
MICHIGAN STATE
UNIVERSITY
MICHIGAN STATE COLLEGE
OF
AGRICULTURE AND APPLIED SCIENCE
University Archives
&
Historical Collections

When the Need for an Institutional Repository Gives Rise to a Federation

Lisa Schmidt

lschmidt@msu.edu

Michigan Academic Library Council

March 18, 2011

Overview

- Institutional Background
- Why an Institutional Repository?
- Why Archivists and Librarians?
- MSU Digital Curation Planning Project
- A Federation of Repositories
- Tips for Achieving Sustainable Success

Michigan State University

- Est. 1855 by act of Michigan Legislature to create agricultural college
- Nation's pioneer land grant college
- Tier one research university with significant national and global impact
- Leader in science and technology
- 47,131 students: 36,058 undergrad, 11,073 graduate/professional

MSU Archives & Historical Collections

- University Archives reports to Vice Provost of Libraries, Computing & Technology (LCT)
- Est. 1969 by Board of Trustees mandate
  - Collect and preserve MSU's historical records
  - Provide university community, scholars, and general public with access to records
  - Approve final disposition and destruction
- 33,000 cubic feet of university records

MSU Archives & Historical Collections

- New UAHC director in 2008, new vision for department's role on campus
- Focus on offering a "service" to campus
- Expanding Records Management Program to address both analog and digital material
- Assist MSU units in the *intentional management* of university business records
- Includes developing new policies, procedures and best practices for new enterprise business systems and digital information

MSU Libraries

- Reports to Vice Provost of Libraries, Computing & Technology
- Main library plus four branch libraries
- Provides resources and services to MSU community, Michigan residents, public
- Collections exceed 5 million volumes
  - Special strengths in Africana, Turfgrass, Comic Arts, voice sound recordings

MSU Libraries

- Annual materials budget approaching $12M
- More than 60% expended on electronic acquisitions
- Growing program area: Collaboration with faculty through creation and curation of digital content and services to enhance learning, teaching, and research
- Recently hired Digital Curation Librarian

Why an Institutional Repository?

- "Digital collections that capture and preserve the intellectual output of a single or multi-university community"—SPARC
- Contain scholarly content
  - Research journal articles
  - Theses and dissertations
  - Other university publications
  - Other digital assets
    - Administrative documents
    - Learning objects

Why an Institutional Repository?

- Global visibility for scholarly research
- Single collection point for scholarly research and other digital assets
- Store unpublished and other "grey" literature that might easily get lost
- Trusted digital repository will ensure that the content is preserved and remains accessible over time

Why an Institutional Repository?

- Michigan State does not have an institutional repository—yet
  - Limited budget
  - Limited staffing options
- Moving in the right direction
  - Library staff lobbying for institutional repository
  - New synergies: Digital Curation Planning Project
  - New repository projects: Spartan Archive
  - Opportunities to leverage access to Central IT

Why Librarians and Archivists?

- Traditional stewards of information
- Trained in management and preservation of all types of media and formats, including digital information
- Can take a neutral stance– the information/ data doesn't belong to us

Why Librarians and Archivists?

"The institutional archive needs to assume more of a policy role, identifying records throughout the campus and working to ensure that digital records are both maintained by their creators and kept ready for research use."

Richard Cox, "The Academic Archives of the Future," *EDUCAUSE Review Magazine*, Volume 43

Digital Curation Planning Project

- Michigan State's growing body of digital assets
  - Faculty and student research
  - Theses and dissertations
  - University publications
  - Institutional records
  - Multimedia collections
  - Digital surrogates of cultural material
  - Learning objects and course materials

Digital Curation Planning Project

- Valuable digital resources created through much time, effort, grant funding, human capital, and research
- Changing technology likely to render digital assets inaccessible absent a long-term management and preservation plan
- Storage limitations decreasing but costs to meaningfully curate these growing collections are increasing

Digital Curation Planning Project

- Some campus units have created their own digital repositories
- No comprehensive, campus-wide digital preservation strategy or guidelines
- MSU does not yet have an institutional repository

