MICHIGAN STATE
UNIVERSITY

Final Report
MSU Digital Curation Planning Team
July 1, 2009 – June 30, 2010

July 26, 2010

Table of Contents

Digital Curation Planning Team Members

Cynthia Ghering | *Director, University Archives & Historical Collections* | ghering@msu.edu
Lisa Schmidt | *Electronic Records Archivist, MATRIX & University Archives* | lschmidt@msu.edu
Nancy Fleck |*Assistant Director, Technical Services, MSU Libraries* | fleckn@msu.edu
Catherine Foley |*Digital Preservation Librarian, MATRIX* | Catherine.Foley@matrix.msu.edu
Traci Gulick |*Associate Registrar for Student Academic Records* | Gulickt2@msu.edu
Shawn Nicholson |*Assistant Director, Digital Information, MSU Libraries* | nicho147@mail.lib.msu.edu
Cindy Straus |*Knowledge Repository Manager, MSU Extension & ANR* | strausc@msu.edu
Don Ries | *Director, Enterprise Information Stewardship* | dries@msu.edu

Executive Summary

Michigan State University, like other research universities and large organizations, has amassed a growing body of digital assets and information. Much time, effort, grant funding, human capital and research has gone into creating MSU's digital resources—some of which *only* exist in digital form. Without an active, well-considered long-term plan for managing and preserving these resources, they will eventually become inaccessible due to technology obsolescence and digital media fragility, and could cost researchers time and frustration searching for data or trying to locate files for sharing. There are also digital storage limitations due to the increasing file sizes and the increasing volumes of files in general. In response to the need to manage digital content, some MSU colleges and departments have started their own digital repositories. MSU has no comprehensive, campus-wide digital preservation strategy or set of guidelines to date and the university has not established an "institutional repository"—commonly defined as "digital collections that capture and preserve the intellectual output of university communities."[1]

The Digital Curation Planning Team leveraged earlier research on preserving the university's digital multimedia collections to design the project, which included a campus-wide, web-based, self-selective questionnaire, a series of interviews with select units, analysis of the results, and suggestions for next steps in providing digital curation guidance. Parallel activities included a preliminary analysis of the University Relations unit's digital photos and video and new university-wide approaches to digital storage, including "tiered storage" options. During the course of the project, similar initiatives at other universities were launched, several institutions expressed an interest in MSU's methodologies, and opportunities for partnering with other members of the Committee on Institutional Cooperation (CIC) to develop common digital curation guidelines and tools were presented.

The interviewed units demonstrated curation practices that trended towards the positive. Some problems were identified, however, and these include:

- Backups located too close to production servers
- Maintenance of preservation copies of files
- Low incidence of file integrity checking
- Inconsistent use of file naming conventions
- Minimal creation of descriptive and administrative metadata
- Minimal documentation of digital curation policies

As revealed by the interviews, the types of digital content at Michigan State University and the needs and expectations of content creators and users vary. Development of common digital curation guidelines for the university will be an iterative process. Four types of digital content identified, each with its own curation considerations, include:

1. **MSU digital publications, including e-journals and electronic theses and dissertations.** In planning for the curation of this material, the university should consider establishing an institutional repository and exploring the use of a distributed preservation solution.
2. **Digital content that documents the history of MSU**, which requires guidelines for assessment and curation as well as mechanisms for its transfer and storage with UAHC.

3. **Non-MSU-specific digital content**, which would benefit from general curation guidelines.
4. **Research data** generated by the university.

Suggested next steps in digital curation planning at Michigan State University include:

Digital Content Assessment, Transfer, and Storage

- Assistance in assessment of digital content, especially for those units holding digital assets of historical value that should be transferred to UAHC for long-term preservation
- Development of tiered storage plans
- Development of transfer workflows to UAHC for digital content of archival value

Toolkits

- Guidelines or decision trees that help units determine if digital assets should be transferred to UAHC for permanent preservation
- Digital/data curation toolkits that offer guidance in file formats, documentation, intellectual property rights, sharing/dissemination, and preservation

Good Practice

- Development of general best and good practices in digital curation recommendations and guidance, keeping in mind differences among unit missions and types of digital material
- Creation of a Digital Preservation Officer position at MSU to coordinate curation and preservation resources across campus for both academic and administrative data types

Collaborations

- Foster "Communities of Practice," in which MSU units share digital curation experiences, generate new ideas, and collaborate on initiatives
- With other CIC member institutions, obtain grant funding to explore the digital curation problem across institutions and develop common best and good practices guidelines

What is Digital Curation?

In keeping with current trends in the field of electronic management of digital resources, the team chose to use the term "digital curation" rather than "digital preservation" in the name of the project. Digital curation includes preservation but also takes into account the life cycle of the data so as to encompass its creation and management. As defined by the Digital Curation Centre: "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."[2]

By offering practical guidance and influencing policy, MSU's archivists and librarians would expand on their traditional role as custodians of physical material. As stated by Richard Cox in a 2008 *EDUCAUSE Review* article: "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."[3] MSU's archivists and librarians can provide guidelines and best practices in digital preservation and management while the material itself remains in the custody of the creating units. To quote Steve Bailey from his book *Managing the Crowd: Rethinking Records Management for the Web 2.0 World,* the focus would be to "stop disciplining data and start herding it,"[4] to provide guidelines rather than to issue mandates.

Project Background

Recognizing the need for guiding the management and preservation of MSU's digital assets, MSU's Office of the Vice Provost of Libraries, Computing and Technology (LCT) approved a year-long Digital Curation Planning Project. In July 2009, a collaborative team led by the University Archives & Historical Collections (UAHC), the MSU Libraries, and MATRIX: Center for Humane Arts, Letters, and Social Sciences Online embarked on this project to explore and evaluate the university's existing digital content and curation practices. This final project report describes the following:

- The methodology and results of a self-selective, campus-wide survey of digital assets and technological infrastructures using a web-based questionnaire
- The methodology, results, and analysis of a series of one-on-one interviews with select units regarding their digital curation practices
- Results of interviews with MSU University Relations regarding curation of digital photographs and video, including content assessment, transfer, and storage solutions
- A brief look at digital storage solution planning at Michigan State University
- Next steps in digital curation planning for the university

The team realized that it was not feasible to complete an exhaustive survey of all university digital assets within one year and that such a comprehensive inventory would result in diminishing returns due to inevitable redundancies. The team was also concerned about creating the perception that the project's desired outcome was to build a one-size-fits-all data repository that might result in individual units' loss of control of their digital assets. Instead, a

campus-wide, self-selective web-based questionnaire provided a sampling of digital assets and repositories. In-depth interviews were conducted with select units, with consideration given to learning more about the digital repository solutions already implemented across campus. Technical infrastructures, storage needs, metadata schemes, and file naming conventions were also reviewed.

During the course of the Digital Curation Planing Project, a project website (http://msudcp.archives.msu.edu) was created to host information about project developments.

Survey Tools

Baseline Data Questionnaire

In October 2009, the MSU Digital Curation Planning Team designed and conducted a voluntary, web-based survey using a baseline data questionnaire created with the SurveyMonkey online survey tool. The questionnaire was publicized to potential participants through MSU's internal IT Exchange network, the weekly *MSU News* online newsletter, and the project website. Survey participants were asked questions about the types of digital content created by their units, the approximate volume of content, storage media used, file formats, online storage capacity and expansion plans, and the use of content management systems and/or digital repository software. Refer to Appendix 2 for the full questionnaire.

The questionnaire resulted in 90 responses from academic departments, administrative services units, research centers/institutes, and technology services units across the university. Types of digital content making up the largest proportion of a given unit's content varied considerably, as did the file formats created. Nearly all of the participating units store digital content on hard drives. Seventeen of them plan to increase online storage capacity in the near future, with 6 units planning to add 30 TB or more. Twenty-three units responded that they have implemented or plan to implement content management system and/or digital repository software. Note that neither "content management system" nor "digital repository" was defined in the questionnaire, so unit representatives responded based on their understanding of the terms. Refer to Appendix 3 for more details of questionnaire results.

One-on-One Interviews

Ten of the units that responded to the questionnaire were chosen for in-depth, one-on-one interviews, with a focus on those that had reported use of a content management system and/or digital repository solution. The team was also interested in talking to offices that created digital content documenting MSU history or generated university records that fell under existing institutional records retention schedules. Units chosen for interviews included Broadcasting Services, the Center for Research on Mathematics and Science Education (CRMSE), the Confucius Institute at Michigan State University (CIMSU), the Department of Art & Art History, the Department of Theatre, MATRIX, MSU Extension/Agriculture and Natural Resources (ANR) Technology Services, the National Superconducting Cyclotron Laboratory (NSCL), Physical Plant Division, and the Turfgrass Information Center (TIC) at the MSU Libraries.

Team members Schmidt, Ghering, and Nicholson conducted the one-on-one interviews during the months of January, February, and March 2010. Discussions included how their digital content relates to the mission of the unit; whether it was of ongoing use or of archival value; the file formats used; and the storage infrastructure, including any space issues. Regarding the unit's content management system and/or digital repository, the team asked about the system(s) used, why they were chosen, and how they are used. The team also asked about processes and workflows of ingesting or adding data and/or digital content into the system, archival storage/preservation of content, and content retrieval. Finally, the units were asked about the use of metadata stored with or related to the content as well as file naming conventions. (See Appendix 4, "Interview Ticklers" for the types of questions asked. Refer to Appendices 5-14 for detailed reports on each interview.)

Analysis of Survey Findings

In analyzing the results of the interviews, several observations were made. Each of the units had devised solutions that fit the mission of the unit, the nature of the data, and the needs of users—similar to what the Blue Ribbon Task Force on Sustainable Digital Preservation and Access refers to as "context-specific attributes."[5] Some units use commercial applications, and some use open-source software. The TIC at the MSU Libraries, for example, has long used the commercially available Cuadra STAR QuadraStar Library database/content management system,[6] and the Department of Theatre uses the relatively new open-source ResourceSpace digital repository solution.[7] Some units—such as Broadcasting Services—hold digital content of archival value to the university. Other units create and store digital material produced on behalf of partner cultural organizations, such as the Department of Art & Art History's digital artwork and MATRIX's digital libraries—important types of digital resources, but not identified as institutional records that should adhere to a defined retention schedule.

The interviewed units exhibited many positive trends in the areas of digital preservation and curation in keeping with guidelines put forth by the Digital Curation Centre's *Curation Reference Manual*,[8] the Center for Research Libraries (CRL) and Online Computer Library Center Inc. (OCLC)'s *Trusted Repositories Audit & Certification (TRAC): Criteria and Checklist*,[9] the Digital Preservation Coalition's *Digital Preservation Handbook*,[10] the United Nations Educational, Scientific and Cultural Organization (UNESCO) *Guidelines for the Preservation of Digital Heritage*,[11] and other resources. First, most units backed up their data in some manner. Many of them demonstrated sound use of metadata, and many were using repository software with good access and discovery interfaces to manage their content. Importantly, many of the units have strong support from their administrative management and stable funding.

Nearly all of the interviewed units store preservation (archival) masters of at least some of their content. MATRIX maintains TIFF files of images and preservation masters of audio and video content that has been converted to access formats for use in the KORA digital repository.[12] The TIC stores scans of printed material and slides as TIFF files, but makes them available online in PDF and JPEG formats respectively. Likewise, the Department of Art & Art History keeps TIFF master files while providing JPEG files as access copies in its Visual Resources Library (VRL).[13] The Department of Theatre, on the other hand, has chosen to convert digital photos from the original RAW format to JPEGs for use in its DOT::Media repository. Preservation masters of MSU Extension's bulletins are stored as TIFF files in a dark archive at the MSU Libraries, with PDF

versions available through its Knowledge Repository.[14] Both Broadcasting Services and the Confucius Institute maintain some audio files in the archival WAV format. Physical Plant currently stores and scans documents in the TIFF format, but would like to move to PDF/A as a preservation master format. CRMSE wishes to convert data sets to XML and survey instruments to PDF/A files for long-term preservation.

