DIGITAL REPOSITORY
MANAGEMENT UNCOVERED

PRECONFERENCE
WEBWISE 2010
DENVER, CO – MARCH 3, 2010

Jessica Branco Colati – Colorado Alliance of Research Libraries
Sarah L. Shreeves – University of Illinois at Urbana-Champaign
Who we are...

I’m Sarah...

I’m Jess...

…and we’re repository managers! (and, today, so are YOU!)
Workshop goals

- Identify key areas of repository management
- Understand best practices in key areas
- Identify potential strategies for managing dig repos
- Understand range of tools available
- Connect with other repo managers
Meant to be..

- As interactive as possible
- High-level (we only have 3.5 hours after all!)
- Relevant
- Actionable

https://www.ideals.illinois.edu/handle/2142/15082
It’s not...

- A preservation workshop
- A metadata workshop
- A scholarly communication workshop
- A building digital collections workshop
- A copyright workshop
- An identity management workshop

It is a “101” workshop focused on general principles and areas of repository management
Workshop agenda (and housekeeping)

- Exercise 1
- Repository Management 101: Getting Started
- Exercise 2
- Repository Management 101: Supporting Discovery and Use

~Break~

- Intellectual Property and Identity Management in Repositories
- Managing Repositories for the Long-term: Sustainability and Durability
- Exercise 3
- Wrapping up Uncovered
Exercise 1

Understanding Your Repository
Exercise One

Each table has two repository scenarios. With your tablemates choose one scenario; you will be using this scenario throughout the rest of the afternoon.

Based on the description:

-- Who are likely to be the major stakeholders in the repository?
-- What are the areas of repository management that are likely to be of special concern?
-- What kind of policies should it have in place?
Getting Started
What is a repository anyway?
Dear Diary
Repository functions

Ingest
Store
Manage
Access
Support preservation services
Support use of content
Others?
experience with focus group evaluations, end-user usability testing, or other recognized methodology for assessing usefulness and benefit of library applications and/or services.
Repository management is not really just about the repository, but about the services that the repository supports.

It is also not all or nothing.
Repository components

- Hardware
- Software
- Trust
- Relationships
- Content
- Controls
Hardware: the parts nobody* sees

- Servers
- Storage
- Back-up Appliances

And Infrastructure...
- Networking /Connectivity
- Power
- HVAC
- Security

*Well, ALMOST Nobody...
Software: The parts people see

- Vendor-purchased
- Open Source
- Hybrid Products

May be used to support back-end, front-end, and “middleware” applications
Content: The things people want

- Active scholarship
- Records of enduring value
- Unique resources

A.K.A.

Digital Assets, Digital Objects, Items, Resources, Files...

OR

Stuff
### General Farm Information

**Location**

- **County:** Champaign
- **Upcoming growing season:** 2008

**Water**

How does your farm get its water?  
- **Ground/Municipal:**

**Non Organic Production**

- **Full Time Employees:** 0.00
- **Full Time Wage:** $10.00

**Organic Production**

- **Full Time Employees:** 0.00
- **Full Time Wage:** $10.00

**Debt Payment**

- **Total Loan Payments Per Month:** $50.00

**Energy Sources**

- **What does your tractor burn?**
  - Diesel
- **Do you dry your own grain?**
  - Yes
- **How do you dry grain?**
  - Dutch irrigation
- **What fuel do you dry with?**
  - Natural gas
An ever-changing web of personal, organizational, cooperative, conceptual, and contractual understandings, agreements, and consensuses.
Controls: Keep functionalities in order

 Deposit
 Review
 Access
 Reproduction
 Security
 Disaster/Risk Mitigation
Trust: What keeps the repository going

- Reliable, long-term access
- Responsibility for long-term maintenance of digital resources
- Commonly accepted conventions and standards

Policies, practices, and performance that can be audited and measured
Evaluating Trustworthiness

