

# Notes on Operations

## Digital Curation Planning at Michigan State University

Lisa Schmidt, Cynthia Ghering, and Shawn Nicholson

*Recognizing the need for guiding the management and preservation of Michigan State University's digital assets, a team led by university archivists and librarians conducted a digital curation planning project to explore and evaluate existing digital content and curation practices. The team used data gathered in this study to identify next steps in digital curation planning, including the recommendation to collaborate with other universities to develop solutions. While the findings were specific to Michigan State University, the process of assessing practices and identifying needs may be replicated elsewhere.*

Research universities and other large organizations continue to create and amass large volumes of digital assets and information. The inherent fragility of digital material because of technology obsolescence and physical threats such as media instability put these valuable digital assets at risk of eventual inaccessibility. In 2009, Michigan State University (MSU) confronted the problem of management and long-term preservation (curation) of its digital assets by implementing a digital curation planning project. The University Archives and Historical Collections (UAHC), MSU Libraries, and MATRIX: Center for Humane Arts, Letters, and Social Sciences Online collaborated in this exploration of digital asset management at the university with the goal of developing campus-wide guidelines for good practices in digital curation.

This paper describes the MSU project beginning with the institutional context and challenges of digital asset management faced by a major research university and explains the plans, methods, and results of the study. The project team developed and implemented a self-selective, campuswide survey of digital assets and technological infrastructures using a web-based questionnaire. After the project team analyzed the results of the questionnaire, they selected eleven units for in-depth, one-on-one interviews regarding their digital curation practices. The paper concludes with recommended next steps in digital curation planning for MSU. The MSU study may serve as a model for similar research at other universities; the MSU recommendations also may be relevant.

Throughout this paper, the term “digital” is defined as “representing information through a sequence of discrete units, especially binary code.”<sup>1</sup> The terms “digital content,” “digital resources,” “digital data,” “digital assets,” “digital material,” “digital objects,” and “digital information” are essentially interchangeable. “Digital media” refers to the physical electronic media that holds digital information, such as magnetic tape, CDs, and computers used as file servers. “Digital asset management” and “digital curation practices” refer to the handling of digital material. “Digital preservation” is defined by the California Digital Library as “the managed activities necessary for ensuring the long-term retention and usability of digital objects.”<sup>2</sup> “Digital curation” includes preservation but also takes into

**Lisa Schmidt** (lschmidt@ais.msu.edu) is Electronic Records Archivist, University Archives and Historical Collections, Michigan State University; **Cynthia Ghering** (ghering@ais.msu.edu) is Director, University Archives and Historical Collection, Michigan State University; and **Shawn Nicholson** (nichol147@mail.lib.msu.edu) is Assistant Director for Digital Information, Michigan State University Libraries, East Lansing, Michigan.

Submitted July 6, 2010; returned to authors August 8, 2010, with request to revision and submit; revision submitted October 6, 2010, and reviewed; tentatively accepted November 21, 2010, pending modest revision; revision submitted Dec. 20, 2010, and accepted for publication.

account the life cycle of the data to include 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.”<sup>3</sup>

## Literature Review

Cultural institutions have been concerned about the increasing number of digital objects for more than two decades. Writing to the museum community in 1994, Zorich warned of the pervasive problems associated with “converting their paper-based records into digital form” resulting in “millions of records.”<sup>4</sup> She drew upon the nascent experience of the U.S. Department of Energy (DOE), as that government agency proposed the then novel idea that electronic data created from the Human Genome project be “curated.” She extended the notion of digital curation first used by the U.S. DOE so that it could be applied to the digital documentation of museum collections. Focusing on the need for “consistency, long-term quality, and relevance over time,” and pointing out that any changes require verification and authentication, Zorich presciently described challenges that institutions continue to face today.<sup>5</sup>

