In Pursuit of Common Digital Curation Guidelines:
An Exploration of Current Practices at Michigan State

Lisa Schmidt, Electronic Records Archivist

Michigan State University

Archives & Historical Collections

March 29, 2010

Overview

- Michigan State University and the MSU Archives & Historical Collections

- Archives 2.0: Policymakers vs. Custodians

- MSU Archives Electronic Records Initiatives

- Digital Curation Planning (DCP) Project

Michigan State University

- Established in 1855 by act of the Michigan Legislature to create an agricultural college

- Nation's pioneer land grant college

- Tier one research university with significant national and global impact

- Leader in innovation and technology

- 46,648 students: 36,337 undergrad, 10,311 graduate/professional

Michigan State University

MSU Archives & Historical Collections

- Official repository for the historical archives of Michigan State University
- Established by Board of Trustees mandate in 1969
  – Collect and preserve historical records of MSU
  – 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

- Subjects:
  - Administration
  - Athletics
  - Campus buildings and grounds
  - Student groups and activities
  - Faculty papers and research
- 700+ historical collections related to Michigan and the Great Lakes region

MSU Archives & Historical Collections

- Actively assists MSU units in efficient administration and management of official university records

- Includes management, collection, and preservation of electronic records

Archives 2.0

"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

Electronic Records Initiatives

- Document management system
  - Exploring both enterprise DMS and guidelines for local level DMSs

- Spartan Archive
  - NHPRC-funded project to develop a governance structure and technical infrastructure to accession, provide access to, and preserve electronic records

Electronic Records Initiatives

- Enterprise Business Systems Project (EBSP)

  - Multi-year initiative to create streamlined business processes and interconnected administrative systems for MSU's finance, human resources, and research administration

- Digital Curation Planning Project

Digital Curation Planning Project

- The Problem

- Digital Curation Internship

- Original Digital Preservation Plan Proposal

- New, Current Digital Curation Plan

The Problem

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

The Problem

- 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

The Problem

- Some campus units have created their own digital repositories

- But—no comprehensive, campus-wide digital preservation strategy or guidelines

- No institutional repository

Digital Curation Internship

- Winter 2009

- Intern from University of Michigan School of Information

- Assessed problem space in relation to digital multimedia collections

- Interviewed 7 units

Digital Curation Internship

- Recommendations
  - More comprehensive survey needed
  - Guidance on best practices in selection, formats, naming conventions, metadata
  - Better long-term storage options
  - Institutional repository

The Solution: Original DP Proposal

- Digital preservation plan rooted in best practices to provide trustworthy stewardship of digital assets and intellectual property

- Collaboration of MSU Libraries, University Archives, and MATRIX

- Top level buy-in: VP of Libraries, Computing and Technology funding digital preservation analyst position

The Solution: Original DP Proposal

- Engaging digital preservation analyst for one year

- Planning team
  - Representatives from other units
  - Monthly meetings
  - Buy-in and reality check beyond Archives, Libraries, and MATRIX

Original Digital Preservation Plan

- Conduct an environmental scan of the university's digital assets

- Survey of MSU's existing digital repositories and technical infrastructure

- Identify best preservation, management, and access practices already on campus

Original Digital Preservation Plan

- Develop policies, procedures, and work flow to standardize MSU's approach to digital asset management and preservation

- Explore potential collaborations with other institutions and consortia—such as HathiTrust, LOCKSS, CIC

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 of digital assets at unit level

New Digital Curation Plan

- Campus-wide, self-selecting survey using web-based questionnaire

- In-depth interviews with select units

- Inventory and appraise digital assets of select units

- Evaluate technical infrastructures, storage needs, metadata schemes, and naming conventions

New Digital Curation Plan

"Stop disciplining data and start herding it."

Steve Bailey, *Managing the Crowd*

Anticipated Outcomes

- Guidelines for electronic records appraisal, preferred file formats, metadata, and file naming conventions

- Layered storage solution and file transfer methodologies

- Foundation for the establishment of an institutional repository or institution-wide federation of digital repositories

Storage

- Central IT supports administrative business systems, e-mail, academic support functions

  – Pro: More efficient management of electronic records and digital assets

- Tradition of local IT staff managing unit systems

- Poor economy merits closer look at central vs. local IT

Storage

- Central IT developing virtual server environments to local units

- Layered storage, a variety of storage types or levels to meet diverse needs

  – Local storage for files of temporary, short-term use

  – Permanent long-term storage environment, possibly under custodianship of the Archives

What is 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)

What is 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)

Digital Curation in the Zeitgeist!

- Initiatives at other universities
  - Penn State, Ohio State, Duke, Yale
- Interest in project from other institutions
  - University of Utah, James Madison, Smithsonian
- Invitation to present at ALA Midwinter
- Invitation to submit journal article
- Approached by NEH to develop proposal with other CIC institutions

Baseline Data Questionnaire

- Informal, web-based survey

- Publicized to potential participants through IT Exchange, MSU News, project website/blog

- Encouraged participation of technology staff and content creators

- Available for two weeks in October 2009

Baseline Data Questionnaire

- Types of digital content

- Digital content making up largest percentage

- Approximate volume of digital content in TB

- Storage media

- File formats

- Formats making up largest percentage

Baseline Data Questionnaire

- Online storage capacity and expansion plans

- Content management systems used

- Digital repository software used

- Presence of confidential data

- Additional comments

Questionnaire Results

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

Questionnaire Results

- Types of digital content varied considerably

- File formats varied considerably

- Storage mostly on hard drives, some combination of removable media and networked storage

Questionnaire Results

- 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

- Large problem space—how to break down?