- TRAC: Trustworthy Repositories Audit & Certification: Criteria and Checklist
  - Builds on Open Archival Information System reference model
  - Checklist divided into three sections:
    - A. Organizational infrastructure
    - B. Digital object management
    - C. Technologies, technical infrastructure, and security
  - Basis for comparing local capabilities against a set of core criteria
  - Can be used as an objective measure or a planning document – or both!
What policies and procedures are important in your repository scenario?
Collection policy

- What’s in scope?
- What’s out of scope?
- Who makes decisions about collections?
- Do you turn down content?
Participation and management policies

- Who can direct development?
- Who can contribute content?
- Who determines who can contribute content?
- Are pieces of the repository managed by other groups?
Preservation policy & procedures

- What commitment are you making? (Hint: it doesn’t have to be everything for ever and ever)
- What’s your mandate / responsibility for making those commitments?
- Who is responsible?
- Will you take any file format into repository? What does this mean for preservation?
- Do you have migration policies and procedures?
Metadata policy

- What are your use cases? Access? Preservation?
- What are the needs of the stakeholders?
- What is system created? User created? Required?
- What standards, best practices, schemas do you implement? When and how do you deviate?
How do you manage copyright issues?
If content is user contributed, what’s your license agreement?
What’s your procedures on potential copyright violations? Is there a take down policy?
Are there privacy / confidentiality issues for the content?
Are there student records? Is there sensitive information?
Access and use policy

- What can users access and what can’t they?
- Are there use restrictions due to copyright or other reasons?
- What are your accessibility commitments?
- Do you allow machine indexing? Harvesting?
- Do you track users on your site? Do you disclose that you track users?
Communication policy

- Feedback processes?
- How do you communicate outages?
- How do you communicate your policies? Is there a welcome packet?
- Who is responsible for communication with users?
Service Level Agreements / MOUs

- What commitments are you making in terms of level of service?
- What’s your scheduled downtime?
- Emergency contact information?
- What dependencies exist?
Policies drive procedures and practices.

Policies need to be revisited regularly.

Flexibility is important.

Policies and procedures will change.
Where do I find example repositories?


- Repository software websites

- Ask on listservs for examples
  - Subscribe to repository- and software-specific lists

- Poke around well-established repositories
  - Professional literature scans
  - There’s always Google…
Exercise 2

Keeping Your Repository on TRAC
Exercise Two

As a table, review each of the six criteria and choose one from each section to evaluate in terms of your repository scenario.

--What kinds of documents, policies, or activities would need to be defined and undertaken to meet the criteria?

--Who would the repository manager need to work with to see they were created, approved, and/or carried out?

--What would be the technical and/or financial considerations of meeting the criteria?

--If you had to prioritize resources to meet the three selected criteria, how would you determine which was most important to your institution?
Supporting Discovery, Access, and Use
Components of Access and Use Policies

- Dissemination of Content
- Restrictions on Access
- Restrictions on Use
- Privacy Policy and Tracking Use of Content
- Linking to Content
What’s your audience?
How do they find content?
Is your content where they’re looking?
Discovery: Does your repository support:

- XML sitemaps?
- OAI Protocol for Metadata Harvesting?
- Google Scholar indexing?
- RSS feeds for new additions?
- Zotero and other citation managers capture?
- Open API’s for machine to machine work?
- Metadata integrated into your catalog or federated/aggregated search system?
Economic and fiscal impact analysis: a primer

Title: Economic and fiscal impact analysis: a primer

Author(s): Campbell, Harrison S.; Crihfield, John Brevard, 1952-

Contributor(s): University of Illinois at Urbana-Champaign. Agricultural Experiment Station.

Subject(s): Urban impact analysis

Geographic Coverage: Illinois.

Issue Date: 1992

Publisher: [Urbana, Ill.]: Agricultural Experiment Station, College of Agriculture, University of Illinois at Urbana-Champaign,

Series/Report: University of Illinois bulletin; 807

Type: Text

Language: English

URI: http://hdl.handle.net/10111/UIUCOCA:Archive/economicfiscalim00camp

http://hdl.handle.net/2142/8627

Rights Information: Copyright 1992 Board of Trustees University of Illinois.