Only three units shared their means for verifying file integrity, a key digital curation practice. MATRIX's KORA repository software includes a message digest algorithm that can generate a unique number for an ingested file and then periodically check that number; any change would indicate that the file had become corrupt. Through the use of Adobe Bridge, the Department of Art & Art History can detect file corruption when viewing thumbnail photos; likewise, the Photoshop script that creates contact sheets of thumbnails will stop running if it encounters a corrupt file. In the near future, the Department of Theatre will add a file integrity test to the code for the DOT::Media repository, accompanied by the capability to store parity files with the original files to simplify restoration in case of corruption.

MSU Extension and the Department of Art & Art History have formal file naming conventions in place. Physical Plant's Meridian system[15] automatically assigns file names that include the project number, document type, and a brief, metadata-based code. Although the Cyclotron has a systematic method of assigning project numbers to file directories for each experiment, researchers have some latitude in naming the actual data files. MATRIX develops file naming conventions with its partners on a project-by-project basis.

Most of those interviewed expressed interest in appraisal and curation guidelines, and they could use guidance: Although these units back up their data, most of the backups tend to be located very close to production servers—often in the same building, if not the same room. The high incidence of maintenance of preservation copies is encouraging, but not practiced by all units and for all file types. Alternatively, the practice of checking file integrity is disappointingly low. Some of the units create only minimal metadata for their digital content, and the team found little in the way of documented digital curation policies. The lack of good digital curation practice is all the more unfortunate considering that out of the entire MSU campus, the interviewed units are more likely to have invested in digital asset management. The team suspects that the majority of campus units are either unaware of or unable to address digital curation at this time.

Schmidt also performed a comparison of metadata schema used among the units interviewed with the Dublin Core metadata element set,[16] a standard in the information science field for describing digital objects. Six of the ten units interviewed had metadata schema to share. MSU Extension, MATRIX, and the Department of Theatre use metadata schema based on or similar to Dublin Core, with slight variations to reflect local needs. The Department of Art & Art History uses the Image Resource Information System (IRIS) cataloging utility for description of its art images with metadata based on the Visual Resources Association (VRA) Core[17] and the Cataloging Cultural Objects (CCO)[18] guide to good cataloging practices; this metadata maps roughly to Dublin Core. On the other hand, the metadata used by Physical Plant and the TIC does not correspond directly to the Dublin Core metadata set. Physical Plant utilizes the metadata customization capabilities of the Meridian facilities assets management system to specify its own locally controlled vocabulary suited to project transactions. Likewise, the

Turfgrass Information Center uses its own indexing terms specified in the QuadraStar system for cataloging bibliographic information of all things turfgrass. (See Appendix 15, "Metadata Comparison.")

University Relations: An Appraisal Prototype

In parallel with the Digital Curation Planning Project interviews, UAHC held interviews with members of several University Relations groups—including University Photography and Videography, the Big Ten Network office and the Media Communications office. University Relations, which creates and maintains digital photographic and video records, is quickly filling server space and expressed an urgent need for help in appraising its digital content. As with the units interviewed for the Digital Curation Planning Project, UAHC expected that curation guidelines resulting from the University Relations analysis would be applicable to digital assets held by other units on campus.

University Relations stores hundreds of thousands of digital photos on 4.6 TB of networked servers. Most of these photos are in the Nikon RAW NEF, TIFF, and high-resolution JPEG formats, and some of the images have embargoes and other use restrictions. Approximately 21,000 images are indexed in the Extensis Portfolio media management system, and 5,100 are publicly available through NetPublish Portfolio.[19] In 2010 University Relations began using Zenfolio[20] as an online delivery system for client work, with more than 12,000 images currently available for clients and the public to view. As many of the photos and videos document current events at the university, these records have long-term historical value and, per university retention schedules, should be transferred to the custody of UAHC. Other files, such as duplicate images and "outtakes," have little business or historical value and should be disposed of to free up server space.

In terms of digital video content, University Relation holds large volumes of footage, including the *MSU Today* show on the Big 10 Network. This show is shot in HD XDCAM, and episodes run 30-60 minutes. File formats include Avid, Open Media Framework, and QuickTime, with Avid used for editing and QuickTime for access copies. There are 16 TB of proprietary non-networked dedicated workspace for HD production only, 6 TB of non-networked HD production storage, 4 TB combined internal workspace across four HD editing stations, and 2 TB of networked "temporary storage " used during video production and for distribution. All programs are closed captioned and supplied to the Big Ten Network for public broadcast. MPEG-4 versions of full shows and segments are uploaded to YouTube for public access, and copies of the shows are sent to contributors, the Office of the Provost, and the Office of the President. University Relations keeps two edited versions of each program: one show master, including text overlays, and one clean sub-master. Most of the shows are used within six months of creation.

Later in 2010, University Relations' digital photos and video will be inventoried, permanent archival records identified, retention schedules created, and selection guidelines recommended for new records as they are created. Metadata structures and file naming conventions will also be proposed to facilitate the day-to-day media management and the future transfer of files to the University Archives for long-term preservation. In addition, the Digital Curation Planning Team will work with Administrative Technical Services (ATS) to develop storage, transfer, and access solutions for the files taken into UAHC's custody as well as those flagged for storage at

University Relations. This will include preservation space for archival masters of the digital photos and public access space for database and low-resolution files.

Digital Storage Solution Planning at Michigan State University

Currently, LCT is developing virtual server environments and price structures as a storage solution to local units. LCT is also looking at the idea of providing "tiered storage" options, which entail a variety of storage types or levels to meet a diverse array of needs. The storage solution would be tiered depending on the nature of the content. For example, storage space for files with temporary, short-term use might be provided locally while capacity and infrastructure efficiencies would be leveraged by developing a centralized, shared long-term preservation storage environment.

Within the Committee on Institutional Cooperation (CIC) community, collaborative storage solutions that would better enable effective, efficient stewardship of campus assets and cutting-edge scholarship are in development. The resulting common architecture, infrastructure, and operating environment would increase economies of scale, permit shared management, and provide for the research and the development of such value-added services as community tagging and annotation, citation tracking, and digital curation.

Next Steps in Digital Curation Planning

During the course of the one-on-one interviews with the select MSU units, it became clear that the types of digital content and the needs and expectations of content creators and users of even the small sampling of units interviewed are quite varied, and there is still much to learn about digital curation practices at MSU. The team realized that the development of common digital curation guidelines for the university will be an iterative process. The project has raised awareness across campus and helped shape the next steps in digital curation planning, including the development of digital asset transfer and storage mechanisms and the establishment of a repository for some institutional records.

One outcome of this initial phase of the project has been the identification of four types of digital content at Michigan State University. These may be addressed as follows:

- **MSU publications, including e-journals and electronic theses and dissertations.** In planning for the curation of this digital material, the university should strongly consider establishing an institutional repository and should explore the use of Controlled LOCKSS (CLOCKSS)[21] or another Lots of Copies Keep Stuff Safe (LOCKSS)[22]-based distributed solution to preserving the content.
- **Digital content that documents the history of MSU**, such as the photos and video of University Relations and some of Broadcasting Services' audio and video programming. This material requires digital curation and appraisal guidelines, as well as mechanisms for its transfer to and storage with UAHC.
- **Non-MSU-specific digital content**, such as that created and managed by MATRIX and the Department of Art & Art History. This type of digital material would benefit from general curation guidelines.

- **Research data** generated by the university. Beginning in October 2010, the National Science Foundation (NSF) will require that grant proposals include a data management plan that addresses preservation and access.[23] For this purpose and to ensure good stewardship of these assets, creators of research data require access to information on Michigan State's existing curatorial infrastructure along with clear, easily understood guidance on data management best practices.

Suggestions for next steps in ensuring optimal management and preservation of MSU's digital assets, many of which correspond to the requirements of the four types described above, include the following:

Digital Content Assessment, Transfer, and Storage

- Provide assistance in the assessment of digital content to university departments and units, especially those holding digital assets of historical value that should be transferred to UAHC for long-term preservation, such as University Relations and Broadcasting Services.
- Work with LCT on the development of tiered storage plans.
- Work with LCT on plans for transferring digital content of archival value to UAHC.

Toolkits

- Develop guidelines to quickly determine if digital assets should be transferred to UAHC for permanent preservation.
- Develop digital/data curation toolkits that acknowledge researchers and units as information producers as well as consumers. Topics could include file formats, documentation, intellectual property rights, sharing/dissemination, and preservation.

Good Practice

- Develop general best and good practices in digital curation recommendations and guidance, keeping in mind the differences in the missions of the units and the types of digital material that they create and manage.
- Encourage the funding of a Digital Preservation Officer position at MSU to coordinate curation and preservation resources across campus for both academic and administrative data types. A dedicated individual at this level could raise the visibility of the need for digital curation and direct the dissemination of guidelines and best practices.

Collaborations

- Foster "Communities of Practice" through online forums and meetings, in which MSU campus units and other institutions have the opportunity to share digital curation experiences, generate new ideas, and collaborate on initiatives.
- Work with other CIC member institutions to obtain grant funding to explore the digital curation problem across institutions and develop common best and good practices guidelines.

Beyond Michigan State University

Scholarly presentations on the project will continue through the summer of 2010, and the manuscript for an article has been submitted to the scholarly journal *Library Resources and Technical Services (LRTS),* with a tentative publication date of April 2011. Several other universities have recently launched projects similar to MSU's Digital Curation Planning Project, and some institutions made inquiries about MSU's project methodologies. A program officer from the National Endowment for the Humanities (NEH) has asked that MSU partner with other CIC institutions to submit a grant on developing common curation guidelines and tools, and MSU is spearheading this effort with the goal of submitting a grant proposal in 2011.

During the course of the project, the Digital Curation Planning Team discovered that several other universities have embarked on their own digital curation projects—including Pennsylvania State University, Ohio State University, Duke University, Yale University, and James Madison University. Other institutions made inquiries regarding MSU's project methodologies, including the Smithsonian Institution, the University of Utah, and Notre Dame University.