Research universities are creating and storing digital information in volumes unthinkable to early writers such as Zorich. In the 2009 inaugural issue of the *International Journal of Digital Curation*, Tindermans described the concerns of the digital curation community, observing that “the sheer volume of digital data, the bewildering variety of formats and digital objects, which sometimes turned up in fast-changing sequences of new versions, soon became a cause of great concern.”<sup>6</sup> Without active, well-considered, long-term plans for managing

and preserving these resources, these digital assets eventually will become inaccessible because of technology obsolescence and digital media fragility. In a 2004 article describing electronic records strategies for small institutions, Cook stated, “once these digital records have been created or captured, they must then be preserved . . . each has an optimistic shelf life in digital format of perhaps 20 years before significant archival intervention is needed . . . before it either disappears as unreadable or self-destructs physically.”<sup>7</sup> Many organizations also face digital storage limitations because of the increasing size of video and other multimedia files, as well as the increasing volumes of files in general. Although storage costs have decreased, budgets have tightened in the current economy. Litigation liabilities and search difficulties result when increasing volumes of digital assets are created and stored indefinitely.

In the late 1990s and into the twenty-first century, the digital curation community produced several guidelines and best practices documents for digital preservation and curation. Among them are the Digital Curation Centre’s “Curation Reference Manual,” the Digital Preservation Coalition’s “Digital Preservation Handbook,” and the United Nations Educational, Scientific, and Cultural Organization’s (UNESCO) *Guidelines for the Preservation of Digital Heritage*.<sup>8</sup> These handbook-type materials represent an important stage in the evolution of the community. They bring focused attention to practices familiar to archivists and librarians, such as appraisal, selection, metadata creation, storage, security, and future planning, and place them into the digital context. The handbooks are also important because they contain advice and provide best practice for areas not often as comfortably familiar to librarians and archivists, including selecting file formats, using open-source

software, performing bit-level integrity checks, and employing migration and emulation techniques.

As the evolution continued, the community created two critical documents that developed a broad framework to understand what is required to provide long-term stewardship of digital information. The Center for Research Libraries (CRL) and Online Computer Library Center (OCLC) issued *Trusted Repositories Audit and Certification (TRAC): Criteria and Checklist*, based on the International Standards Organization’s Open Archival Information System (ISO OAIS) reference model, which approaches the problem of good digital preservation practice by articulating eighty-four criteria for a trusted digital repository.<sup>9</sup> The criteria are organized into three sections: “Organizational Infrastructure,” which includes administrative, management, and financial support, as well as copyright; “Digital Object Management,” which includes the accession, storage, and provision of access to digital objects; and “Technologies, Technical Infrastructure, and Security,” which includes backup and disaster recovering planning. All criteria require evidence of compliance, preferably as written documentation and policies.

Research data poses a different set of challenges and its own growing body of literature. As it relates to this project, Gold’s review of the library’s role in data curation highlighted milestones such as the establishment of a data curation education program at the University of Illinois and the Johns Hopkins Sheridan Library’s leadership of a multi-institutional effort to build a national infrastructure for curation of research datasets.<sup>10</sup> Offering a more granular view, Walters provided a retelling of Georgia Institute of Technology’s journey toward curating scientific data and developed a roadmap that can easily be adapted to other institutions.<sup>11</sup> Mirroring the broader community’s evolution,

Abrams, Cruse, and Kuntz emphasized the need to constantly reevaluate and refocus curatorial activities.<sup>12</sup> They discussed how the California Digital Library has moved away from a singular focus on preservation toward a more encompassing “programmatic, rather than a project-oriented approach; and a renewed emphasis on services, rather than systems.”<sup>13</sup>

As new tools and expectations arise, richer exploration options and an expansion of existing best practice are needed. The results of two large-scale surveys that explored perceived needs and current practices were published in 2009. Sinclair and colleagues surveyed 172 national libraries, archives, and other cultural heritage organizations in Europe to “better understand the organization’s digital preservation activities and needs.”<sup>14</sup> While the respondents expressed a high level of awareness of the challenges of preservation, the authors found no corresponding level of policy and budgetary development. An Educause Center for Applied Research study by Yanosky examined data management in U.S. higher education.<sup>15</sup> Yanosky employed a multifaceted research design consisting of a quantitative web-based survey, qualitative interviews, and case studies; 309 responded to the survey portion and 23 participated in targeted interviews. The results confirmed the vast amount of data (described as “data big bang”) that is being created and that data stewardship and security are significant concerns. When asked about enterprise-wide content management, “only about 12 percent of 304 respondents [said that their institution] has an integrated enterprise content management solution.”<sup>16</sup>

Although the literature is replete with discussions of digital curation theory development and evolving best practices, no known studies describe an institution-wide effort to understand curatorial activities across the full range of digital objects.