Date Available in IDEALS: 2003-05-29

Has Version(s): http://hdl.handle.net/10111/UIUCOCA:economicfiscalim00camp

http://www.archive.org/details/economicfiscalim00camp

Identifier in Online Catalog: 3497685

OCLC Identifier: (OCoLC)ocm28276751
Economic and fiscal impact analysis: a primer

Author: Harrison S Campbell

Publisher: [Urbana, Ill.]: Agricultural Experiment Station, College of Agriculture, University of Illinois at Urbana-Champaign, [1992].

Series: Illinois bulletin (University of Illinois at Urbana-Champaign Agricultural Experiment Station), 807.

Edition/Format: eBook: Document: State or province government publication: English

Rating: ★★★★★ (not yet rated) 0 with reviews - Be the first.

Find a copy online

Find a copy in the library
<table>
<thead>
<tr>
<th>Title</th>
<th>Date</th>
<th>Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete conditioning for hockey / Peter Twist</td>
<td>c2007</td>
<td>Books Upper Level - GV848.3</td>
</tr>
<tr>
<td>All roads lead to hockey: reports from northern Canada to the Mexican border / Bill Boyd</td>
<td>2006</td>
<td>Books Upper Level - GV847</td>
</tr>
</tbody>
</table>

**Search Found In**
- **Title**: (1317)
- **Subject**: (231)
- **Author**: (36)

**Format**
- **BOOKS/JOURNALS**: (172)
- **DVD/VIDEOS**: (16)
- **DISS/THESSES**: (1)

**Collection**
- **PEAK Digital Repository**: (1277)
- **The Cable Center**: (1)
- **Penrose/Music**: (188)
- **Reference**: (2)
- **Journals/serials**: (5)
+ more
It is not an overstatement to say that the transition from analog to digital technologies is changing everything – content creation, distribution, content packaging,
Managing access and use

**Supporting Access**
- Machine indexing and mining?
- Persistent identifiers?
- Downloading?
- Streaming?
- Viewers and players?
- Web 2.0 features
  - Tags, Comments, Annotations

**Controlling Use**
- Restricting access to IP ranges and/or authorized users?
- Include access control policies in objects?
- Generate “lesser” quality access versions?
- Disabling “right-click” functions in players and viewers?
- Provide information for obtaining access?
- Require agreements for use?
The frog prince
1972
Running length: 52 minutes

For details about this item, see the Libraries' Catalog »

SUMMARY

CREDITS

Muppet performers, Frank Oz, Jerry Nelson, Carl Banas, Richard Hunt, John Lovelady, Daniel Seagren; puppet costumes, Bonnie Lewis, Caroly Wilcox, Kermit Love.

SUBJECTS

» Fairy tales
» Kermit the Frog (Fictitious character)
» Made-for-TV movies
» Muppets (Fictitious characters)
» Performing Arts, Music
» Puppet television programs
» Television specials

To cite or link to this item, use this identifier:
http://hdl.handle.net/1903.1/363
Welcome Back!
Access criteria from Exercise Two?

B6.2 Policy for recording access actions
Repository has implemented a policy for recording all access actions (includes requests, orders etc.) that meet the requirements of the repository and information producers/depositors.

B6.4 Policy for access
Repository has documented and implemented access policies (authorization rules, authentication requirements) consistent with deposit agreements for stored objects.
Repository Management 101

Intellectual Property and Identity Management
Do you have copyrighted material available? If so, is it clear that it is under copyright?

Do you have copyrighted material available under a Creative Commons license? Is this clear?

Do you have public domain content available? Is it clear that it is in the public domain?

Equally important, is it clear what users may do in both cases under public domain and fair use provisions?*

Do you have a license agreement when users submit content?

