MICHIGAN STATE UNIVERSITY

University Archives & Historical Collections

A Digital Curation Planning Project at Michigan State University

Cynthia Ghering & Lisa Schmidt

Research Forum

Society of American Archivists

August 10, 2010

Overview

- Project Background
- Research Methodology
- Results and Findings
- Conclusions
- Future Impact

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

Michigan State University

"Belle Sarcastic"

MSU Archives & Historical Collections

- University Archives reports to Vice Provost of Libraries, Computing & Technology (equivalent to CIO)
- 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 and a new vision for the 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 policy, procedure and best practice for new enterprise business systems and digital information

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

- **Enterprise Business Systems Project (EBSP)**
  - Multi-year, streamline business processes and connect administrative systems for MSU's finance, human resources, and research administration
- **Enterprise Document Management System (EDMS)**
  - Implement guidelines for local DMSs and future EDMS
  - Opportunity to standardize business workflow:
    - link digital content to retention schedule
    - Develop university-wide classification schemes
    - file formats, names and version control
  - Incorporate principles of Trusted Digital Repositories

Electronic Records Iniatitives

- **Spartan Archive**
  - NHPRC-funded project to develop workflow and technical infrastructure to accession, provide access to, and preserve electronic records
- **Digital Curation Planning Project**
  - Internally funded, one year project to investigate MSU's growing body of digital assets and information
    - Institutional records, faculty and student research, theses and dissertations, university publications, multimedia collections, learning objects and course materials, digital surrogates

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

- MSU does not have an Institutional Repository
  - we can learn from early implementers
- Some campus units have created their own digital repositories
- No comprehensive, campus-wide digital preservation strategy or guidelines

Digital Curation Internship at MSU

- Intern from School of Information, University of Michigan, in Winter 2009
- Focus on digital multimedia collections
- Interviewed 7 units and intern compiled results
- 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*

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

Digital Curation Planning

"Stop disciplining data and start herding it."

—Steve Bailey, *Managing the Crowd*

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 *Planning*?

"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

Other Digital Curation Initiatives

- Penn State
- Ohio State
- Duke University
- Yale University
- James Madison University

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

Metadata Comparison

Subject	X	X	Creation User Description	Keywords
Description	Description Classification Work Type Materials and Techniques	X	Extended Description	Caption Extracted Text Named Persons
Publisher		X	X	
Contributor		Contributing Institution	Other Contributors	Credit
Date	Date Date/Period *(creation or publication date)* Graduation Year Work Date	Date Original Date Range Date Digital	Creation Date Pub Date	Date Expiry Date
Type		Resource Type	Resource Type	
Format	Format Measurements Duration	X	X	File Format Image Size

Sample Findings: University Relations

- Public relations for Michigan State
- Example of unit holding digital content with permanent retention and serious storage issues
  - Digital photo and video content with historical value to the university
  - Storage servers full at more than 16TB of content
- Staff asked for help from Archives and IT

Sample Findings: University Relations

- Photos
  - Nikon RAW NEF, TIFF, JPEG formats
  - 21,000 images indexed in Extensis Portfolio
  - 5,100 publicly available through NetPublish Portfolio, 12,000 Zenfolio
- Video: *MSU Today* show, Big 10 Network
  - Shot in HD XCAM
  - Avid, Open Media Framework, QuickTime
  - MPEG-4 access versions on YouTube

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
- No magic bullet or one-size-fits-all solution
- Digital curation is part of larger university Records/Information Management Program
  - Must include policy and procedure framework

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

University Relations: An Example

- Unit creates two types of content :
  1. MSU Publications
  2. Digital content that documents history of MSU
- Highlights link between curating digital content and information/records management
  - Records inventory
  - Appraisal/selection guidelines
  - Format recommendations
  - Content management
  - Storage options

Digital Storage Solution Planning

- Evolving balance between central and local IT services (remember secretary pools?)
- 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 but tide is turning at MSU
- Central IT developing virtual server environments for local units