Notes

[1] Crow, Raym, "The Case for Institutional Repositories: A SPARC Position Paper," The Scholarly Publishing & Academic Resources Coalition (SPARC), 2002, http://scholarship.utm.edu/20/1/SPARC_102.pdf. Retrieved 28 June 2010.
[2] Digital Curation Centre, DCC Charter and Statement of Principles, http://www.dcc.ac.uk/about-us/dcc-charter. Retrieved 9 April 2010.
[3] Cox, Richard, "The Academic Archives of the Future," *EDUCAUSE Review*, Vol. 43, No, 2, March/April 2008, http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume43/TheAcademicArchivesoftheFuture/162683. Retrived 3 June 2010.
[4] Bailey, Steve, *Managing the Crowd: Rethinking Records Management for the Web 2.0 World* (London: Facet), 2008.
[5] Blue Ribbon Task Force on Sustainable Preservation and Access, "Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information," Final Report, February 2010, p. 30, http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf. Retrieved 7 April 2010.
[6] Cuadra STAR QuadraStar content management system, http://www.cuadra.com/products/products.html. Retrieved 14 May 2010.
[7] ResourceSpace digital asset management system, http://www.resourcespace.org/. Retrieved 14 May 2010.
[8] Digital Curation Centre, Curation Reference Manual, http://www.dcc.ac.uk/resources/curation-reference-manual. Retrieved 21 June 2010.
[9] The Center for Research Libraries (CRL) and Online Computer Library Center Inc. (OCLC), *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, Version 1.0, February 2007, http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf. Retrieved 2 June 2010.
[10] Digital Preservation Coalition, Digital Preservation Handbook, http://www.dpconline.org/advice/digital-preservation-handbook.html. Retrieved 22 June 2010.
[11] National Library of Australia, *Guidelines for the Preservation of Digital Heritage*, United Nations Educational, Scientific and Cultural Organization (UNESCO), March 2003, http://unesdoc.unesco.org/images/0013/001300/130071e.pdf. Retrieved 21 June 2010.
[12] KORA digital repository software application, http://www2.matrix.msu.edu/research/technology/kora/. Retrieved 12 May 2010.
[13] Visual Resources Library (VRL), Department of Art & Art History, Michigan State University, http://www.art.msu.edu/?page_id=140. Retrieved 23 June 2010.
[14] MSU Extension Knowledge Repository, http://www.msue.msu.edu/portal/.
[15] InnoCielo Meridian Enterprise engineering content management solutions, http://www.cyco.com/products/ice/. Retrieved 14 May 2010.
[16] Dublin Core Metadata Element Set, Version 1.1, http://dublincore.org/documents/dces/. Retrieved 14 May 2010.
[17] Visual Resources Association (VRA) Core data standard, http://www.vraweb.org/projects/vracore4/. Retrieved 14 May 2010.
[18] Cataloging Cultural Objects (CCO) http://www.vraweb.org/ccoweb/cco/about.html. Retrieved 24 May 2010.
[19] Extensis, a Celartem Company, http://www.extensis.com/en/home.jsp. Retrieved 24 May 2010.
[20] Zenfolio, http://www.zenfolio.com/.
[21] Controlled LOCKSS (CLOCKSS), http://www.clockss.org/clockss/Home/. Retrieved 23 June 2010.
[22] Lots of Copies Keep Stuff Safe (LOCKSS), http://lockss.stanford.edu/lockss/Home. Retrieved 25 May 2010.
[23] National Science Foundation, "Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans," Press Release 10-077, 10 May 2010, http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928&org=NSF. Retrieved 2 June 2010.

Appendix 1

MSU Digital Preservation Proposal – April 2009

Project: Preserving MSU's Digital Assets

I. Project Overview

Like other research universities, Michigan State University has amassed a growing body of digital information, including institutional records; faculty and student research, theses, and dissertations; university publications; multimedia collections; collections of digital surrogates of cultural material; learning objects; and more. Units across the MSU campus have started to create their own repositories of some of this material, but there is currently no comprehensive preservation strategy for the university's growing collection of digital assets.

Much time, effort, grant funding, human capital, and research has gone into creating MSU's digital resources—some of which *only* exist in digital form. Without an active, well-considered plan for managing and preserving these resources, they will eventually become inaccessible due to the ever-changing nature of technology.

MSU must develop a preservation plan rooted in current best practices that will provide trustworthy stewardship of the university's digital assets and intellectual property. We propose, through collaboration of the MSU Libraries, University Archives, and MATRIX, and with the support of the Office of the Vice Provost of Libraries, Computing and Technology, to engage a digital preservation analyst to develop a digital preservation plan with a focus on practical solutions using the resources MSU has in place right now.

The digital preservation plan will include:

- An environmental scan of the university's current digital assets and an estimate of the potential creation rate of new digital material in the next five years. (*Note: This process has already begun with a spring 2009 digital preservation internship project focused on MSU's digital multimedia collections. The project report will be distributed in early May.)*
- Survey of MSU's existing digital repositories and technical infrastructure across departments and units.
- The application of best preservation, management, and access practices to the university's digital collections. Using tools provided by Nancy McGovern and the Inter-University Consortium for Political and Social Research (ICPSR) and Cornell University's digital preservation program (see http://www.icpsr.umich.edu/dpm/ for details), we will determine institutional readiness for long-term preservation of digital resources through a "gap analysis."
- Development of policies, procedures, and work flow to standardize MSU's approach to digital asset management and preservation across departments and units.

- Potential collaborations with other institutions and consortia—including HathiTrust, Lots of Copies Keeps Stuff Safe (LOCKSS), Controlled LOCKSS (CLOCKSS), Portico, and the Michigan Library Consortium (MLC)—to share resources and services such as distributed storage and archiving.

- The use of emerging standards for digital repositories, including the Trusted Repositories Audit & Certification (TRAC): Criteria and Checklist assessment tool and the Open Archives Initiative-Preservation Metadata Harvesting (OAI-PMH) and OAI-Object Reuse Exchange (ORE) protocols for exposing metadata and participating in data aggregations.

II. Project Scope

The project will consider all of the digital assets and collections at MSU for inclusion in the digital preservation plan, such as:

- Content collections—including audio, video, and print
- Institutional records, including business records
- Campus/university journals and other publications
- Digital surrogates of historical and cultural collections
- Podcasts and vodcasts
- Websites
- E-mail
- Research—including data, outputs, and working papers
- Electronic theses and dissertations (ETDs)
- Library instructional media
- Learning objects
- Scholarly journals

III. Deliverables

The project includes the following deliverables:

- Project work plan and timeline
- Survey of digital content (electronic records, multimedia collections, digital surrogates, and more) created by MSU
- Scan of current technical infrastructure for digital collections at MSU
- Assessment of institutional readiness, including a gap analysis of the current environment to comprise the three legs of digital preservation—technology, organization, and resources[1]
- Five-year preservation plan for MSU's digital assets, to include:
  —Preservation, management, and access strategies for each digital asset type
  —Implementation strategy with achievable phases of implementation
  —Policies and procedures that support digital asset management and preservation

[1] Digital Preservation Management Workshop, November 2008. MAC Fall Symposium.

IV. Timeframe
The digital preservation plan will take approximately 12 months to develop and begin to implement. The project work plan will detail project milestones. Implementation will be staged over a five-year period. The digital preservation analyst will review project status on a monthly basis with a core working group including representatives from MSU Libraries, University Archives, and MATRIX. In addition, the digital preservation analyst and working group will interview and solicit feedback from university stakeholders (see Section V: Potential Collaborators and/or Content Providers) during this initial stage of exploration and development.

V. Potential Collaborators and/or Content Providers
The following MSU organizations and groups create and currently host digital content. Meetings will be scheduled with them to arrange for the inventorying of their current digital content, projected future content, and to review their current and planned technical infrastructures.

- MSU Archives
- MSU Libraries
- MATRIX
- MSU Museums
- MSU Collections Committee (chaired by Kurt Dewhurst and Susan Bandes)
- MSU Extension
- Sports Broadcasting Services
- University Relations
- Virtual University Design and Technology (VU-DAT)
- WKAR
- Colleges and schools
- Administrative Information Services (AIS)
- Academic Technology Services (ATS)
- Enterprise Business Systems Project (EBSP)
- Institution for Cyber-Enabled Research (ICER)
- MSU Press
- Campus institutes
- Area studies centers

VI. Digital Preservation Analyst
Focused on their own needs and resources, the IT staff and content creators of MSU's various units and departments have either postponed critical preservation decisions due to lack of resources and staff or devised short-term asset management solutions that do not have long-term viability. The university-wide scope of this proposal requires the dedicated focus of an individual who can broadly view and consider the needs and resources of all campus units managing digital assets.

This individual must have a working knowledge of digital preservation issues and best practices. Training in the design and implementation of a digital preservation strategy such as that provided by ICPSR's Digital Preservation Management Workshop is highly recommended. The ICPSR approach addresses the development of policy, technology, and resource frameworks for planning and building a

successful digital preservation program. Familiarity and practical experience with important emerging preservation tools and standards such as the Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist is also a requirement.

With this knowledge and experience, the digital preservation analyst will be well-equipped to inventory and access the university's digital resources and technical infrastructure and develop a digital preservation plan that leverages common requirements and services while allowing appropriate customization to meet the needs of each unit and digital asset type.

The digital preservation analyst must hold a Master's degree in library or information science from an American Library Association-accredited institution with a focus in electronic records and digital preservation. In addition, the digital preservation analyst must have at least two years of practical experience and a working knowledge of digital preservation practices, emerging standards, repository software, digitization, and intellectual property issues.

VII. Future State – an Institutional Repository for MSU?

As Clifford Lynch states in a seminal report on institutional repositories, "…an institutional repository is a recognition that the intellectual life and scholarship of our universities will increasingly be represented, documented, and shared in digital form, and … a primary responsibility of our universities is to exercise stewardship over these riches: both to make them available and to preserve them."[2] The *Census of Institutional Repositories in the United States* published February 2007 indicated that large numbers of educational institutions have moved forward with building institutional repositories, with 94.5% of the 83 research universities that participated in the survey in the planning or implementation stages; 62.5% of them had begun implementation.[3] A 2008 survey of Committee on Institutional Cooperation (CIC) members by the CIC University Archivists Group found that 12 of the 13 CIC institutions—all but MSU—have an institutional repository.

By immediately addressing critical digital preservation needs on campus, MSU can better position itself for a future institutional repository that provides for the university-wide collection, management, and preservation of valuable digital resources while ensuring their availability to current and future faculty, students, researchers, and the public. An institutional repository may offer the potential to control collection development costs and curb data storage expenses across campus. If time allows, the digital preservation analyst may also explore the potential utility of an institutional repository for MSU by:

---

[2] Lynch, Clifford, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age, Association of Research Libraries (ARL) bimonthly report, no 226, February 2003, http://www.arl.org/resources/pubs/br/br226/br226ir.shtml. Retrieved 19 January 2009.

[3] Markey, Karen, Soo Young Rieh, Beth St. Jean, Jihyun Kim, and Elizabeth Yakel, *Census of Institutional Repositories in the United States*, MIRACLE Project Research Findings, Council on Library and Information Resources (CLIR), Washington, D.C., February 2007, 18, http://www.clir.org/pubs/reports/pub140/pub140.pdf.

- Benchmarking early adopter institutions' experiences with planning and implementing institutional repositories.

- Examining issues such as:
  —Intellectual property management practices that better protect the rights of contributing faculty and students.
  —Promotion of the repository to contributors and potential users.
  —Publishing, distribution, and other possible value-add services.

- Investigating the development of a CIC-wide collaborative institutional repository.

VII. Bibliographic Sources

- Census of Institutional Repositories in the United States (CLIR), http://www.clir.org/pubs/reports/pub140/pub140.pdf
- Cornell University Digital Preservation Policy Framework, http://commondepository.library.cornell.edu/cul-dp-framework.pdf.
- *Creating an Institutional Repository: LEADIRS Workbook* (Mary R. Barton, MIT Libraries), http://www.dspace.org/implement/leadirs.pdf
- Digital Preservation Management Workshop (ICPSR), http://www.icpsr.umich.edu/dpm/
- *Identifying Factors of Success in CIC Institutional Repository Development* (CIRSS, University of Illinois at Urbana-Champaign), http://www.cic.net/Libraries/Reports/PalmerEtAlMellonReport.sflb.
- *The Institutional Repository* (Richard Jones, Theo Andrew, John MacColl; Chandos Publishing)
- *The Research Library's Role in Digital Repository Services: Final Report of the ARL Digital Repository Task Force* (ARL), http://www.arl.org/bm~doc/repository-services-report.pdf
- Reference Model for an Open Archival information System (OAIS) http://public.ccsds.org/publications/archive/650x0b1.pdf
- *Trusted Digital Repository: Attributes and Responsibilities* (RLG-OCLC) http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf
- *TRAC: Trusted Repositories Audit & Certification: Criteria and Checklist* (OCLC), http://www.crl.edu/PDF/trac.pdf
- *The University's Role in the Dissemination of Research and Scholarship—A Call to Action* (ARL), http://www.arl.org/bm~doc/disseminating-research-feb09.pdf
- *Vision Task Force Report* (MSU Libraries), http://img.lib.msu.edu/about/VisionTaskForceReport.pdf

Appendix 2
Michigan State University Digital Curation PlanningProject
Baseline Data Questionnaire

Welcome

Welcome to the Michigan State University Digital Preservation Planning Baseline Data Questionnaire—the first step towards participating in a university-wide initiative that will help you preserve and maintain the accessibility of your unit's data.