*See M. Schlosser’s “Unless Otherwise Indicated: A Survey of Copyright Statements on Digital Library Collections” College and Research Libraries vol. 70 no. 4 371-385
Quick case study: Institutional Repository

**content coming in**
- Generally under copyright
- Often copyright transferred to publisher
- Self submission so don’t always know what is coming in

**content going out**
- Accessed all over world
- Can state copyright but can’t control use
- Complicated by different versions of content
Resources

Copyright Term and the Public Domain in the United States: [http://www.copyright.cornell.edu/resources/publicdomain.cfm](http://www.copyright.cornell.edu/resources/publicdomain.cfm)

Fair Use of Copyrighted Materials: [http://www.utsystem.edu/ogc/intellectualproperty/copypol2.htm](http://www.utsystem.edu/ogc/intellectualproperty/copypol2.htm)

SHERPA / RoMEO Publisher Copyright Policies: [http://www.sherpa.ac.uk/romeo/](http://www.sherpa.ac.uk/romeo/)
Active Shibboleth Federation Users LDAP Access Authorization Authentication Controls Community Directory XACML Workflows Accounts Local
Identity management

- “People Metadata”
- Range from manual user account creation to inter-institutional identity assertions
  - SAML, eduPerson, LDAP, AD, Shibboleth
- Complexity and approach dependent on institutional factors
  - Purpose, scale, end-user convenience, existing infrastructure, partnerships with IT
  - How content is ingested, managed, accessed and used will impact identity management needs
# Current InCommon Participants

A community of more than 4 million end users.

(November 2009. Source: Higher Education Students, Faculty, and Staff, Integrated Postsecondary Education Data System.)

<table>
<thead>
<tr>
<th>Higher Education Participants (153)</th>
<th>Government and Nonprofit Laboratories, Research Centers, and Agencies (6)</th>
<th>Sponsored Partners (51)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona State University</td>
<td><strong>Energy Sciences Network (ESNet)</strong></td>
<td>Absolute Software, Inc.</td>
</tr>
<tr>
<td>Augsburg College</td>
<td><strong>Lawrence Berkeley National Laboratory</strong></td>
<td>Apple - iTunes U</td>
</tr>
<tr>
<td>Baylor University</td>
<td><strong>Moss Landing Marine Laboratories</strong></td>
<td>Atlas Systems, Inc.</td>
</tr>
<tr>
<td>Brown University</td>
<td><strong>National Institutes of Health</strong></td>
<td>Blatant Media Corporation</td>
</tr>
<tr>
<td>California Institute of Technology</td>
<td><strong>National Science Foundation</strong></td>
<td>Burton Group</td>
</tr>
<tr>
<td>California Maritime Academy</td>
<td><strong>TeraGrid</strong></td>
<td>Cengage Learning, Inc.</td>
</tr>
<tr>
<td>California Polytechnic State University - San Luis Obispo</td>
<td></td>
<td>Colorado Alliance of Research Libraries</td>
</tr>
<tr>
<td>California State Polytechnic University, Pomona</td>
<td></td>
<td>Davie County Schools</td>
</tr>
<tr>
<td>California State University, Channel Islands</td>
<td></td>
<td>Digital Measures</td>
</tr>
<tr>
<td>California State University, Chico</td>
<td><strong>Ebook Library - EBL</strong></td>
<td>e2Campus by Omnilert, LLC</td>
</tr>
<tr>
<td>California State University, Dominguez Hills</td>
<td></td>
<td>Elsevier</td>
</tr>
<tr>
<td>California State University, Fresno</td>
<td><strong>EBSCO Publishing</strong></td>
<td>Good Steward Software</td>
</tr>
<tr>
<td>California State University, Fullerton</td>
<td></td>
<td>Houston Academy of Medicine - Texas Medical Center Library</td>
</tr>
<tr>
<td>California State University, Monterey Bay</td>
<td></td>
<td>Identit-e</td>
</tr>
<tr>
<td>California State University, Northridge</td>
<td></td>
<td>Internet2</td>
</tr>
<tr>
<td>California State University, Office of the Chancellor</td>
<td></td>
<td>JSTOR</td>
</tr>
<tr>
<td>California State University, Sacramento</td>
<td></td>
<td>Kuali Foundation</td>
</tr>
<tr>
<td>California State University, San Bernardino</td>
<td></td>
<td>Learn.com</td>
</tr>
<tr>
<td>California State University, San Marcos</td>
<td></td>
<td>lynda.com</td>
</tr>
<tr>
<td>California State University, Stanislaus</td>
<td></td>
<td>MCNC</td>
</tr>
<tr>
<td>Carleton College</td>
<td><strong>National Institute for Technology and Liberal Education (NITLE)</strong></td>
<td>Microsoft</td>
</tr>
<tr>
<td>Carnegie Mellon University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case Western Reserve University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Piedmont Community College</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clemson University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College of William and Mary</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Repository Management 101