- Decided to start by focusing on units with content management systems and/or digital repositories

- Informal, two-hour conversations rather than formal interviews

- Held at unit office

One-on-One Interviews

- Digital content, relation to mission of unit

- Content that must be preserved
    - Of ongoing use
    - Archival, documents activity of unit or university

- File formats

- Storage, including any issues

One-on-One Interviews

- Content management system and/or digital repository
  - System used and why it was chosen
  - What it's used for
- Ingest, archival storage/preservation, access processes

One-on-One Interviews

- Metadata stored with or related to content

- File naming conventions

One-on-One Interviews

- MSU Extension/Agriculture and Natural Resources (ANR) Technology Services
    - DotNetNuke, SharePoint, Intrafinity Portal (written for MSUE)

- Art & Art History Department
    - Master image files stored offline
    - Access files stored in MDID
    - Metadata catologed using IRIS

One-on-One Interviews

- Confucius Institute
  - Promotes Chinese Language/Culture Education
  - Version Cue, Subversion (SVN)
- Department of Theatre
  - 75% digital photos, 15% CAD drawings
  - In-house CMS based on LAMP
  - ResourceSpace digital repository

One-on-One Interviews

- Broadcasting Services

- Center for Research on Mathematics and Science Education (CRMSE)

- Turfgrass Information Center (TIC)

- MATRIX

- National Superconducting Cyclotron Laboratory (NSCL)

- Physical Plant

Prototype: University Relations

- Public relations arm of Michigan State
- Hold records of historical value to the university
- Servers bursting at the seams with digital photos and video

University Relations: Digital Photos

- Hundreds of thousands of digital photos

- 4.6 TB on networked servers

- Nikon RAW NEF, TIFF, JPEG formats

- Some embargoes and use restrictions

- 21,000 images indexed in Extensis Portfolio

- 5,100 images publicly available through NetPublish Portfolio

UR Digital Photos: Value?

- Some of historical/archival significance

- Many of temporary use/value

- Many of no current value, should be disposed of

UR Digital Photos: Curation Needs

- Records inventory

- Appraisal of currently held files
  - Identify permanent records of archival value

- Storage
  - Preservation space/environment for archival masters

  - Public access space for database and low resolution files

University Relations: Digital Video

- MSU Today show on Big 10 Network

- Shot in XDCAM HD

- Shows run 30-60 minutes

- Avid, Open Media Framework, QuickTime formats
  - Avid for editing
  - QuickTime for access copies

University Relations: Digital Video

- 2 TB networked storage

- 6 TB non-networked storage

- 4 TB internal storage on editing machines

- 16 TB "scratch" storage on Avid Unit network

University Relations: Digital Video

- Access versions uploaded to YouTube with closed captioning

- Tapes sent to Provost's office

- UR keeps two edited versions
  - Show master, including text overlays
  - Clean master

- Most usage within 6 months of production

UR Digital Video: Curation Needs

- Preservation guidelines, including format recommendations

- Archival storage

- File transfer workflow

- Ability to provide access or reproduce as needed

University Relations: Next Steps

- Records inventory and appraisal/selection guidelines
  - Meet with content creators and users
- Storage options
- Custody and file transfer workflows

Analysis

- Units developed solutions that fit nature of their data, needs of their users

- Some using 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 backing up data in some fashion

- Many demonstrate good use of metadata

- Many using repository software to manage digital content

- Many had good access/discovery interfaces

- Many had strong support from management, stable funding

- Open to digital curation guidelines

Analysis: The Not-So-Good

- Little emphasis on/plans for preservation

- Backups/server mirrors too close to production

- Some create and use little or no metadata

- Little in the way of digital curation policies

- Question of support, sustainable funding

- Cultural and financial intertia

Metadata Comparison

- Investigate metadata approaches and schemas used by prototype units

- Compare to Dublin Core metadata elements

Metadata Comparison

- Six units had metadata to share
  - MATRIX, Theatre, and MSU Extension based on Dublin Core
  - Art & Art History uses IRIS data standard for cataloging and management of art images, metadata based on VRA Core and CCO
  - Physical Plant metadata from engineering content management solution used to manage facilities assets
  - TIC uses bibliographic indexing terms in Cuadra Star system

Digital Curation: Next Steps

- Best/good practices recommendations and guidance

- Foster "Community of Practice" meetings

- NEH grant proposal

- Development of layered storage plans, including transfers to Archives

- More appraisal help
  - University Relations
  - Other units, especially those with content to transfer to Archives

Digital Curation: Next Steps

- Exploration of unrepresented units with other types of digital content and curation practices

- Three tiers
  - Curation plan for MSU publications, to include investigation into the options of creating an institutional repository or using CLOCKSS or other LOCKSS-based distributed preservation solution
  - Curation guidelines for material that should be transferred to Archives
  - Curation guidelines for non-MSU-related material

Deliverables & Dissemination

- Final report to University for this phase of project

- Project website: http://msudcp.archives.msu.edu/

- Presentations

  - ALA Midwinter Meeting, January 2010

  - SAA Student Chapter, School of Information, University of Michigan, March 2010

  - SAA Annual Meeting, August 2010

- Publications

  - *Library Resources & Technical Services (LRTS)* journal article, date TBD

References

- Cox, Richard. "The Academic Archives of the Future," EDUCAUSE Review Magazine, Volume 43, http://www.educause.edu/EDUCAUSE+Review/

- Digital Curation Centre (DCC), http://www.dcc.ac.uk

- Michigan State University Archives & Historical Collections, http://www.archives.msu.edu/

- Michigan State University Digital Curation Planning Project, http://msudcp.archives.msu.edu/