1. What is the name of your MSU unit or department?

2. What is your title?

Digital Content

3. What types of digital content does your unit produce? Please check all that apply.
    - Word Processed Documents
    - Imaging—Paper Documents
    - Imaging—Photos
    - Imaging—Non-Photos (e.g., maps, drawings)
    - Digital Photos
    - Digital Graphical Images (e.g., maps, drawings)
    - Audio
    - Video
    - Spreadsheets
    - Databases
    - Presentations
    - Web Pages
    - CAD Drawings
    - Data Sets
    - Other

4. Of the digital content types checked in the previous question, which type(s) make up the largest proportion of the total digital content produced at your unit? Please indicate approximate percentage(s) of total proportion of digital content.

5. Approximately how much digital content does your unit maintain?
(multiple choice, one answer)
    - < 1 TB
    - 1-5 TB
    - 5-10 TB
    - > 10 TB

6. How is your digital content stored? Please check all that apply.

Hard drive
Removable Magnetic Media (e.g., floppy discs, Zip discs)
Optical Media (CD/DVD)
Digital Tape
Solid State (e.g., flash drive)
Other

File Formats

7. What file formats are created and/or maintained by your unit? Please check all that apply.

MS Word
Text
PDF
HTML
TIFF
JPEG
WAV
MS Powerpoint
MS Excel
MS Access
MS Publishes
Other (please specify)

8. Of the file formats checked in the previous question, which make up the largest proportion of files produced at your unit? Please indicate approximate percentage(s) of total proportion of files.

Technological Infrastructure

9. What is your unit's current storage capacity?

10. Does your unit plan to expand this capacity in the next year?

Yes
No

11. If so, approximately how much capacity will be added?

12. Does your unit use any content management or other specialized software systems to manage digital files? (e.g., SharePoint, Luna, Extensis Portfolio, etc.)

Yes
No

13. If so, which digital asset management system(s) are used?

14. Does your unit maintain a digital repository?
    Yes
    No

15. If so, what digital repository software is being used? (e.g., DSpace, Fedora, ContentDM)

Confidentiality Issues

16. Is any of your digital content of a confidential or sensitive nature?
    Yes
    No

17. If so, what is the proportion of confidential to non-confidential content?

Contact Information

20. Please provide the following contact information. The MSU Digital Preservation Planning team may contact you shortly to schedule a more in-depth interview.

    Name:
    Email Address:
    Phone Number:

Thank you!

Thank you for participating in this questionnaire. If you have any questions about the MSU Digital Preservation Planning initiative, please contact Lisa Schmidt, digital preservation analyst, at lisa.schmidt@matrix.msu.edu.

Appendix 3
Michigan State University Digital Curation PlanningProject
Digital Curation Planning Survey Results

The baseline data questionnaire administered recently by the MSU Digital Curation Planning Project team in October 2009 yielded 90 responses: 23 from academic departments, 31 from administrative services units, 9 from research centers, and 27 from technology services units.

Represented academic departments covered a wide range of fields, from agricultural economics, nursing, and veterinary medicine to math and science education, physics and astronomy, telecommunications, business, athletics, and the arts. Likewise, administrative units ranged from the Controllers Office, Inventory/Capital Asset Management, the Office of Planning and Budget, the Office of the President/Board of Trustees, and the MSU Libraries to Broadcasting Services, University Relations, and Virtual University Design and Technology (vuDAT), among others. The research centers included the Cyclotron, the Julian Samora Research Institute, and MATRIX. In contrast, all of the technology services responses came from 6 units: Academic Technology Services (ATS), Administrative Information Services (AIS), Agriculture and Natural Resources (ANR) Technology Services, Enterprise Business Systems Projects (EBSP); Enterprise Information Stewardship (EIS); and Health Information Technology.

The types of digital content making up the largest proportion of a given unit's content varied considerably. Digital and scanned photos and images, word processing documents, and research data sets topped several of the academic departments' lists, while administrative units reported large proportions of paper imaging documents, word processing and spreadsheet documents, and databases. Research data, audio/video, word processing documents, and programming code predominated at the research centers. As might be expected, technology services units noted that most of their digital content consisted of code, databases, and web pages.

File formats comprising the largest proportion of a given unit's digital content were similarly varied. Among the academic departments surveyed, PDFs, SPSS and SAS statistical formats, TIFFs, JPEGs, MySQL, and Camtasia video formats were all noted. Various database formats, TIFFs, text, MS Office formats, as well as audio and video formats, predominated at the administrative units. The research centers reported sizeable concentrations of video formats, PHP code, MS Word, and SAS, and the technology services units carry large proportions of text and programming code formats.

In terms of storage, nearly all of the units store digital content on hard drives, and most also use some combination of different types of removable media as well as network storage; one unit even reported storing data on cassette tapes. Seventeen units plan to increase online storage capacity in the near future, most from 1-10 TB, with 6 units planning to add 30 TB or more.

Twenty-three units responded that they have implemented or plan to implement content management system and/or digital repository software. As neither "content management system" nor "digital repository" was defined in the questionnaire, unit representatives responded based on their understanding of the terms. Content management systems noted include Sharepoint, Alfresco, Mura CMS, Drupal, Cascade CMS, Document Viewer, and DotNetNuke, as well as in-house-developed solutions; one unit reported using Trac Project, an issue tracking system for software development projects. The Physical Plant Division uses the Facilities Administration Management Information System (FAMIS) to manage operations, maintenance, and repair projects for the university's physical environment. Digital repository solutions included KORA, the Madison Digital Image Database (MDID), ResourceSpace, Portfolio Server 9, and MSU Extension's custom-built Knowledge Repository system. In some cases, the same software solution was listed as the content management system and the digital repository application.

Tools more properly known as version control software were identified as content management systems and/or digital repository software by some units, including Concurrent Versions System (CVS), Git, Adobe Version Cue, and Subversion. The database application FileMaker was listed as a content management system by one unit. Some units did not use a software tool, but instead noted that they managed content on web and file servers, or that they used "homegrown" solutions such as wikis.

Many of the respondents provided additional comments stating great interest and enthusiasm in the digital curation planning project's goal of establishing naming conventions and other digital curation standards. One administrative unit noted, "This is a timely survey, because our unit is at a point where we HAVE [sic] to choose which data to delete off our servers, as we are accumulating more than we can afford to store. We need university guidelines and related archival resources." Another asked for guidelines on how to handle archive-worthy files at the time of creation, rather than storing everything up front and subjecting the unit to an arduous appraisal process later. Interest in guidance on choosing a digital asset management system was also expressed.

Appendix 4
Michigan State University Digital Curation Planning Project
Unit Interview "Tickler" Questions

Describe the mission of your unit
Describe your digital content
How does the digital content relate to the mission of your unit?
What content must be preserved
    Of ongoing use to unit and/or partners
    "Archival" in the local sense, documenting the activities of the unit
Is any of the content archival in the sense that it documents the history of the university and should be in the custody of the Archives?

File formats
    Describe
    Different preservation and access formats?

How stored
Do they have storage issues?

Discuss CMS and/or DR
What are they using?
What are they doing with it?
What digital content are they storing in it?
Who uses it?
Why did they choose that solution?
How is it working for them?
Does the system provide preservation functionality, such as checksum calculations?
Are preservation masters stored in the CMS/DR?
    If not, where are they stored?
Are they happy with it, or are they looking at implementing another solution?

Describe workflows
    Ingest
    Archival storage/preservation processes
    Access

Metadata
    Information stored with or related to content
    Any particular metadata schema?

File naming conventions
    Consistent?
    Describe

Appendix 5
MSU Digital Curation Planning Interview
Broadcasting Services
March 11, 2020

Broadcasting Services
Harold Beer, Chief Engineer—Radio
Gary Blievernicht, Director of Engineering
Kim Geiger, Information Technologist
Bill Richards, Media Producer/Webmaster, WKAR
Rob South, News Reporter
Jason Thomson, Multimedia Producer, Interactive Video Services (IVS)

Digital Curation Planning Team
Cynthia Ghering
Lisa Schmidt

MSU Broadcasting Services[1] has as its mission "to connect people to powerful ideas that inspire personal growth, exploration of our world and positive change," to educate, entertain, and inform. The unit serves the MSU campus and the Mid-Michigan region with the WKAR radio and television stations, instructional video production, and other campus communications services; the Dubai campus with video conferencing; and worldwide via the Internet. Broadcasting Services has produced radio content for 75 years and television content for 50 years. Media technology has changed drastically and rapidly in the last several years, and now most of the content is delivered and stored digitally.

As a PBS and NPR affiliate, Broadcasting Services is allowed to store content received from those national broadcasters for the duration of a "rights" period, which is typically up to three years. This content is not archived, however, and is of little interest in terms of digital curation planning for MSU. The rest of this report will cover content created by Broadcasting Services.

In the early years of television broadcasting, the station simply taped over locally produced video week after week. Some select TV shows from the 1960s on still exist, and these have been

migrated to DVDs. Minimal metadata for these shows (such as date and title), originally stored on a Rolodex, was transferred to a database a few years ago. Broadcasting Services holds approximately 300 DVDs of this older digitized video, some of which is not of broadcast quality.

Broadcasting Services began producing born-digital video content in the late 1990s. Current production technology involves shooting straight to video recording onto Sony Betacam SX in the compressed MPEG2 format, and Apple Final Cut Pro is used for editing. The content may be exported in a variety of formats, including QuickTime. To date, Broadcasting Services has primarily stored its streaming online video content in Real Media formats, although online video distribution is expected to convert to MPEG-4 or Flash in the near future. Again, minimal metadata is provided for this content, as news producers tend to focus on their next story rather than on the curation of the show that they just finished.

As a PBS member, the TV station will begin to participate in PBSCove, a comprehensive video environment that provides the illusion of a seamless interface for the end user of any member's website. Broadcasting Services will enter locally-produced content into PBSCove. PBSCove requires MPEG-4 compressed video at an approximate rate of 500kbs. Content is not easily searchable and is intended primarily for access, not as an archive. A process for digital storage of higher resolution copies is still in development.

The radio station at MSU started in 1922, and most of the earliest content produced was never recorded. Some recordings were made to old media formats over the years, such as transcription disks and analog tape. Beginning in the early 1990s, digital recordings of some concerts, news features, and local talk programming were created in the DAT format. Although most pre-1999 digital audio does not seem to have been retained, some producers may have individually collected some of the DAT recordings. None of these earlier digital recordings were formally

cataloged, although some producers added minimal metadata for the tapes to the DAT labels. The digital content on the DATs has not been migrated to a more current format. Broadcasting Services has two working DAT machines that can read the tapes.

Broadcasting Services' current digital audio content exists in three formats for three purposes:

- Audio played on the air, which is centrally stored in Broadcasting Services' own automated "AudioVault" content management system. The limited metadata embedded in these files is in the MPEG-1 Audio Layer II or WAV Cart Chunk audio broadcasting formats.
- Audio available on the WKAR website, which is stored on a server. The original broadcast files are converted to an access format—formerly Real Media, now MP3. More metadata is provided, as well as transcripts for some scripted stories. Locally produced stories are also being made available in a public interactive management system, an aggregator from which they can be served up to the websites of other NPR stations. Weekly webinars provided by the aggregator train producers on how to create metadata.
- Preservation masters of complete audio features are kept only at a producer's discretion. WAV is the preferred format for preservation masters, as the files can be edited before conversion to access copies.