### **Institutional Context and Challenges of Digital Asset Management**

Like other research universities, MSU has amassed a growing body of digital assets and information—including institutional records, faculty and student research, theses and dissertations, university publications, multimedia collections, digital surrogates of cultural material, learning objects, course materials, and more. Much time, effort, grant funding, human capital, and research have gone into creating these digital resources—some of which only exist in digital form. In response to the need to manage digital content, some MSU colleges and departments have started their own digital repositories. No comprehensive, campuswide digital preservation strategy or set of guidelines exists, however, and MSU has not established an institutional repository.

The University Archives and Historical Collections (UAHC) (<http://archives.msu.edu/index.php>), the official repository for MSU’s historical archives, began looking into the problem of cross-campus digital preservation in early 2009. Along with its mandate to collect, preserve, and provide access to the historical records of the university, the UAHC also is charged with assisting departmental and administrative units in the efficient administration, management, and preservation of all official university records, including electronic records.

In the winter of 2009, the UAHC engaged a digital preservation intern from the University of Michigan’s School of Information to research a strategy for preserving and making accessible MSU’s multimedia productions, including born-digital film, audio, and still images. This internship was funded through an Institute of Museum and Library Services (IMLS) grant.<sup>17</sup> Under the direction of the UAHC director, the intern identified file formats, recommended workflows,

and addressed the long-term preservation of electronic media. During the course of the project, UAHC staff and the intern held interviews with stakeholders at seven campus units: the Digital Media Center/Vincent Voice Library at MSU Libraries; MATRIX; Michigan Government Television, a public affairs initiative of Michigan’s cable television industry; Sports Broadcasting; University Relations, which handles communications and public relations; Virtual University Design and Technology, which specializes in creating online courses and e-learning tools; and Broadcasting Services, including the WKAR public television and radio affiliates.

The project concluded with a report to the UAHC director that included several high-level recommendations for the creation, storage, preservation, and long-term access to multimedia university records. The intern suggested surveying more campus units to broaden the analysis of the university’s digital assets, creating best practice guidelines for selecting and appraising content, file formats, naming conventions, and metadata; providing better long-term storage options; and establishing an institutional repository. This digital preservation internship project, while not groundbreaking, raised staff awareness and led to a more formalized and significant environmental scan of the institution’s strengths and weaknesses in this area.

Building on this initial research, the director proposed developing a more comprehensive environmental scan and gap analysis inspired by the Inter-University Consortium for Political and Social Research (ICPSR) digital preservation management workshop.<sup>18</sup> The purpose of the project was to develop a digital preservation plan with a focus on practical solutions using MSU’s current resources; the initiative recommended a collaborative effort of the UAHC, the MSU Libraries, and MATRIX.<sup>19</sup> The

resulting proposal received top-level support, with the university's vice provost of libraries, computing, and technology committing funding for a half-time digital preservation analyst to manage the project for one year beginning in July 2009.

To that end, the director of UAHC created a cross-university Digital Preservation Planning Project team and brought in an electronic records archivist then working for MATRIX to serve as the digital preservation analyst. The team, which served to guide the analyst's work and to encourage participation from institutional stakeholders, consisted of staff from UAHC, the MSU Libraries, MATRIX, MSU Extension/Agriculture and Natural Resources Technology Services, campus Information Technology's (IT's) data services, and the Office of the Registrar. During monthly meetings, the team reviewed activities and progress, anticipated next steps and potential challenges, and analyzed findings.

As the project got underway, the team quickly realized that the initial plan of an all-campus environmental scan was overly ambitious. The team could not complete an exhaustive survey of all university digital assets within one year. Even if more staff resources were available, an all-encompassing inventory would require a significant effort and result in diminishing returns because of inevitable redundancies. Perhaps most important, the team was concerned that a comprehensive inventory would create the perception that the project's desired outcome was to build a one-size-fits-all data repository. In MSU's heavily decentralized environment, both academic and administrative units are apprehensive of solutions that may result in loss of control of the digital assets of their college or unit. Therefore, instead of conducting a comprehensive survey, the team decided to focus on the variety of digital content and the disparate

needs of the different campus units.