Digital Curation Internship at MSU

- Intern from School of Information, University of Michigan, Winter 2009
- Focus on digital multimedia collections
- Interviewed 7 units
- Compiled and analyzed results

Digital Curation Internship at MSU

- Recommendations included:
  - More comprehensive survey needed
  - Guidance on selection and retention
  - Best practices for formats, naming conventions, descriptive and technical metadata
  - Better long-term storage options
  - Institutional repository

Proposed a *Planning Project*

- Collaboration of MSU Libraries, University Archives, and MATRIX digital humanities research center
- Top level buy-in: Vice Provost of LCT funded digital preservation analyst position

Proposed a *Planning Project*

- Engage half-time digital preservation analyst for one year to manage the project
- Invited university-wide participation in team
  - Buy-in and reality check beyond partners
  - Representatives from Registrar's Office, Central IT
  - Hindsight – should have included faculty on team

Proposed Methodology

- Conduct environmental scan of the university's digital assets
- Survey existing digital repositories and technical infrastructure
- Identify best preservation, management, and access practices on campus

Proposed Goals and Deliverables

- Develop policies, procedures and workflows to standardize MSU's approach to digital asset management and preservation
- Explore potential collaborations with other institutions and consortia—such as HathiTrust, LOCKSS, CIC
- Explore Institutional Repository for MSU

Overly Ambitious!

- Would eventually reach saturation point with broad, all-encompassing inventory
- Impossible to complete in one-year timeframe
- Concern over perception of creation of one-size-fits-all data repository, loss of control over digital assets at unit level

Revised Planning Project to …

- Digital *curation* not preservation
- Campus-wide, self-selective survey using web-based questionnaire
- In-depth interviews with select units
- Evaluation of preservation practices and technical infrastructures
- Recommendations and next steps in digital curation planning

Why Digital Curation?

"Digital curation is maintaining and adding value to a trusted body of digital information for current and future use… the active management and appraisal of data over the life-cycle of scholarly and scientific materials."

—Digital Curation Centre, www.dcc.ac.uk

Why Digital Curation?

"Implicit… are the processes of digital archiving and preservation but it also includes all the processes needed for good data creation and management, and the capacity to add value to data to generate new sources of information and knowledge."

—Digital Curation Centre, www.dcc.ac.uk

Baseline Data Questionnaire

- Informal, web-based survey
- Publicized through IT Exchange, *MSU News*, project blogsite
- Encourage participation of technology staff and content creators
- Available for two weeks, October 2009

Baseline Data Questionnaire

- Types of digital content
- Digital content making up largest percentage
- Approximate volume in TB
- Storage media
- File formats
- Formats making up largest percentage

Baseline Data Questionnaire

- Online storage capacity / expansion plans
- Content management systems
- Digital repository software
- Presence of confidential data
- Additional comments

Questionnaire Results

- 90 responses
  - 23 academic departments
  - 31 administrative units
  - 9 research centers
  - 27 technology services units

Questionnaire Results

- Types of digital content varied
- File formats varied
- Storage mostly on hard drives, some combination removable media and networked storage
- 17 units planned increase of storage capacity, most from 1-10 TB
- Several CMS and/or digital repository implementations

Questionnaire Results

- Great interest and enthusiasm in project
- Anecdotal comments
  - "Accumulating more than we can store!"
  - Requests for guidance on identifying and handling archive-worthy files at time of creation
  - How to choose digital asset management system

One-on-One Interviews

- Many respondents – how to select units for follow up interviews?
- Focus on units with established CMSs and/or digital repositories
- AND/OR units with records of enduring value to the university
- Informal, two-hour conversations
- Team members went to the unit's office

One-on-One Interviews

- Digital content, relation to unit mission
- Content that must be preserved
  - Active records – still used/needed by unit
  - Permanent retention – a.k.a. archival
- File formats
- Storage, including any issues

One-on-One Interviews

- CMS and/or digital repository
  - System used and why chosen
  - What it's used for
- Ingest, archival storage/preservation, access processes
- Metadata
- File naming conventions