Broadcasting Services also holds about 6 GBs of 1000-2000 born-digital photographs taken during productions, averaging 40 still images per hour of programming. These digital photos, most of which are in JPEG format, are stored on a file server. They are uncataloged, relying for identification and management on a directory structure that includes names and dates of the subjects. Most of the digital photos were taken by the promotions staff over the last seven years and are used for public relations purposes. Reporters and producers also take photos that are not always added to the file directory on the server. Broadcasting Services also still possesses

photographs on film and paper used for promotional purposes previous to 2003. To date, there are no plans to digitize them.

Since September 2009, Broadcasting Services has retained PDFs of Stay Tuned, its monthly radio and television program guide that is published online. The table of contents for each issue functions as an index to the programs aired during that month.

In terms of backup, Broadcasting Services mirrors the AudioVault on multiple local servers in the Communication Arts building as well as on an offsite mirror at their transmitter building in Okemos, Michigan. The unit is also working on mirroring the video content off-site. There are no current plans to back up or migrate the content stored on DVDs. Still images, website content, and Real Media audio and video content are backed up onto DLT tape and stored off-site by Academic Information Services (AIS).

Broadcasting Services also hosts WMSU, a public webcasting station that streams live and stored video and audio of public events and other programs. This content is stored on its own server and streamed in the Real Media format. Virtual University Design and Technology (vuDAT) backs up this server. Video of East Lansing City Council meetings (copyright held by City of East Lansing) are also provided to WMSU (on DVD), and served online at WMSU.org in Real Media format.

Interactive Video Services (IVS), a department within Broadcasting Services, creates digital interactive course content for MSU faculty. Lectures are recorded in the MP4 format and audio versions of the lectures are made available in the MP3 format for podcasts. Both video and audio files are streamed online and are also available for download from Angel. The files are kept in long-term storage on an external hard drive, along with Word files containing metadata for the

multimedia content. Although IVS's multimedia files are not formally backed up, redundant copies of the files run on the streaming server.

As some of the material produced by Broadcasting Services documents the history of Michigan State University, that material rightly belongs at UAHC after it has passed through its time of active use. In the past, Broadcasting Services has deposited film and tapes of programs with UAHC. Now UAHC needs to work with the unit to develop a system of appraisal and transfer of its digital content, similar to processes being developed with University Relations.

---

[1] Broadcasting Services, Michigan State University, http://www.wkar.org/. Retrieved 14 May 2010.

Appendix 6
MSU Digital Curation Planning Interview
The Confucius Institute at Michigan State University (CIMSU)
January 22, 2010

The Confucius Institute at Michigan State University (CIMSU)
Ken Dirkin, Information Technologist
Roman Stotland, Information Technology Professional

Digital Curation Planning Team
Cynthia Ghering
Shawn Nicholson
Lisa Schmidt

The Confucius Institute at Michigan State University (CIMSU)[1] is part of a network of more than 60 such institutes at universities in the United States and more than 300 worldwide. Developed in collaboration with the Chinese Language Council International (Hanban) in Beijing, a branch of the Chinese Ministry of Education, the Confucius Institute seeks to strengthen educational cooperation with China, develop Chinese language education, and increase intercultural understanding with the peoples of China. Established in 2006, CIMSU is the oldest Confucius Institute in the state of Michigan.

Considered one of the best CIs in the US, CIMSU is led by executive director Dr. Yong Zhao, University Distinguished Professor in the College of Education and advocate of merging Eastern and Western education philosophies. CIMSU focuses on creating online materials for teaching Chinese to K-12 students, especially those living in rural locations. The CIMSU courses use interactive e-workbooks and include weekly synchronous audio chats via Adobe Connect. To date, approximately 3,500 students in Michigan have taken the courses, as well as students in five other states and Mexico.

CIMSU also supports the training and certification of Chinese language teachers and creates curriculum materials. Among the media provided by the CI is Zon, a Chinese language learning

game, for which CIMSU created much of the original artwork and content. CIMSU also uses the digital content in its instantiation of Second Life.

Dr. Zhao is also the director of the College of Education's Center for Teaching and Technology, which supports the College's goal of helping to train students in the effective use of technology in the classroom. Thus, many of the methods and techniques used at CIMSU are also applied at the Center for Teaching and Technology. Eight full-time employees, a dozen graduate students, visiting scholars, and several consultants work at CIMSU.

Digital content created by the CIMSU includes digital video files, audio, and programming files. Through 2006, most of the video was shot in the MiniDV format and converted for online use; since then, it has been shot straight to HDV and is accessed using QuickTime Player.

Audio content, including podcasts, songs, and sound effects for games and textbooks, are recorded as WAV files and compressed into MP3s for access purposes. CIMSU stores both versions. Most digital images are created in Photoshop and Illustrator and stored in high-resolution JPG, PNG, and EPS formats rather than in the raw TIFF format. Programming files include Adobe Authorware, Flash, PHP, and MySQL, all of which are used to create the Zon game. Images created in the vector-based EPS and Flash formats are lossless and can be scaled infinitely. Final course packages are in HTML.

File names are generally based on the project name, type of digital object (such as video, audio, image), the date, and a version number. Podcasts are named for the date of production or distribution.

The CIMSU has been using Adobe Version Cue 3 and shared drives as its content management systems for the curriculum packages it develops. Part of the Adobe Creative Suite, Version Cue enables users to track and manipulate file versions and metadata. Migrating from one version of Version Cue to the next has been difficult due to connectivity differences between CS3, which uses the Adobe Bridge interface, and CS4, which uses Adobe Drive. Due to this significant change, CI is not currently comfortable migrating content to the newer version.

SubVersion (SVN) is used to manage the individual programming pieces of digital content used to create the curriculum packages. The Trac system manages the SVN repositories, performing such functions as bug reporting and task assignment. File types stored in SVN include text, video, sound, and image files. SVN does not read Flash files, although it can store them. The ideal content management system would read all file types, including Flash.

Since October 2009, CIMSU has been testing the open-source Alfresco content management system[2] as a collaborative document management solution to share digital content. (To date, shared drives have been used for this purpose. The College of Education also has a SharePoint installation.) Alfresco offers many promising features for the management of digital content as well as business documents. Its Java-based API ties into WordPress, which is used as a front-end to Zon, as well as Moodle, which is used for CIMSU's K-12 courses. Version history, including the user who created the new version, is recorded. Documents may be scanned, e-mailed, encrypted, and stored in the system. Workflows may be created and tasks assigned, with various levels of permissions granted, including public access. A personal dashboard tracks all of the documents, content, and tasks of interest to an individual. CIMSU has already started to store business and MS Office-related documents in Alfresco shares. To date, the center is not using it for large-scale multimedia storage due to space and backup concerns. In addition to using

Alfresco as a document/content management system, CIMSU is considering creating a help desk environment for student IT support.

CIMSU has 32 TB of data storage capacity. The center stores its content on RAID servers, including mirrors distributed across other servers, and performs tape backups of critical data that are stored in a safe. The College of Education backs up and stores its core business infrastructure data on external RAIDs that can be transferred off-site to ACD.net, a commercial vendor in Lansing that offers environmentally controlled storage.

To date, CIMSU has focused on making its content available for re-use in current projects and not so much on ensuring its long-term preservation into the future. The center is seriously looking for long-term data storage and digital asset management solutions.

---

[1] Confucius Institute at Michigan State University, http://www.experiencechinese.com/. Retrieved 14 May 2010.
[2] Alfresco Software, http://www.alfresco.com/.Retrieved 14 May 2010.

Appendix 7
MSU Digital Curation Planning Interview
Department of Art & Art History
January 21, 2010

Department of Art & Art History
Tom Berding—Department Chair
Sue (Susan) Morris—Curator, Visual Resources Library
Alex Nichols—Academic Technology Coordinator

Digital Curation Planning Team
Cynthia Ghering
Shawn Nicholson
Lisa Schmidt

The MSU Department of Art & Art History[1] provides the Visual Resources Library (VRL) as a service to faculty and students. Primarily intended as a teaching collection, the VRL includes an analog collection of 180,000 images and 30,000 digital images. In 2005, Art & Art History hired an academic technology coordinator to create a repository of digitized images for the VRL, as well as to support technology in other areas of the Department. The VRL now includes 30,000 digital images, sources of which include student work, commercial vendors, books, faculty gifts, faculty research, gifts from visiting artists, and digital surrogates of images from the analog collection.

Faculty and students in art history, studio art, design, and other areas of study make use of the VRL. On an informal and limited basis, the Department of Art & Art History also uses images in the VRL for public outreach and stores images in the VRL that document historic departmental events, such as the 75th anniversary celebrations in 2006. (The VRL is investigating the possibility of formalizing such services in the future.)

Art & Art History uses the Madison Digital Image Database (MDID), an open-source, web-based content management system developed at James Madison University[2] to manage, share, and

organize digital images in the VRL. Designed as a teaching tool to encourage active learning, MDID allows faculty, students, and other users to curate collections and slideshows online as well as offline using an image viewer utility. MDID allows users to pull images into PowerPoint and Angel, generate flashcards, browse collections, create galleries of favorite images, and tag images. Due to copyright issues, images in the collection may not be used for publication in journals or books. It is recommended by the VRL that those seeking to publish images consult with the MSU Libraries regarding permissions. To that end, a copyright disclaimer with a link to the Libraries' website displays when a user logs in to MDID.

Master files of the images are created in the TIFF format and stored on a file server, with copies made to offline external hard drives and to CDs and DVDs in some cases. All copies of TIFFs are stored in the Kresge Art Center. Access copies are generated from the TIFFs in the JPEG format and stored in the MDID repository, and the MDID server is backed up to tape stored offsite. Housed at MSU Academic Technology Services (ATS), the MDID server is managed by ATS staff and Art & Art History's academic technology coordinator. Anyone with MSU login authorization may access the online VRL. Note that images created in other formats are converted to TIFF and JPEG.

Although MDID does not provide checksum validation, Art & Art History has other means to verify the integrity of the files in the repository. Corruption of a file is indicated if the Adobe Bridge image viewer cannot read a thumbnail. Also, a Photoshop script that creates a contact sheet of thumbnails will stop running if it encounters a corrupted file. Art & Art History has found few corrupt files in the MDID repository, with more problems arising from licensed images than those created by the department.

More than 90 institutions use MDID to manage digital images, and NASA also uses the software for its image exchange. These institutions can perform aggregate cross-institution searches and share collections using Dublin Core metadata fields to match up requests.

The current version of MDID is Java-based on a MySQL database. A new Python-based version with the capacity to support video and multimedia will be released in summer 2010. The new version of MDID will broaden searches outside of the MDID institutions to include Flickr, ARTstor, and other large image collections.

With MDID as the public facing interface to the VRL, Art & Art History uses the Image Resource Information System (IRIS) data standard for cataloging and metadata management of art images. Created by a consortium of image librarians and curators, IRIS is a FileMaker Pro-based relational database and cataloging tool that uses Visual Resources Association (VRA)[3] Core and Cataloging Cultural Objects (CCO) standards.[4] IRIS supports the elaborate metadata needs of the department and exports easily into MDID as a CSV file. Although IRIS institutions do not share authority tables, each implementation may be customized to use controlled vocabularies based on standard thesauri. The VRL primarily uses the Getty Art & Architecture Thesaurus (AAT), the Getty Union Catalog of Artist Names (ULAN), the Getty Thesaurus of Geographic Names, Grove Art Online, and Library of Congress name and subject authorities. Institutions may also create local subject headings.