To address the concerns stated above, the project team refined the scope of work and deliverables. The team decided to change the name of the project, replacing the term "digital preservation" with "digital curation." This new project name emphasized the focus of the study on the life cycle of the university's digital assets, including preservation, creation, and management. Instead of conducting a comprehensive inventory, the team administered an online survey that would provide a sampling of digital assets and repositories based on a campus-wide, self-selective, web-based questionnaire. The team later conducted in-depth interviews with a limited number of 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 also were reviewed.

By offering practical guidance and influencing policy, MSU's archivists and librarians sought to expand on their traditional role as custodians of physical material. As Cox stated 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."<sup>20</sup> The information professional as policymaker is not a new idea. Bearman maintained in a 1991 paper that in the interests of gaining respect as professionals, archivists and records managers should move away from the custodial role and "reposition themselves as policy makers and regulators whose task it is to assure that managers throughout the organization demonstrate awareness of the institutional significance of information by retaining and destroying information at appropriate times and in appropriate ways."<sup>21</sup> Thus 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.

## Self-Selective, Campuswide Survey

### Research Method

In October 2009, the digital curation planning project team designed and conducted a voluntary, web-based survey using a questionnaire created with the online SurveyMonkey tool. UAHC publicized the survey through MSU's internal e-mail-based communications network for information technology employees, the weekly *MSU News* online newsletter, and the project website (<http://msudcp.archives.msu.edu>). These notifications emphasized the need for digital curation planning and encouraged involvement of technology staff and content creators in the survey as a first step. At the same time, the strictly voluntary nature of participation was stressed. The digital curation planning project team made the questionnaire available for two weeks.

Survey participants were asked about types of digital content created by their units, the approximate volume of digital content in terabytes, storage media used, file formats created, online storage capacity and storage expansion plans, and any content management system (CMS) or digital repository software used. (Neither "CMS" nor "digital repository software" was defined in the questionnaire, allowing survey takers to respond according to their own interpretations of the terms.) The questionnaire appears in appendix A.

### Findings

The questionnaire received 90 responses, including 23 responses



from academic departments, 31 responses from administrative services units, 9 responses from research centers and institutes, and 27 responses from technology services organizations. This is a good response rate for a large research university such as MSU. As of fall 2009, MSU offered more than 200 programs from 17 academic degree-granting colleges, had a student body of more than 47,000, and supported more than 50 administrative services units and 70 research centers.<sup>22</sup> Responses that came from organizations that provide technical services were placed in the “technology services” category, even if those organizations officially belonged to academic or administrative units. Because the link to the questionnaire was distributed through an e-mail list and made publicly available through the online MSU staff newsletter, multiple staff per unit could respond.

Academic departments represented in the questionnaire responses 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. Participating administrative units ranged from the Controller’s Office, the Capital Asset Management Department, 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. The research centers included the Confucius Institute, the Cyclotron, the Julian Samora Research Institute, and MATRIX. In contrast, all of the technology services responses came from only 6 organizations.

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 imaged or scanned paper documents, word processing and spreadsheet documents, and databases. Research data, audiovisual material, word processing documents, and programming code predominated at the research centers. The technology services organizations noted that most of their digital content consisted of code, databases, and webpages.

File formats comprising the largest proportion of a given unit’s digital content also varied. The academic departments surveyed noted the existence of PDFs, Statistical Package for the Social Sciences (SPSS) and Statistical Analysis System (SAS) statistical formats, TIFFs, JPEGs, MySQL, and Camtasia video formats. Various database formats, TIFFs, text, MS Office formats, as well as audio and video formats, prevailed at the administrative units. The research centers reported sizeable concentrations of video formats, PHP code, MS Word, and SAS, and the technology services organizations carried large proportions of text and programming code formats.

Most of the units reported that they 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 reported storing data on cassette tapes. Seventeen units plan to increase online storage capacity soon, most from 1 to 10 terabytes, with 6 units planning to add 30 terabytes or more.

