- Backup procedures
- Persistent identifiers
Attributes of a Trusted Digital Repository

- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technological and procedural suitability
- System security
- Procedural accountability
- OAIS compliance
How to Implement an Institutional Repository: Part III

Resource Issues
Develop a business plan

- Easier said than done
- What do you take into account?
- How far and how long are you willing to subsidize the IR?
Staffing needs

➢ Who will be involved?
➢ How much service will you provide?
How much service will you provide?

- Assist with or handle submissions?
- Convert files on ingest?
- Clean up or review metadata?
- Digitize hard copy?
- Develop supplemental pages or web forms?
- Develop customized search interfaces?
- Assist with copyright investigation?
How much service will you provide?

- Provide use statistics?
- Provide current awareness services?
- Integrate with other services?
- Develop marketing tools?
- Set up focus groups?
- Long-term preservation of files?
- User support and problem resolution?
Can you afford it?

- Will you absorb the work?
- Can you hire new people?
- Will you use volunteers?
Other costs

- Server capacity
- Operating system
- Database set-up
- Licenses and certificates
- Registration with other services
- Marketing expenses and supplies
- Attending meetings and conferences
How to Implement an Institutional Repository: Part IV

Policy Issues
Policy issues

- Division of responsibilities and clarification of roles
- Structure of the archive
- Definition of communities, collections, and users
- Control of content
- Submission and withdrawal
- Metadata standards
- Institutional commitment
- Copyright, permission, and access
- Going it alone or multi-institutional
Division of responsibilities and clarification of roles

• Your target group and users can help with:
  ➢ Identifying and submitting content
  ➢ Verifying copyright
  ➢ Serving as advisers
  ➢ Letting you know of service problems
Structure of the archive

- Depends on software and technical support
  - Communities, collections, titles, files
  - Direct or mediated submissions
  - Need for local modifications
- Depends on your definition and resources
  - Target group
  - Name
  - Type of content
Structure: software implications

• Communities
  – Sub-communities
• Collections
  – Titles
   » Files
Definition of the repository: UO

- Somehow affiliated with the UO
- Academic content, or in support of the academic mission
- Cumulative and perpetual
- Open access
- Interoperable
Definition of communities, collections, and users

- Definition of community
- Limits on the number and type of collections?
- Different rules for different groups?
- Commercial use permitted or not?
Limits on the number or type of collections

- Will you limit?
- Who makes the decision?
- What are the determining factors?
Types of collections

- Informational
- Administrative
- Primary resources for research or study
- From the faculty
- From students
- Groups only or individuals also
Different rules for different groups

- One set of rules and services for students
- Another set of rules and services for faculty
- Other groups
Control of content

➢ Who sets the standards for the content?
➢ What type of material is acceptable?
➢ Who owns the content once deposited?
➢ Will you restrict access to any content?
Submission and withdrawal

- Mediated submission
- Author self-submission
- Review of submissions
- Replacement of files
- Withdrawal
Policies for submission and withdrawal: UO example

- Every community determines the specific submission policies for its collections.
- Someone, whether from the library or the community itself, will revise new submissions to make sure the content is appropriate for the collection where they were submitted.
- Authors may submit their own work or they may ask the library for assistance.
Policies for submission and withdrawal: UO example

➢ At the start of every new collection, someone from library reviews first few submissions

➢ Library reserves the right to remove content in cases of copyright violation

➢ Library reserves the right to withdraw content and return it in the event that the repository cannot be maintained

➢ Library reserves the right to change its guidelines without telling users
Revision or replacement of files

➢ Will you permit the replacement or revision of files?
  ➢ If so, in what circumstances?
Withdrawal of content

➢ Will you withdraw submissions to your IR?
➢ If so, in what cases?
➢ Will you leave any markers or placeholders for that missing content?
Metadata standards

- Based in Dublin Core or some other standard
- Determined by software limitations
- Mediated versus self-submission
- Controlled vocabulary issues
- Whatever is necessary to find, organize, and display the files appropriately
  - It’s not cataloging
Institutional commitment

- Permanence of the archive
- How to guarantee the integrity of the files
- How to guarantee the stability of the archive
Copyright, permission, and access

University of Oregon
PERMISSION FORM FOR STUDENT SUBMISSIONS TO SCHOLARS' BANK

(BEFORE filling out this form, please be sure you have consulted the Student Options for Submitting Dissertations, Theses, or Class Projects to Scholars' Bank page)

- Your name: 

- Your email address: 

- Title of your dissertation, thesis, or class project: 

- Class or program for which your work was done 

I grant the University Libraries permission to digitize my dissertation/thesis/class project and make it freely and publicly available in the digital archive known as Scholars' Bank (https://scholarsbank.uoregon.edu).