Art & Art History uses the accession number of an image for its file name. With images from graduate student work, the file name consists of the student's last name, the year that he/she graduated, and a sequential number within the body of work. For example, file names for the first two images in the work of a student named Nichols would be "Nichols2001_1" and "Nichols2001_2."

Art & Art History increases storage space for digital files as necessary. The addition of videos to the online collection may create storage problems in the future, however.

The VRL has not yet produced formal documentation of workflow. Instructors request images, which are then digitized in the department's Digital Imaging Lab and cataloged using IRIS. Faculty may also upload their own images with custom metadata and make the collections publicly available as they wish. Student artwork included with a thesis—usually 15-20 slides—is digitized or converted to TIFFs if submitted digitally. (Art & Art History is looking into an acceptable archival format for moving images.) The thesis itself is submitted on a CD or DVD, as an MS Word or PDF file, accompanied by images, videos, or other files documenting the associated artwork. Art & Art History uploads the student work into MDID with basic metadata, including the student's name, title of the thesis or work, media used, dimensions of the work, year of graduation, year artwork was created, and the student's area of study.

---

[1] Department of Art & Art History, Michigan State University, http://www.art.msu.edu/. Retrieved 14 May 2010.
[2] Madison Digital Image Database (MDID) system, http://mdid.org/overview.htm. Retrieved 14 May 2010.
[3] Visual Resources Association (VRA) Core data standard, http://www.vraweb.org/projects/vracore4/. Retrieved 14 May 2010.
[4] Cataloging Cultural Objects (CCO) http://www.vraweb.org/ccoweb/cco/about.html. Retrieved 24 May 2010.

Appendix 8
MSU Digital Curation Planning Interview
Department of Theatre
February 26, 2010

Department of Theatre
Brian Adams, Technical Director
Alison Dobbins, Integrated Media and Performance Design Professor
Kirk Domer, Acting Chair and Head of Design

Digital Curation Planning Team
Cynthia Ghering
Lisa Schmidt

Michigan State University's Department of Theatre's mission: "To provoke and sustain interest in the theatre as a source of truth and insight into the human condition."[1] Digital content that documents the context of productions supports the department's educational and teaching goals. This content includes show photos, press releases, promotional materials, stage and set designs, design resources such as animations and images, and reviews of productions in local newspapers. (Due to copyright issues with playwrights and their agents, live productions may not be shown publicly or distributed online, although videos of performances may be used in controlled teaching environments.)

The Department of Theatre vetted several repository software solutions for storing and managing its digital content before choosing the open-source ResourceSpace application.[2] Brian Adams, technical director of the Department of Theatre, felt that PHP-based ResourceSpace was well written and liked that it was easy to modify, had a good user interface, and was freely available. Originally developed by Montala Limited for the British non-profit Oxfam, ResourceSpace was released under a BSD-style license in 2006. Theatre hosts it own instantiation of ResourceSpace, the DOT::Media repository, on an Apache web server. In the spirit of the open source software movement, Adams offers code modifications back to the ResourceSpace community. One such enhancement called "NameGame" allows students to tag photos and other digital objects. Adams

also runs a ResourceSpace repository for the College of Arts and Letters, as well as a custom build of ResourceSpace for managing and storing archaeological dig data and documents such as photographs and notebooks.

DOT::Media went live in 2008, and the repository has hosted digital content related to shows produced every season since then. The department holds additional digital content dating back to 2005, but it has not all been ingested into DOT::Media at this time. In honor of the 50th anniversary of the department's free outdoor play series, Summer Circle Theatre, retired alumni and other volunteers are scanning old photographs to create a digital "memory book." Adams plans to ingest this content into DOT::Media.

ResourceSpace allows for the management of a variety of file formats. Theatre ingests and stores digital photos as high-resolution JPEG files; the original RAW formats are converted directly to JPEG rather than TIFF. (Select original RAW images are stored on the workstation of the integrated media professor.) Proprietary AutoDesk CAD files of stage set drawings are stored in both DWG and DWF formats, and PNG thumbnails are also available as previews. PDFs of press releases, promotional materials, and reviews are also ingested, as well as DOC and HTML files. ResourceSpace can optionally convert uploaded DOC files, such as the Production Firearms Policy, into PDFs to render them uneditable. Audio files, created by students or licensed (with restricted access), are present in WAV and MP3 formats. Student-created video files not subject to licensing restrictions are available in AVI, WMV, and MPEG formats. Flash video and audio recordings may also be ingested into the repository in the future.

To date, the workflow for adding content to the DOT::Media repository is very controlled. Typically, department faculty, staff, and students provide Adams with digital objects requiring

ingest. Staff members with appropriate privileges upload the material, which is kept in a holding status until Adams previews, approves, and releases it to the main collection.

The metadata used in DOT::Media is based on Dublin Core elements. Adams and other staff manually enter much of the metadata using some controlled vocabulary in dropdown menus. On ingest, metadata stored in the digital photos through the exchangeable image file format (Exif) is extracted and automatically added to the repository record for that digital object. All digital cameras store Exif-based metadata with images, which includes technical specifications for the camera. In the case of smart cameras, the Exif metadata may include geo-location data indicating where the photo was shot. With this geo-location data, a map could be generated to display where photos related to a production were taken. All metadata is stored in a MySQL database, which can be exported as XML.

Regarding digital preservation features, ResourceSpace developers are working on adding a file integrity test to the code. Adams hopes that this will be accompanied by the capability to store parity files with the original files for easier fixes if corruption occurs.

Primarily intended as a resource for Theatre students, ResourceSpace allows them to download performance photos to add to their portfolios. MSU students, faculty, and staff use their Net ID to access DOT::Media. From there, several options for viewing and downloading the content are available. Individual records and files may be viewed and downloaded. Groups of images may be zipped together with an index and then e-mailed to the user. Private, public, and themed collections may be created, along with a contact sheet of each collection as a PDF of thumbnails, names, and identifiers of individual objects.

No file naming conventions have been established for DOT::Media. A file retains its original name on ingest. ResourceSpace automatically assigns a unique identifier for use within the system.

The DOT::Media repository resides on a 112 GB server, and only 39% of this space is currently in use. The server is backed up internally using RAID.

In addition to the ResourceSpace repository, the Department of Theatre is experimenting with using the StoreMedia file manager software for streaming video and downloadable Flash files in the classroom. StoreMedia was developed by MSU's Virtual University Design and Technology (vuDAT) unit.

---

[1] Department of Theatre, Michigan State University, http://www.theatre.msu.edu/. Retrieved 14 May 2010.
[2] ResourceSpace digital asset management system, http://www.resourcespace.org/. Retrieved 14 May 2010.

Appendix 9
MSU Digital Curation Planning Interview
MATRIX: The Center for Humane Arts, Letters and Social Sciences Online
February 11, 2010

MATRIX
Matt Geimer, Chief Technical Officer
Alicia Sheill, Project Manager/Office Manager
Dean Rehberger, Director

Digital Curation Planning Team
Cynthia Ghering
Shawn Nicholson
Lisa Schmidt

MATRIX: The Center for Humane Arts, Letters and Social Sciences Online is a digital humanities research center at Michigan State University dedicated to the application of new technologies for teaching, research, and outreach.[1] As one of the premier humanities computing centers in the United States, MATRIX creates and maintains online resources, provides training in computing and new teaching technologies, and creates forums for the exchange of ideas and expertise in the field.

MATRIX houses major digital library repositories, including the African Online Digital Library (AODL), Detroit Public Television's American Black Journal video archives, Historical Voices, and The Quilt Index. These repositories run on KORA, an open-source, database-driven, online digital repository application developed by MATRIX.[2] MATRIX also hosts H-Net, an international scholarly networking community. Future plans include research into ways of using computational power to look at and analyze large stores of data.

Designed primarily for creating online collections for small cultural heritage institutions, KORA is a highly flexible repository application that may accommodate a variety of digital object types, including text, images, audio, and video. Although nearly all file formats may be ingested into

KORA, certain display and access formats are encouraged—including JPEG, MP3, Flash video, and PDF. Other formats found in KORA include DOC, GIF, HTML, Real Media (audio/video), WAV, and QuickTime. An associator feature allows record types to be linked together, such as individual pages in a larger document and various pieces of complex multimedia objects.

KORA stores data and digital objects in MySQL tables. Metadata fields are based on Dublin Core, but can be modified to meet client needs. An OAI-PMH mapping function may be activated within a KORA project to enable the harvesting of Dublin Core metadata. KORA also has the means to establish fixity and run integrity checks on digital objects for preservation purposes using the SHA-256 message digest algorithm.

MATRIX project managers work with clients to create an online collection that meets their specifications and unique needs. With guidance from MATRIX, the client defines their own metadata, controlled vocabularies, and file naming conventions; determines what types of digital objects will be included in the collection; and decides on the end-user format. (KORA assigns its own unique identifier used behind the scenes.) During the course of project development, these aspects may be tweaked as needed. The client also decides whether they wish to store preservation copies of the digital objects within KORA. To date, they have only wanted to ingest access copies, and the fixity mechanism has yet to be implemented for a KORA project.

Note that preservation copies of much of the audio and video content available through KORA-based collections is stored offline on the MATRIX premises. As audio and video technology changes rapidly, these preservation copies are required to make new access copies as needed.

Clients have two basic options for ingesting data: manual and XML import. Manual ingest involves filling out an ingest page form with drop-down menus and uploading digital objects.

With the XML import option, the client provides the data and files are matched to the chosen schema. After import, the client may check the data for accuracy.

MATRIX can work with the client to develop a public website for a project, and KORA plug-ins for WordPress that will automate the creation of front ends for digital collections are currently in development. Alternatively, clients have the option of creating their own web-based front ends using PHP. Most KORA-based collections are publicly available for access through websites, although permission levels within KORA limit content editing.

As clients usually own the digital objects hosted by KORA, in most cases there are no copyright issues with the collections. Some projects, such as the Detroit Public Television and the "Studs Terkel: Conversations with America" projects, have required the signing of permission statements. Ownership statements are provided on the various websites for collections.

MATRIX hosts three KORA repositories as well as a "sandbox" implementation for testing code changes and new features. KORA is available for download on SourceForge. In 2009, there were 800 downloads and several reported usages of the software outside of MATRIX.

KORA self-stores files on servers that MATRIX backs up regularly to LTO tape, with incremental backups occurring daily and full backups on a weekly basis. These tapes are stored at the Cyclotron building on the MSU campus. Long-term backups to tape are performed every four months, and the tapes are currently stored on the MATRIX premises with tentative plans to transfer them off-site to Conrad Hall in the future. MATRIX also has a reciprocal storage arrangement with the Inter-University Consortium of Political and Social Research (ICPSR) at the University of Michigan, which involves the daily copying of 3 TB of MATRIX data to ICPSR

servers and vice versa. In the near future, MATRIX will increase its onsite redundant storage capacity to 10 TB for better availability and uptime.

---

[1] MATRIX: Center for Humane Arts, Letters, & Social Sciences Online, http://www.matrix.msu.edu. Retrieved 14 May 2010.
[2] KORA digital repository software application, http://www2.matrix.msu.edu/research/technology/kora/. Retrieved 12 May 2010.

Appendix 10
MSU Digital Curation Planning Interview
MSU Extension/Agriculture and Natural Resources (ANR) Technology Services
January 14, 2010

MSU Extension/Agriculture and Natural Resources (ANR) Technology Services
Erica Ciupak, CIO, ANR Technology Services
Cindy Straus, Extension Specialist

Digital Curation Planning Team
Cynthia Ghering
Lisa Schmidt

Michigan State University Extension (MSUE)

Michigan State University Extension (MSUE) brings knowledge-based educational programs to Michigan's 83 counties with programming focused on agriculture and natural resources; children, youth and families; and community and economic development. By facilitating access to the information created for these programs, MSUE's digital repository directly relates to the unit's land-grant outreach mission to the people of Michigan.