One-on-One Interviews

Broadcasting Services	MATRIX
Center for Research on Mathematics and Science Education (CRMSE)	MSU Extension/Agriculture and Natural Resources (ANR) Technology Services
Confucius Institute	National Superconducting Cyclotron Laboratory (NSCL)
Department of Art & Art History	Physical Plant Division
Department of Theatre	Turfgrass Information Center (TIC)
University Relations	

Analysis: General

- Units developed solutions that fit nature of data, needs of users
- Some use commercial software, some open source
- Some hold content of archival value to university and/or the unit
- Need for appraisal and preservation guidelines

Analysis: The Good

- Most units back up data
- Some demonstrate good use of metadata
- Many use repository software
- Many have good access interfaces
- Many had strong support from management, stable funding

Analysis: The Good

- Nearly all store preservation masters of some digital content
- Three had means of verifying file integrity
- Some had file naming conventions
- Open to digital curation guidelines

Analysis: The Not-So-Good

- Little emphasis on preservation
- Backups too close to production
- Maintenance of preservation copies not practiced by all units or for all file types
- Practice of checking file integrity low
- Some create/use little or no metadata

Analysis: The Not-So-Good

- Mixed bag on use of file naming conventions
- Little in the way of digital curation policies
- Question of support, sustainable funding
- Cultural and financial inertia
- Interview sample likely best of lot

Metadata Comparison

- Six units had metadata to share
  - MATRIX, Theatre, and MSU Extension: Based on Dublin Core
  - Art & Art History: IRIS data standard for cataloging/management, VRA Core, CCO
  - Physical Plant: Metadata from engineering CMS used to manage facilities assets
  - TIC: Bibliographic indexing terms in Cuadra Star system

Conclusions and Future Impact

- Types of digital content, needs and expectations vary significantly
- Development of common digital curation guidelines an iterative process
- Must be practical, quick and easy – content creators have little time for additional processes

Conclusions and Future Impact

- No silver bullet or one-size-fits-all solution
- Digital curation is part of larger university Records/Information Management Program

Conclusions and Future Impact

- Four types of digital content:
  1. University publications, including e-journals, electronic theses and dissertations
  2. Digital content that documents history of MSU
  3. Non-MSU-specific digital content
  4. Research data
- Unique solutions based on content type and curation needs

Next Steps

- Good Practice
  - Develop *new* workflow for intentional management of records throughout life cycle
  - Develop guidelines to determine whether digital assets should be transferred to Archives or remain in unit custody
  - Develop digital/data curation toolkits for file formats, documentation, intellectual property rights, sharing/dissemination, preservation
  - Link digital repositories to records retention schedules, if appropriate

Next Steps

- Collaborations
  - Foster “communities of practice” of MSU units and other institutions through online forums and meetings
  - Big Ten Universities exploring collaborative storage (and curation!) solution
  - Work with other Big Ten institutions to obtain grant funding for digital curation research

A Federation of Repositories

- Nurturing of repositories that suit the needs of the unit and the content
- Possibly institute a common base repository software, such as Fedora
- Use common metadata, with customization as necessary
- Institute preservation and access guidelines and/or policies as appropriate

A Federation of Repositories

- "Institutional repository"—MSU Libraries
- Permanent preservation environment—MSU Archives (Spartan Archive project)
- Repositories holding course-related content (Theatre, Art & Art History)
- Repositories holding non-MSU content (MATRIX)
- Repositories holding research data

Tips for Achieving Success

- Partner with depositors and other stakeholders
- Offer guidelines rather than rules
- Clearly articulate the benefits
  - It's their repository
  - Global visibility!
- Keep intellectual property rights in mind
- Make it easy for them

References

- Cox, Richard, "The Academic Archives of the Future," *EDUCAUSE Review Magazine*, Volume 43, http://www.educause.edu/EDUCAUSE+Review/
- Crow, Raym,"The Case for Institutional Repositories: A SPARC Position Paper," http://scholarship.utm.edu/20/1/SPARC_102.pdf
- Digital Curation Centre (DCC), http://www.dcc.ac.uk
- Michigan State University Digital Curation Planning Project, http://msudcp.archives.msu.edu/


MICHIGAN STATE
UNIVERSITY
University Archives
&
Historical Collections