University Archives & Historical Collections
Spartan Archive:  
*New Workflows for an Expanding Program*

Digital Preservation Tools and Workflows  
Cynthia Ghering  
Lisa Schmidt  
November 4, 2010
Overview

- Michigan State University and University Archives (UAHC)
- Spartan Archive Project
- Tools and Workflow in Transition: The “Digital Shelf”
- Tools to assess for use in Spartan Archive
- Conclusions
Michigan State University

- Founded in 1855 by Act of Michigan Legislature
- Nation’s pioneer land grant university
- 47,238 students / 4,985 faculty / 6,335 staff
- Tier One research university
  - National Superconducting Cyclotron Lab
  - Bio/computational Evolution in Action Consortium
  - Three medical colleges, industrial and organizational psychology
  - Elementary education, supply chain and logistics
University Archives (UAHC)

- Established in 1969 by MSU’s Board of Trustees
- Official repository of MSU’s permanent records
- Provides support and training in records management, storage, and retrieval
- Assists units in efficient administration and management of official records, active and inactive, and regardless of format (paper or electronic records)
Paper-based Tools and Workflows

- Procedures had changed little since the 1980s and 1990s:
  - Reactive collecting activities; no strategy
  - Out-of-date records retention schedules
  - Paper forms requiring signatures
  - MicroMARC: DOS collection management system
  - Digital files on removable media were either separated from collection and stored in file drawer or kept with collection and ignored
Spartan Archive Project

“Spartan Archive: An Electronic Records Archive at Michigan State University”

NHPRC Collaborative Electronic Records Projects grant

Three year project, Apr 2010 - Mar 2013

Funding allows UAHC to hire IT staff

Part of the vision for expanded program
Spartan Archive Project

- Develop electronic records archives for university’s born-digital records of enduring historical value
- Use proof-of-concept approach to test sustainability of archival solution for long-term access and preservation
Spartan Archive Project

- Capture four electronic records series produced by the Registrar’s Office:
  - Catalog of Academic Programs
  - Description of Courses
  - Annual Student Directory
  - Schedule of Courses
Spartan Archive Project

- 71,200 logical database records
- Vital records that document essential business functions
- Originally produced on paper, now entirely digital
Goals of Spartan Archive

- Develop a governance structure and technical infrastructure for the accession, access, and preservation of electronic records
  - Repository policies and procedures
  - Electronic records processing workflow
  - Institutional metadata standards
- Implement new collection management system
  - Replace MicroMARC with Archivists’ Toolkit
Goals of Spartan Archive

- Utilize iRODS for long-term storage and management
- Develop web access for researchers
  - Collection level description
  - Item level access
- Scale to include other university records of enduring value
Spartan Archive Conceptual Model
Tools and Workflow in Transition

- Why a “Digital Shelf”? 
  - Curation required for digital records of historical value 
    - Campus units retiring older material 
    - Content stored on removable media unstable and/or unreadable over time 
    - Some historical content stored in periodically purged CMS 
  - No public access system
Digital Shelf: Tech Infrastructure

- Shared network drive managed by service contract with central IT unit
  - Restricted to use of UAHC staff
- Servers in secure location across campus
- Daily incremental backups, weekly full backups to local SAN
- Twice weekly backups to tape
  - Taken to undisclosed campus location
Digital Accession: Administration I

- **Transfer**
  - Removable media: CD, floppy disc, Zip disc
  - E-mail, download, FTP
- **Create Accession Record**
  - Use information from transmittal form to create record in MicroMARC (collection management)
  - Basic metadata
- **Paperwork!**
# Digital Accession: Administration I

<table>
<thead>
<tr>
<th>Location</th>
<th>Record Group</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Donor or Department</th>
<th>Received from if different from above</th>
<th>Date Received</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Archives Staff Member</th>
<th>Collection Title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BOX</th>
<th>FOLDER</th>
<th>FILE FOLDER TITLE</th>
<th>DATES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|     |        |                   |       |
|     |        |                   |       |

|     |        |                   |       |
|     |        |                   |       |

11/10/10
Digital Accession: Administration I
Digital Accession: Administration II
Digital Accession: Ingest

- Insert removable media into PC
- Run virus scan
  - Tool: Kaspersky
- Examine and appraise files
  - Tool: Gemulator for reading Mac files on PC
  - Appraisal criteria depends on collection, tech
    - Ex., QuarkExpress files for printed publication may not be preserved
Digital Accession: Ingest

- Copy “good” files that meet appraisal criteria to item folder
- Normalize files as necessary
  - Ex., MacWrite to MS Word with MacLinkPlus
  - Ex., MS Word to PDF/A
  - Keep original bitstream (?)
Digital Accession: Administration III

- Remove media from PC
- Label w/ accession number, title (optional), virus notice (if applicable)
- Copy “good” files to new CD (?)
- Place all media related to accession into folder and file in drawer
Between “Shelf” & Spartan Archive

- Ingest all removable media
- Limit staff access
- Virus check deposits in pre-ingest folder
- Streamline administrative workflow
- Implement Archivists’ Toolkit, some tools
  - Metadata extraction
  - Integrity checking
- Document policies
Ingest Tools: Spartan Archive

- Metadata extraction
  - JHOVE
  - File Information Tool Set (FITS)
- BagIt
- DSpace
  - Ingest workbenches
Ingest Tools: Spartan Archive

- Tufts Accessioning Program for Electronic Records (TAPER)
  - Automated accessioning
  - Submission agreements
- Software-Independent Archiving of Relational Databases (SIARD)
Other Tools: Spartan Archive

- Integrated Rule-Oriented Data System (iRODS)
- Web interface
- Fedora
  - Hydra content management framework
- Archivematica
Conclusions

- MSU Archives between Stage 2 and Stage 3 of the 5 stages of digital preservation
  - Digital shelf as response to current threats of loss of digital content
  - Spartan Archive started as project, will scale to digital preservation program
    - Comprehensive policy framework
    - Technology investment and infrastructure
    - Sustainable funding