Twenty-three units responded that they have implemented or plan to implement a CMS, digital repository software, or both. Again, neither “CMS” nor “digital repository” was defined in the questionnaire, so unit representatives responded using their understanding of the terms. CMSs 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 CMS and the digital repository application. Tools (including Concurrent Versions System (CVS), Git, Adobe Version Cue, and Subversion), more properly known as version control software, were identified as CMSs or digital repository software by some units. The database application FileMaker was listed as a CMS 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 respondents provided additional comments stating great interest and enthusiasm in the digital curation planning project’s goal of providing curation guidance. One administrative unit noted, “This is a timely survey, because our unit is at a point where we have 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 unit asked for guidelines on how to handle archive-worthy files at the time of creation rather than storing everything and later subjecting the unit to an arduous appraisal process. Respondents also expressed interest in guidance on choosing a digital asset management system.

## One-on-One Interviews

### Research Method

The digital curation planning project team chose 11 units that responded to the survey to interview, focusing on the units that reported using a CMS, digital repository solution, or both. The team determined that these units would have concentrations of digital content they intentionally manage or preserve and thus would be good resources for the project. These units might already have solutions in place that could be folded into guidelines useful to other units. Conversely, these units might need the most help and guidance from the project. The team also was 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 interviewed were

- Broadcasting Services ([www.wkar.org](http://www.wkar.org)), including the WKAR public radio and TV stations, which produces digital content of historical value to the university that should be transferred to UAHC.
- Center for Research on Mathematics and Science Education (CRMSE), which focuses on grant-funded research into the teaching of mathematics and science. CRMSE is primarily concerned with the preservation of survey instruments and research data, and the unit is interested in transferring these files to UAHC for long-term preservation.
- Confucius Institute at Michigan State University (CIMSU) ([www.experiencechinese.com](http://www.experiencechinese.com)), which is part of a worldwide network that promotes the teaching of Chinese language and culture. The unit creates multimedia content for use in online training and educational games.
- Department of Art and Art History ([www.art.msu.edu](http://www.art.msu.edu)), which hosts the Visual Resources Library (VRL) of digital art images.
- Department of Theatre ([www.theatre.msu.edu](http://www.theatre.msu.edu)), which stores a mix of digital photos of performances, CAD drawings of sets, and other performance-related digital files.
- MATRIX: Center for Humane Arts, Letters, and Social Sciences Online ([www.matrix.msu.edu](http://www.matrix.msu.edu)), a digital humanities research center and host for major digital libraries of cultural material, including the African Online Digital Library (AODL), Detroit Public Television's *American Black Journal* video archives, Historical Voices, and the Quilt Index.
- MSU Extension/Agriculture and Natural Resources (ANR) Technology Services, which provides communities across Michigan with programming focused on agriculture and natural resources; children, youth, and families; and community and economic development.
- National Superconducting Cyclotron Laboratory (NSCL) ([www.nscl.msu.edu](http://www.nscl.msu.edu)), a National Science Foundation (NSF)-funded, world-leading laboratory for rare isotope research and nuclear science education.
- Physical Plant Division ([www.pp.msu.edu](http://www.pp.msu.edu)), which maintains all buildings and land entities on and off campus and provides utility services. In addition to its operations-related digital management systems, Physical Plant holds architectural drawings and building maps that may

have historical significance.

- Turfgrass Information Center (TIC), MSU Libraries (<http://tic.msu.edu>), which manages the Turfgrass Information File (TGIF) database of 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.
- University Relations (<http://www2.ur.msu.edu>), the unit responsible for public relations, creates and maintains large volumes of digital photographic and video records of archival value, some of which should be transferred to UAHC. Digital video footage includes the *MSU Today* show on the Big Ten Network ([www.bigtennetwork.com](http://www.bigtennetwork.com)), a joint venture between Fox cable television networks and the Big Ten Conference, a U.S. collegiate athletics organization.

Digital curation planning team members conducted the one-on-one interviews January–March 2010. The meetings were informal, two-hour conversations held at the unit offices. Discussion topics included how the unit's digital content relates to the mission of the unit, whether it was of ongoing use or of archival value—that is, whether it documents the activity of the unit or the university file formats used, and storage infrastructure, including any space issues.