MSU Extension uses a digital repository system custom developed by Toronto-based Intrafinity[1] with Teaching and Learning Environment (TLE) funds. Currently known as the Knowledge Repository,[2] MSUE plans to change the name of the repository to the MSU Extension Electronic Library.

The Knowledge Repository contains MSUE print bulletins that have been digitized and reformatted by the MSU Libraries. For a fee, the library scans the bulletins and creates PDF files from the TIFF originals. The TIFFs are stored in a dark archive at the library and the PDFs are provided to MSUE on CDs. From there, they are loaded on a server and ingested into the Knowledge Repository. To date, more than 3,000 bulletins have been digitized; approximately half of those have already been loaded into the repository. Bulletins are available from 1917

through the present, with more recent editions published in digital formats. A bulletin file name consists of the e-bulletin number and year, with "SP" added for Spanish language bulletins.

In addition to the bulletins, the Knowledge Repository contains content created by Area of Expertise (AoE) Teams. For example, fruit information fact sheets are created and posted on the Fruit AoE Team page without having been run through the MSU Bulletin office system. Currently, there are 33 teams covering many areas. After a pending reorganization, four institutes will carry out work in areas determined to be of high priority to the state. As the fact sheets created by the AoE teams were not originally printed by the MSUE print shop, they do not fall under the bulletin numbering system.

The Knowledge Repository also includes MSUE-approved links to standalone websites containing known Extension-generated material, Extension-related material, and material referenced or hosted by another agricultural extension program. This linked material consists of such items as county newsletters and peer reviewed articles, most of which are in PDF, HTML, PPT, and JPEG formats. One such link will be to a set of 3,000 JPEG files soon to be gifted to MSUE, horticultural images created by retired MSU Extension specialist Randy Heatley. Links currently require manual checking to ensure viability.

Several online databases have been added to the Knowledge Repository, including historical databases such as "Home Maintenance and Repair." Originally available in the dBMAN program, the databases have been converted from that DOS format to HTML.

Although no audio or video files have been ingested into the Knowledge Repository thus far, MSUE plans to begin ingesting podcasts created by the counties as WAV files. Links are provided to dozens of Breeze presentations, but the presentations themselves are not yet stored in

the repository. MSUE currently works with Virtual University Design and Technology (vuDAT) to provide access to its Breeze presentations; however, current vuDAT policy is to delete presentations from its server after one year.

The MSUE Knowledge Repository uses slightly modified Dublin Core metadata, with some of the fields renamed for easier input. Controlled vocabulary based on the US Department of Agriculture's National Agricultural Library Thesaurus (NALT) is used for metadata, with the option to use local, uncontrolled vocabulary. Students, part-time librarians, and the Extension specialist add metadata to the digital objects on ingest and to posted links. Outside staff could add content but not metadata, but to date a standard process is not in place and training has not been made available. Exports of records to XML do not include metadata, an unfortunate limitation of the Knowledge Repository

Part of the MSUE portal, the Knowledge Repository runs on three servers: one each for indexing data, for storage, and for the public interface of the repository. The Knowledge Repository was originally intended to serve as both preservation and access tool, although its functionality is currently focused on access. For example, PDFs rather than TIFFs are stored in the repository. Also, rather than ingesting all of the content, links are provided to websites containing it as described above. The Knowledge Repository database contains 3 GB of data.

Users of the Knowledge Repository include MSU staff and faculty as well as the general public. Data on user statistics is not currently available. MSUE tried to use WebTrends, but has not been happy with the results. The unit is considering using Google Analytics instead.

Knowledge Repository policies are scant. A collecting policy states that the Extension specialist must approve new information to be added. Disclaimers are included with historical documents

and bulletins—for example, in the case of outdated pesticide recommendations. In some special cases, researchers are directed to contact the repository manager for access to some of the historical material.

Agriculture and Natural Resources (ANR) Technology Services

Agriculture and Natural Resources (ANR) Technology Services provides technology services support to MSUE, the College of Agriculture and Natural Resources (CANR), and the Michigan Agricultural Experiment Station (MAES).

Expression Engine, which has replaced DotNetNuke as ANR's content management system, manages web-based, public facing information such as press releases, brochures, and some faculty research. File formats include HTML and PDF, as well as audio and video files stored on a media server. Other content created is typical of most academic units, and includes faculty research, grant proposals, and state and Federal grant and contract reports. ANR will soon launch SharePoint as its internal records management system to administer business process workflows and reporting. Governance and control issues are being addressed pre-launch, including the creation of a taxonomy.

In terms of technical infrastructure, ANR has implemented a storage area network (SAN) server system that runs 100% of the unit's administrative functions and centralizes backup and disaster recovery for ANR. The SAN has three parts: an e-mail system, flat file storage, and a database system. ANR is trying to eliminate "under the desk" servers, and to that end 45% of the full-time employees now make use of the SAN.

---

[1] Intrafinity, http://www.intrafinity.com/. Retrieved 14 May 2010.

[2] MSU Extension Knowledge Repository, http://www.msue.msu.edu/portal/.

Appendix 11
MSU Digital Curation Planning Interview
National Superconducting Cyclotron Laboratory (NSCL)
February 19, 2010

National Superconducting Cyclotron Laboratory (NSCL)
Jay Kusler, Department Head

Digital Curation Planning Team
Lisa Schmidt

The National Superconducting Cyclotron Laboratory (NSCL) at Michigan State University[1] is a world-leading laboratory for rare isotope research and nuclear science education. With support from the U.S. National Science Foundation (NSF), the laboratory operates as a national user facility that serves more than 700 researchers from 100 institutions in 35 countries.

Research data is the primary digital content created at the NSCL, which has committed to providing thousands of experimental hours annually to the NSF. Organized months in advance, experiments at the NSCL run from one day to two weeks. Each experiment consists of millions of events. Through the NSCL Data Acquisition System, NSCL's in-house developed open-source general purpose data acquisition software, thousands of files—up to 300 GB of data—can be generated from a single experiment.[2] SpecTcL, an open-source, C++-based data analysis package for nuclear physics data developed at NSCL, is used to read, analyze, and interpret the data and to generate histograms of the events.[3] Both the NSCL Data Acquisition System and SpecTcL packages are published on SourceForge, and labs at Duke University, Kansas State University, and other institutions also use the software.

At the close of an experiment, the data is written to tape and then read back on a different tape drive to a different file system for validation purposes. Once the researcher has ensured the validity of the tape, the original data is deleted from the original file system's staging area to free

up space for new experiments. One copy of the experimental data then resides on the new file system and at least one copy is on tape. Researchers may make as many taped copies as they wish. Typically, a researcher will keep at least one tape and leave a taped copy with NSCL, but there is no requirement to do the latter.

NSCL stores approximately 30 TB of research data on its Unix file servers. For convenience, the data on the servers is available for analysis. NSCL does not back up the servers, as the copy on tape is considered the true end copy. If data on the servers is lost, NSCL turns to either its stored tape copy or the tape retained by the original researcher. NSCL keeps its copies of the data tapes in a "research store" at the facility. Most of these tapes are in the LTO-2 and LTO-3 formats. At this time, there is no media migration strategy for the tapes and no off-site storage for data tapes.

There are currently no plans to migrate data in older format versions of the NSCL Data Acquisition System software. Although newer versions of the software are not necessarily backwards compatible, NSCL maintains earlier versions for reading and compiling older data. As the software is open source, specifications for non-current versions should always be available.

Calibration data saved in a separate file describes the research data. This important corresponding metadata is also in the NSCL Data Acquisition System "Event Data" file format. Without this calibration, the research data would be unusable.

Before an experiment gets "beam time," a five-digit number is assigned to it by the NSCL Program Advisory Committee (PAC). The first two digits of this number indicate the year of the PAC in which the experiment was approved. This number becomes the name of the file directory for the experiment. For each experiment, the researcher decides how he or she wishes to name the files within the directory. During the course of an experiment, a new file will start when the

working file reaches the size limit of 2 GB. Each new file is numbered in consecutive order after the naming scheme assigned by the researcher.

In terms of other digital content, NSCL's design department creates CAD drawings of buildings and cyclotrons, primarily using the proprietary SolidWorks CAD software program; these CAD drawings occupy 400-500 GBs of storage space. In the near future, NSCL is planning to add the SolidWorks Product Data Management (PDM) software module that provides CAD-specific document management and version control.

NSCL also creates business documentation that supports its infrastructure and research. This includes purchasing and procurement documentation, as NSCL does not go through the standard MSU purchasing channels. Other content includes Microsoft Word documents and e-mail correspondence. Scans are made of documents that are not born digital. To date, NSCL has used the Alfresco content management system for the project documentation required by ISO certification programs, the NSF, and the Department of Energy. However, the center plans to move to SharePoint as its content management and document control system in the near future. Although NSCL likes Alfresco—which scales well and easily handles simple workflows—the center lacks the Java programming expertise required to customize the software to its overall needs. As NSCL already uses SharePoint for collaboration among researchers nationwide on Facility for Rare Isotope Beams (FRIB) project design, the center sees value in simplifying on one system for a centralized portal.

Backup procedures for the business data residing on NSCL's servers include weekly full backups to tape. After three months, only data from the first week of the month is retained. Once a year has passed, only data from the month of that year is kept.

[1] National Superconducting Cyclotron Laboratory (NSCL), http://www.nscl.msu.edu/. Retrieved 14 May 2010.
[2] NSCL Data Acquisition System, http://sourceforge.net/projects/nscldaq/. Retrieved 12 May 2010.
[3] NSCL SpecTcL Histogramming System, http://sourceforge.net/projects/nsclspectcl/. Retrieved 12 May 2010.

Appendix 12
MSU Digital Curation Planning Interview
Physical Plant Division
February 12, 2010

Physical Plant Division
Angela Carey, Physical Plant Unit Information Systems Manager
Lorena Griffin, Physical Plant Records Planner/Inspector/Analyst II

Digital Curation Planning Team
Cynthia Ghering
Lisa Schmidt

The Physical Plant Division of Michigan State University[1] "provides and maintains the University's physical environment"—that is, all buildings and land entities on and off campus. The Physical Plant's primary operational purposes include keeping buildings clean and sanitary; operating, securing, maintaining, and repairing building and utility distribution systems; and providing electricity, steam, and potable water. Key services include a 24/7 response service line; engineering and structural architecture design; construction management, including construction estimates and management of contractors; landscape services and snow removal; custodial services, including set up/take down crews for events; transportation services and motor pool, including auto repair, maintenance, and fuel, and bike rental and repairs; recycling; skilled trades specialties, including painting, elevators, locksmithing, masonry, roofing, plumbing, pipefitting, metal working, HVAC, refrigeration, and steam tunnel crews; power plant, including water; and telecommunications and phones. The Physical Plant has more than 1,200 full-time and part-time employees, including students.

The Physical Plant uses the Oracle-based FAMIS enterprise facility management software suite (now owned by Accruent) to manage MSU's facilities operations.[2] The MSU FAMIS implementation includes integrated modules for maintenance management; inventory control used to manage large warehouse inventories, with a complementary wireless bar coding

functionality module; an interface to University Purchasing; key control; tool control; facility assessments, which maintains information about buildings (names and numbers), equipment, and other capital assets; space management; an AutoCAD interface; capital projects budgeting and scheduling system; and mobile software for automating field workers.

Three other campus units share the MSU FAMIS system with the Physical Plant: Facilities Planning and Space Management (to the room level), Campus Planning and Administration, and Residential and Hospitality Services (including Campus Interior Design). A Facility Interface Tool (FIT) may be used to compile reports from information stored in the FAMIS database.