Preservation, Sustainability, and Durability
“...digital curation is characterised as a risk-management activity;

the job of digital curator is to rationalise the uncertainties and threats that inhibit efforts to maintain digital object authenticity and understandability, transforming them into manageable risks.”

Drambora

(http://www.repositoryaudit.eu/about/)
Review: Repository components

- Hardware
- Content
- Software
- Controls
- Trust
- Relationships

Digital Repository
Mitigating risk: Hardware

- Virtualization, reducing physical footprint
  - Lessens resource consumption
- Redundancy
- Distributed storage, back-up, and disaster recovery strategies
- Cloud “options”

Regardless of in-house or hosted hardware scenarios, risks need to be understood and mitigated whenever possible.
Mitigating risk: Software

- Community is KEY
- Understanding how and where changes to policies and desired functionalities impact software component selection and customization
  - Workflows
  - Versioning
  - Access controls
  - Accessibility
  - Usability
Mitigating risk: Content and controls

- Commit to policies, champion consistency
- Capture, manage, and refresh non-descriptive metadata
  - Administrative information
  - Preservation information
  - Technical information
    - Audit trails, fixity, migration…
Mitigating risk: Relationships

Assessment (and re-assessment) is key

Organizations and players change

Contracts and agreements can and should be regularly renegotiated.

IDEALS example – working with campus IT
The Five Organizational Stages

(Anne Kenney and Nancy McGovern)

The five stages of organizational response to digital preservation are:

- Acknowledge: Understanding that digital preservation is a local concern;
- Act: Initiating digital preservation projects;
- Consolidate: Seguing from projects to programs;
- Institutionalize: Incorporating the larger environment; and
- Externalize: Embracing inter-institutional collaboration and dependency.

http://quod.lib.umich.edu/cgi/t/text/text-idx?c=spobooks;idno=bbv9812.0001.001;rgn=div1;view=text;cc=spobooks;node=bbv9812.0001.001%3A11
Marketing – A Key to Sustainability

- Who’s your audience again? What are their needs?

- How does your repository help meet those needs?

- Are you reassessing the needs of your audience / stakeholder / community in order to adapt the repository?

- Are there others that can help you promote the repository? Reference librarians? Committed users? Collection managers and curators?
Say what you do.

Do what you say.

Show that you do what you say.
All of this leads to trust.
Exercise 3

Building a Repository Management Team
Listed are some of the characteristics that libraries (at least) tend to look for in a repository manager position...

--Which of these skills would you look for in a repository manager or in a management team in your scenario?

--Are there roles that need to be filled with different people?

--What kind of training or development might need to happen?

--Are there skills that you don’t see here that you would include?
Wrap-Up
What does your team look like?
Sustainable preservation strategies are not built all at once, nor are they static. Sustainable preservation is a series of timely actions taken to anticipate the dynamic nature of digital information.

Contact information

Sarah L. Shreeves  
University of Illinois at Urbana-Champaign  
sshreeve@illinois.edu

Jessica Branco Colati  
Colorado Alliance of Research Libraries  
jessica@coalliance.org