The team asked about the CMS(s) or digital repository software used, why they were chosen, and how they are used. The team also asked about processes and workflows of ingesting data and digital content into the system, archival storage and preservation of content, and content retrieval, as

well as whether the unit had a means to ensure file integrity. Finally, the team asked about the use of metadata stored with or related to the content as well as file naming conventions. (See appendix B for the types of questions asked.)

### Findings and Analysis

In analyzing the results of the interviews, the team noted several key points. Each unit had devised solutions that fit the mission of the unit, the nature of the data, and the needs of users. Some units use commercial applications and some use open-source software. The Turfgrass Information Center at the MSU Libraries, for example, has long used the commercially available Cuadra STAR database and CMS, and the Department of Theatre uses the relatively new open-source ResourceSpace digital repository solution. See appendix C for a list of the content management, digital repository, and other software used by the interviewed units to manage digital assets.

Some units, such as Broadcasting Services, hold digital content of archival value to the university. Other units, such as the Department of Art and Art History and MATRIX, create and store digital materials that is produced on behalf of partner cultural organizations. These are both important digital resources, but are not institutional records that must adhere to a defined retention schedule.

Many of the interviewed units exhibited good digital preservation and curation practices. Most backed up their data in some manner. Some used detailed metadata to describe their digital assets, and many were using repository software with good access and discovery interfaces to manage their content. Many of the units have strong support from their administrative management and stable funding.

Most 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 audiovisual content that has been converted to access formats for use in the KORA digital repository. The Turfgrass Information Center stores scans of printed material and slides as TIFF files but makes them available online in PDF and JPEG formats. Likewise, the Department of Art and Art History keeps TIFF master files while providing JPEG files as access copies in the Visual Resources Library. 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 the MSU Extension Knowledge Repository. Both Broadcasting Services and the Confucius Institute maintain some audio files in the archival WAV format. Physical Plant currently stores and scans documents in TIFF but would like to move to PDF/A as a preservation master format. The Center for Research on Mathematics and Science Education (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. 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. Using Adobe Bridge, the Department of Art and Art History can detect file corruption when viewing thumbnail photos; likewise, the Adobe Photoshop script that creates contact sheets of thumbnails will stop

running if it encounters a corrupt file. Soon 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, which record the structure of files that need to be protected. If the original files become corrupted, the parity files may be used to restore them.

MSU Extension and the Department of Art and Art History have formal file naming conventions in place. The Physical Plant Division's Meridian system 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 curation guidelines and said 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 project team found little in the way of documented digital curation policies. The lack of good digital curation practice is unfortunate considering that the interviewed units are more likely to have invested in digital asset management compared to the rest of campus. The team suspects that most campus units are either unaware of the need for or unable to address digital curation at this time.

The team also compared the metadata schema used by the units interviewed with the Dublin Core (DC) metadata element set, a standard in the information science field for describing digital objects.<sup>23</sup> The comparison involved examining the metadata element sets used by the units, noting correspondences with the DC element set, and reviewing definitions of elements in data dictionaries for schema when necessary to understand the information that they were intended to represent. For example, MSU Extension's "Author One" and "Author Two" elements correspond to the DC "Creator" element. In some cases, the analyst contacted original unit interviewees for in-depth explanations of particular elements.

Six of the units interviewed had implemented metadata schema that could be considered in the comparison. MSU Extension, MATRIX, and the Department of Theatre use metadata schema based on or similar to DC, with slight variations to reflect local needs. The Department of Art and Art History uses the Image Resource Information System (IRIS) cataloging utility for describing its art images with metadata based on the Visual Resources Association (VRA) Core and the Cataloging Cultural Objects (CCO) guide to good cataloging practices; this metadata maps roughly to DC.<sup>24</sup> The metadata used by Physical Plant and the Turfgrass Information Center, the other two units in the comparison, do not correspond directly to the DC metadata set. Physical Plant uses 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 Cuadra STAR system for creating descriptive metadata for bibliographic information about physical and digital material related to turfgrass.