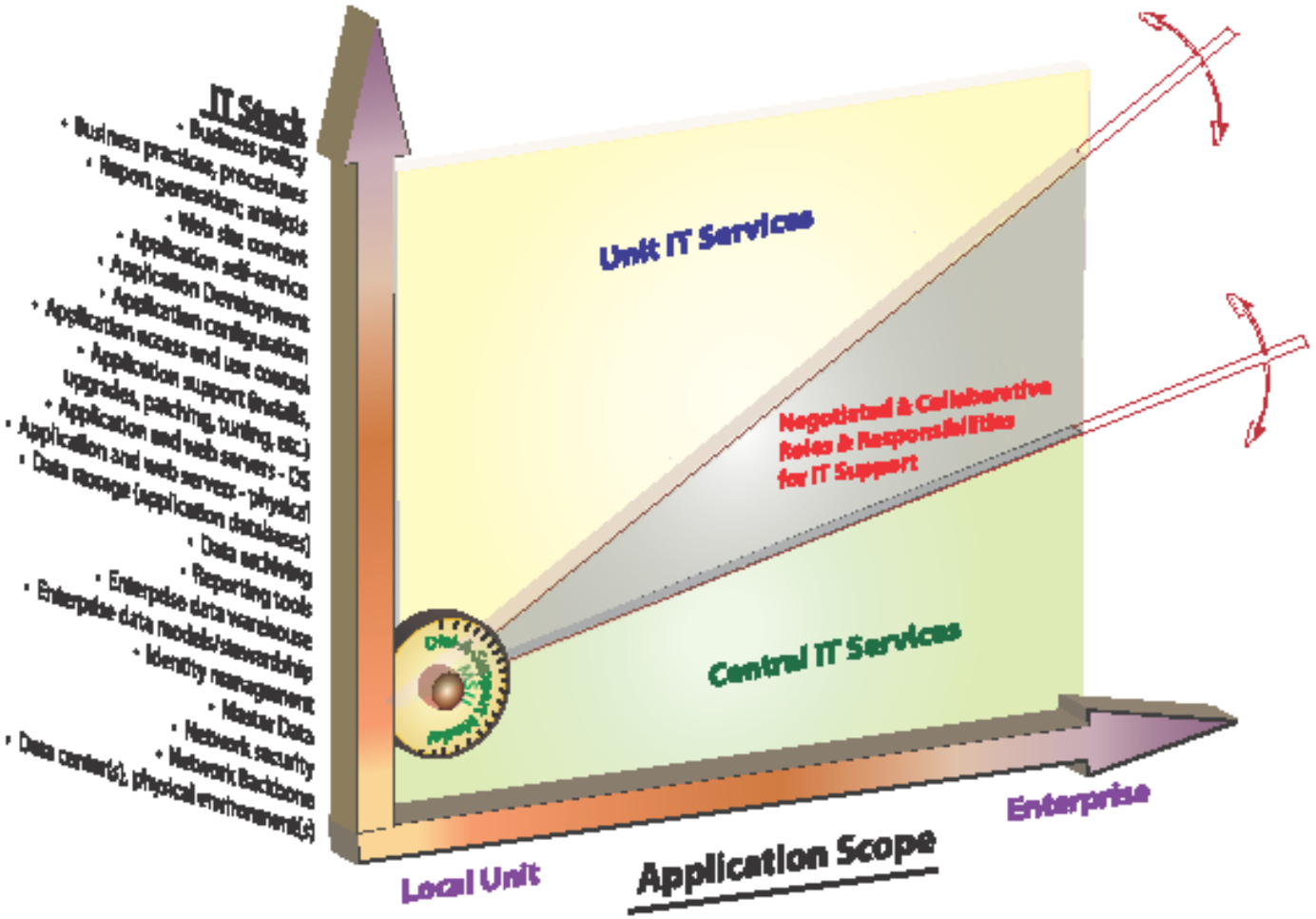
Digital Storage Solution Planning

- Tiered, 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 Archives
- DCP project helped make the case that building a larger “closet” is NOT a long term solution

Digital Storage and *Curation*

Content Type	Curation Needs	Functional Specifications
University Publications		
University records of historical value		
Non-university content		
Research Data		

DRAFT: MSU Application Support Model



Supporting Processes

1. IT and Data Change Control
2. IT and Data Configuration Management
3. Risk Management
4. Documentation
5. Trust Development
   - Training
   - Peer review and approval of access requests
   - Clear, shared notions of what parties will do and how will they do it.
6. Shared Governance
7. ARM Forms
8. Identify Lead Functional Groups

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

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/

Contact

Cynthia Ghering, ghering@msu.edu
Lisa Schmidt, lschmidt@msu.edu


MICHIGAN STATE
UNIVERSITY
University Archives
&
Historical Collections