I grant permission: 

- Yes 

- Please add the following keywords to my work to help people find it: 

---

Scholars' Bank
Home

What is Scholars' Bank?
- Overview
- Scholarly Communication

UO Implementation
- Community Guidelines
- FAQ
- License Agreement
- Submitting to Scholars' Bank
- Resource Page for Faculty
- Resource Page for Students

News & Related Readings
- DSpace in the News
- Related Readings or Sites

People
- Steering Committee
- Advisory Board
How to Implement an Institutional Repository: Part V

Marketing
Initial vision

- Faculty content
- Self-submission
- Buy-in to open access model
First eight months

First Eight Months of Scholars' Bank

- Library: 12%
- Econ Wk Papers: 72%
- PPPM: 5%
- Spouse: 11%
Cultural challenges

- Intellectual property concerns
- Authenticity of submissions
- Concern over quality
- Concern over control
- Terminology
- Technophobia
Technical challenges

➢ Submission template
➢ Metadata
➢ Version control and revision
➢ User interface
➢ Preservation
Practical challenges

➢ Time
➢ Money
➢ Identification of content and communities
➢ Acquisition of content
Strategies

- Develop local context
- Explain the broader context
- Obtain a critical mass
- Obtain interesting and attractive content
- Make connections wherever possible
- Publicity – be creative and shameless
- Improve the suite of services
- Borrow ideas from other archives
Local context

- Provide a general overview without jargon
- Personalize the information
- Answer questions of local interest
- Build in redundancy
- Provide varying levels of detail
- Educate on issues and link to broader movement
Broader context

- Link to worldwide movement
- Educate on issues
Obtain a critical mass

- What is the magic number?
- Different for each institution
- When existing materials start to draw in other materials without active solicitation
How to obtain critical mass

- Expand vision of the IR: examples
  - Faculty output
    - Individual works and scholarly journals
  - Library collections
    - Electronic records, finding aids, historical materials, etc.
  - Campus publications and presentations
    - Newsletters, planning documents, historical publications, streaming video, etc.
- Student works
  - Theses and dissertations, honors projects, etc.
- Miscellaneous
Publicity – be creative and shameless

➢ Promote anywhere and everywhere
➢ Use subject specialists in the effort
➢ Share success stories
➢ Link from the catalog
➢ Work with known allies
➢ Never miss an opportunity
Recent improvements to service

- Full text
- Provide use statistics
- Sub-communities
- Create links between collections
Further improvements

- Digitize documents
- Support the investigation of copyright
- Implement Creative Commons licenses
- Modify the user interface and develop searching guides
- Contribute further to code developments
Collaborative opportunities

- Share information
- Share strategies
- Develop federated search of repositories
- Develop shared collections
Challenges for the future

- Get buy-in at the highest institutional levels
- Refine the preservation program
- Keep abreast of changing technology
- Develop self-submission model more fully
Strategies for future growth: UO

- Continue to acquire content
- Continue to promote shamelessly
- Align more closely with instructional programs
- Establish a campus advisory group
- Develop searching guides
- Integrate the IR in the new website design
How do we measure success?

- Expanding access to materials
- Capturing grey literature
- Enhancing instruction
- Highlighting individual achievement
- Increasing institutional visibility
- Integrating different types of content
- Increasing collaboration
Acknowledgements

- Eliz Breakstone
- Heather Briston
- Mary Grenci
- Corey Harper
- Barbara Jenkins
- JQ Johnson
- Adam Byrd, Rebecca Fischer, Cathy Flynn-Purvis, Nathan Georgitis, Mary Grenci, Hilary Hart, David Landazuri, Marion Obar, Lori Robare