### Digital Storage Solution Planning at Michigan State University

In tandem with the exploration into digital curation practices, the digital curation planning project team kept abreast of and helped to influence digital storage planning at MSU. The Libraries, Computing, and Technology (LCT) division provides central technology support for administrative business systems, e-mail, academic, and network services throughout the campus. A strong tradition of local units maintaining their own IT staff and managing their own systems also exists, however. Like many other institutions, MSU is looking closely at this divide between centralized and local IT to discern how best to use and consolidate storage and other technology functions. One recognized advantage of a strong central IT organization is that it can more effectively manage electronic records and digital assets.

Currently, LCT is developing virtual server environments and price structures as a storage solution to local units. LCT also is considering 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 of 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.

The Committee on Institutional Cooperation (CIC), a consortium of Big Ten universities and the University of Chicago, is developing collaborative storage solutions that would better enable effective, efficient stewardship of campus assets and cutting-edge scholarship. The CIC expects common architecture, infrastructure, and operating environments to increase

economies of scale, permit shared management, and provide for the research and development of such value-added services as community tagging and annotation, citation tracking, and digital curation for scholarly data. Many of these efforts derive from the CIC chief information officers' 2010 report, "A Research Cyberinfrastructure Strategy for the CIC: Advice to the Provosts from the Chief Information Officers."<sup>25</sup> These collaborative storage solutions are separate from HathiTrust ([www.hathitrust.org](http://www.hathitrust.org)), which aims to build a reliable and comprehensive digital archive of library materials converted from print that is co-owned and managed by a number of academic institutions and in which the CIC partners.

### Next Steps in Digital Curation Planning

The digital curation planning project team observed that despite the variety of digital content and formats as well as different approaches to curation, commonalities and patterns between the units interviewed emerged. Studying these patterns led to the identification of four basic types of digital content created at MSU: university publications, including e-journals and electronic theses and dissertations; digital content that documents the history of the university, such as the photos and video of University Relations and some of Broadcasting Services' audiovisual programming; nonuniversity digital content, such as the digital files created and managed by MATRIX and the Department of Art and Art History; and research data.

With an understanding of the types of digital content in hand, the curation needs of the units can be addressed and functional specifications for curation solutions developed and applied as needed. For example, university publications will require deposit in an institutional repository



and the implementation of a distributed preservation solution such as Lots of Copies Keep Stuff Safe (LOCKSS) ([www.lockss.org](http://www.lockss.org)). Digital content that documents the history of the university will require digital curation and appraisal guidelines as well as mechanisms for transfer to and storage with UAHC. Digital content not specific to MSU will benefit from curation guidelines on metadata, file formats, repository software, file integrity checking, consistent file naming conventions, and storage and backup planning. Units that create and manage research data will require support in meeting the new National Science Foundation (NSF) requirements for grant proposals to include a data management plan that addresses preservation, access, and other elements of digital curation.<sup>26</sup>

In keeping with the approach of identifying user needs and developing functional requirements, the digital curation planning project team recommended several next steps to guide the management and preservation of MSU's digital assets. For example, UAHC can provide digital asset appraisal assistance to university departments and units, especially those holding digital assets of historical value that should be transferred to UAHC for long-term preservation, as is planned for University Relations. Work will continue with LCT on the development of tiered storage plans as well as plans for transferring digital content of archival value to UAHC. In that same vein, UAHC will develop guidelines to quickly determine whether digital assets should be transferred for permanent preservation.

UAHC, the MSU Libraries, and representatives from the other members of the digital curation planning project team will work together to explore the digital curation practices of units holding significant digital content that were not represented in this project, particularly those with other types of content and with different

content management practices; this will include further investigation of research data curation across the campus. They will develop general best and good practices in digital curation recommendations and guidance, keeping in mind differences in the missions of the units and the types of digital material that they create and manage.

With the MSU Libraries and other digital curation planning team principals, UAHC will develop metadata standards for university records, including publications and digital content that documents the history and activity of the university. UAHC and the MSU Libraries will work together to develop digital data curation toolkits that acknowledge researchers and units as information producers as well as consumers. Topics covered by these toolkits will include file formats, documentation, intellectual property rights, sharing and dissemination, and preservation.

The team also recommended the fostering of "Communities of Practice" through online forums and meetings, in which campus units and other institutions have the opportunity to share digital curation experiences, generate new ideas, and collaborate on initiatives. UAHC and the MSU Libraries could work with their counterparts in the digital humanities and at other CIC member institutions to obtain grant funding to explore the digital curation problem across institutions and develop common best and good practices guidelines.