For document management, the Physical Plant recently started using InnoCielo Meridian Enterprise software,[3] otherwise known as "Meridian." Scanned paper records as well as born digital documentation related to construction and other operational projects are stored in the Meridian system; once scanned, the physical documents are transferred to the University Archives. The Physical Plant is in the process of migrating 200,000-300,000 documents into Meridian from its previous FileNet document management system. FileNet had good workflow functionality but would have been costly to upgrade. Meridian provides more customization options and integrates with FAMIS. Consideration is also being given to using Meridian to manage the staff's current working files.

The Physical Plant uses Skire project management software[4] to manage business processes digitally, from project order through final payment; hard copies of some capital project documentation are never created. Consultants and contractors with a contract for a capital project are required to use Skire. Plans to integrate Skire with the Meridian document management system are under way.

Other software used by the Physical Plant includes: Munsys spatial data management software, which maps underground utilities and interfaces with the Oracle database;[5] AutoDesk AutoCAD software,[6] to create drawings of buildings and other structures; and eDNA Real-Time Historian from InStep,[7] which the Power & Water Department uses to measure utility usage. The Physical Plant works with a third-party software integration vendor to ensure interoperability among the various software applications.

To date, the Physical Plant has scanned documents stored in FileNet and now Meridian in the TIFF file format. The unit would like to move to PDF/A as its long-term preservation format for both scanned and born-digital documents, including those created in the Skire system. Large drawings, such as CAD drawings, will continue to be created and stored as TIFF files. Hard copies of all CAD drawings are also retained. If a consultant provides CAD files, all files related to the project are held in a Zip file in their native formats. The Zip files are stored in the Project Drawings Vault of the Meridian system.

Although the FAMIS database contains no documents, it interfaces with files stored in the Meridian system. Metadata includes such information as the type of room—office, classroom, or elevator—as well as a unique number for each building. Data stored in FAMIS may be exported as XML. The Meridian system references the metadata in the FAMIS database, adding that metadata to its own records.

Access to the FAMIS system is dependent on role-based authorization levels and requires an Oracle user ID and password rather than the MSU Net ID. Of 600 FAMIS users, only about 200 have data input capabilities; the other 400 are limited to add and change capabilities. Approximately 500 data tables provide the means to generate service and equipment requests, create requisitions, and track requests electronically through an interface with MSU Purchasing.

The online FIT system does not require a password and may be used to generate static reports of FAMIS activity, as noted above.

In terms of file-naming conventions, Meridian automatically assigns a filename based on the metadata provided. This filename includes the project number, document type, and a brief code based on metadata.

The Physical Plant runs nightly incremental and full weekly backups of its data. Backup tapes are stored in a fireproof vault at the power plant. Academic Technology Services (ATS) also stores a mirror of the Physical Plant's data on a NetApp server located at the MSU Computer Center. As part of a disaster recovery plan, the Physical Plant is developing a "hot site" redundant computer server facility at the James B. Henry Center, MSU's executive development center located a few miles south of the main campus.

---

[1] Physical Plant Division, Michigan State University, http://www.pp.msu.edu/. Retrieved 14 May 2010.
[2] FAMIS enterprise facility management software suite, http://solutions.oracle.com/solutions/famis/famis. Retrieved 24 May 2010.
[3] InnoCielo Meridian Enterprise engineering content management solutions, http://www.cyco.com/products/ice/. Retrieved 14 May 2010.
[4] Skire Inc., http://www.skire.com/. Retrieved 14 May 2010.
[5] Munsys Inc., http://www.munsys.com/index.htm. Retrieved 14 May 2010.
[6] AutoDesk AutoCAD software, http://usa.autodesk.com/adsk/servlet/pc/index?id=13779270&siteID=123112. Retrieved 24 June 2010.
[7] InStep eDNA Real-Time Historian, http://www.instepsoftware.com/edna_overview.asp. Retrieved 14 May 2010.

Appendix 13
MSU Digital Curation Planning Interview
Turfgrass Information Center (TIC), MSU Libraries
March 5, 2010

Turfgrass Information Center (TIC)
Pete Cookingham, Project Director—Turfgrass Information Center
Mike Schury, Library Assistant—Turfgrass Information Center
Bill Wheeler, Library Systems

Digital Curation Planning Team
Cynthia Ghering
Shawn Nicholson
Lisa Schmidt

The Turfgrass Information Center (TIC),[1] a specialized unit at the Michigan State University Libraries, was created to manage and grow the Turfgrass Information File (TGIF) database through a discipline-based rather than a geographic approach—to collect, preserve, and provide access to published and unpublished materials relating to turfgrass science, culture, and the management of turfgrass-based facilities such as golf courses, parks, sports fields, lawns, sod farms, roadsides, institutional grounds, and other managed landscapes. As the most comprehensive publicly available collection of turfgrass educational materials in the world, TGIF contains more than 150,000 records, with more than 40 percent linked to the full text of the item. Currently, three full-time employees and seventeen students work at TIC.

In 1961, MSU crop scientist Dr. James B Beard began to build the turfgrass collection with some external funding. The Library's director kept the collection going after Dr. Beard left MSU in 1975. Recognizing that MSU's teaching and research program in turfgrass science was among the top five in the world, the United States Golf Association (USGA) approached the university in the early 1980s about providing bibliographic control over the turfgrass science discipline.

To that end, in 1983 TIC chose Cuadra STAR[2] as the content management system for TGIF. The only database system designed to index and store text data, STAR is used to index all of the materials related to turfgrass management that come into the collection—including monographs, serial content, web pages, podcasts, and images. Note that most of this material is "gray literature" not available through traditional library acquisition programs, and therefore comes to TIC through donations. Approximately ten percent of the records in the database are for unowned items that came to TIC's attention. TIC licenses STAR for roughly $10,000 annually.

The current TIC staff is very happy with STAR's functionality, flexibility, and reliability and cannot conceive of moving to another system at this time. TIC has not overlooked the fact that STAR is proprietary software and may one day require an exit strategy. To that end, the software will allow files to be exported as XML, delimited ASCII, field-tagged, MARC, EndNote, or any other export specification that might be necessary.

The MSU Library's Digital Media Center (DMC) scans printed material to the TIFF format and creates issue-level PDFs that are subsequently broken down to the article level and linked by TIC to records in TGIF. Browsing sites for each serial also link to each article, as well as by sequential PDF blocks, so that pages of advertising may be retained. To date, more than 80,000 PDFs of journals, books, dissertations, and other monographs have been created. Ninety-seven percent of the PDFs are OCRed, and 87 percent of the article PDFs have searchable but non-displaying text within the corresponding TGIF records. The content is retrievable by any of the indexed fields, which includes all data elements of usefulness. Depending on the copyright status of the material or the terms of negotiated permission agreements, displayed full-text PDF content may be restricted to subscribing users rather than made publicly available.

For web material, TIC prints all webpages, including PDFs, as black and white offprints and stores them in vertical files. Links to stored web pages referenced in TGIF are checked weekly, with active links sought when broken links are detected.

TIC uses its own set of index fields (metadata) within STAR, rather than MARC, to catalog the data in TGIF. Student staff members create bibliographic records for the items in TGIF using a controlled vocabulary from TGIF's *Turfgrass Thesaurus*. This custom thesaurus includes more than 23,000 valid and invalid terms with associated syndetic structures.

TGIF also includes integrated reference content within search results, such as item entries for all listings (more than 8,000), including tables and figures, from within *Beard's Turfgrass Encyclopedia for Golf Courses, Grounds, Lawns, Sports Fields* (Beard and Beard. 2005. Michigan State University Press, 522 pp.).

Faculty, students, and practitioners in the field of turfgrass management all use TGIF. In addition to members of subscribed institutions, blanket access agreements with nine professional organizations in three countries provide access to more than 20,000 practicing turfgrass professionals worldwide. Also, anyone in the state of Michigan with a driver's license may access TGIF through the Michigan eLibrary (MEL). Of 1.3 million TGIF search presentations in 2009, 96 percent came from outside of MSU; Asia is a growing audience segment, particularly China. Within TGIF, Boolean, nesting, and proximity searches may be performed. Dynamic searches may also be picked up by Google, as STAR supports deep linking.

The TGIF bibliographic databases comprise 8-10 GBs of data, including indices but not the PDFs or other digital content. Daily incremental and weekly full backups are performed on TGIF, with

full backups kept for one month and each “fifth week” backup (once per three months) kept for three months. Both production and backup servers are located in the Library building.

No written documentation of TGIF’s STAR installation exists other than the Construction Policy Wiki for staff use, but the database structure itself is self-documenting and includes dynamic help screens. Likewise, TIC has not formally documented a collection development policy. The unit’s website lists the kinds of material relevant to turfgrass, but there is no formal policy regarding acceptable digital formats at this time.

In addition to TGIF, TIC hosts the American Society of Golf Course Architects (ASGCA) Architect’s Gallery. This includes 200 individual Real Media audio and video files and 200 architectural drawings in the proprietary DjVu format. The files, which are hosted on the MSU Academic Technology Services (ATS) HELIX server, are streamed to TIC’s website. StoreMedia from MSU’s Virtual University Design and Technology (vuDAT) is used as the front-end interface.

TIC is also creating an online image collection that will be indexed and searched via a STAR database. Slides are being digitized as TIFFs and converted to JPEGs, and the collection’s interface will be created in STAR/Web using Dreamweaver. The online image collection will tie into the TGIF database. Although the image database has yet to launch, 7,000 of 15,000 slides from the main pre-numbered series have been digitized to date. External grant funds are currently being sought to support and speed up this work.

TIC stores 11-12 TBs of digital content on Library servers, including the PDFs digitized by DMC for TIC referenced in TGIF and the JPEGs referenced in the image database along with all original TIFFs. Content on those servers is backed up to the ATS data center.

More than 60 institutions have purchased sliding-scale ($1,250 -$3,700), one-time access subscriptions to TGIF, and hundreds of corporations and individuals have subscribed in addition to the supporting organizations with blanket agreements. As this is not enough funding to fully support TIC, the MSU Library is building an endowment to ensure the unit's financial sustainability. See http://tic.msu.edu/supporters.html for a listing of major donors, ranging from $10,000 to more than $1,000,000.

---

[1] Turfgrass Information Center (TIC), MSU Libraries, Michigan State University, http://tic.msu.edu/. Retrieved 14 May 2010.

[2] Cuadra STAR QuadraStar content management system, http://www.cuadra.com/products/products.html. Retrieved 14 May 2010.

Appendix 14

Michigan State University Digital Curation Planning Project: Metadata Comparison/Crosswalk

Dublin Core Metadata Element	Art & Art History*†	MATRIX	MSU Extension	Theatre
Title	X	X	X	X
Creator	X	X	Author One Author Two	
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
Identifier	Image No. Work No.		Resource Identifier	Resource ID
Source	X	X	X	
Language			X	
Relation			X	Related Resources Website
Coverage	Date/Period Century Location (repository and/or creation) Culture Style or Period		Departments	Geo Location
Rights	Copyright	Rights Management	Copyright Expiration Date	License

Local Metadata	Collection, Area of Study, Contact Info, Checkout			Image Correction, Notes, Status, Access, Rating, Show Type, Source (camera, negative, photo)

* The Art & Art History Virtual Resource Library (VRL) is composed of several collections, accommodated in the local metadata field "Collection." The metadata listed is applicable to one or more collection type but not necessarily to all.

† Note that the Art & Art History VRL metadata uses IRIS, a data standard for cataloging and metadata management of art images based on VRA Core and CCO standards, with MDID as the public facing interface, resulting in a complex display/index duality. For simplicity's sake, the descriptive metadata in the table has been limited to the metadata displayed in browser views of MDID records.