Finally, the team advocated the creation of a senior-level digital preservation officer position at MSU. This individual could continue to raise the visibility of digital curation, focus the coordination of curation and preservation resources across campus for both academic and administrative data types, and direct the dissemination of digital curation guidelines and best practices. Although economic conditions prohibit hiring a senior-level digital preservation officer for the

university now, a new digital curation librarian position has been created in the Digital Information Services unit of the MSU Libraries.

## Conclusion

In response to the need for ensuring the viability of the valuable digital assets created in ever-larger volumes at MSU, a year-long digital curation planning project explored current practices in the creation and management of digital material. Data was gathered and observations made through a self-selective, campuswide online survey and one-on-one interviews with campus units that demonstrated some level of curation practice or held material of historical interest to the university.

Although the digital curation planning project team initially found the variety of digital content and formats overwhelming, patterns emerged that will make addressing the problems of digital curation at MSU easier. Four types of digital content were identified, and user needs can be articulated and functional specifications developed to meet those needs. By investigating the needs of more campus units and continuing to build on these patterns, digital curators can make sense of the jumble of digital content at MSU and develop solutions that also may be of use to other institutions.

Developing digital curation guidelines for Michigan State University will be an iterative process. Recommended next steps include UAHC providing appraisal assistance to units holding material of archival value, such as University Relations; studying and influencing the curation of scholarly research data; developing of common metadata standards and curation toolkits; fostering of "Communities of Practice" within the university and with partner institutions; and working with other universities to obtain grant funding for the study of digital curation practices across institutions. By

surveying and interviewing a variety of departmental and administrative units, the digital curation planning project team began developing an understanding of the digital assets and related needs and concerns of the MSU community, building trust, and establishing new relationships that will aid in moving forward with an institutional approach to digital curation planning. While the MSU findings are specific to one institution, the process of assessing practices and identifying needs can be replicated elsewhere.

### References

1. Richard Pearce Moses, "A Glossary of Archival and Records Terminology," Society of American Archivists, [www.archivists.org/glossary/term\\_details.asp?DefinitionKey=340](http://www.archivists.org/glossary/term_details.asp?DefinitionKey=340) (accessed Nov. 30, 2010).
2. California Digital Library (CDL), "Glossary," <http://www.cdlib.org/gateways/technology/glossary.html?field=institution&query=CDL&action=search>, (accessed Nov. 30, 2010).
3. Digital Curation Centre, "DCC Charter and Statement of Principles," 2010, [www.dcc.ac.uk/about-us/dcc-charter](http://www.dcc.ac.uk/about-us/dcc-charter) (accessed Apr. 9, 2010).
4. Diane Zorich, "Data Management: Managing Electronic Information: Data Curation in Museums," *Museum Management & Curatorship* 14, no. 4 (1995): 431.
5. *Ibid.*, 430–32.
6. Peter Tindermans, "Key Stakeholders Pledge to a Strategic Approach to Preserve the Digital Records of Science," *International Journal of Digital Curation* 1, no. 1 (2009): 84, [www.ijdc.net/index.php/ijdc/article/viewFile/13/8](http://www.ijdc.net/index.php/ijdc/article/viewFile/13/8) (accessed Dec. 23, 2010).
7. Terry Cook, "Byte-ing Off What You Can Chew: Electronic Records Strategies for Small Archival Institutions," *Archifacts* (Apr. 2004), [www.aranz.org.nz/Site/publications/papers\\_online/terry\\_cook\\_paper.aspx](http://www.aranz.org.nz/Site/publications/papers_online/terry_cook_paper.aspx) (accessed May 26, 2010).
8. Digital Curation Centre, "Curation Reference Manual," 2010, [www.dcc.ac.uk/resources/curation-reference-manual](http://www.dcc.ac.uk/resources/curation-reference-manual) (accessed June 21, 2010); Digital Preservation Coalition, "Digital Preservation Handbook," (2009), [www.dpconline.org/advice/preservationhandbook](http://www.dpconline.org/advice/preservationhandbook) (accessed Sept. 20, 2010); National Library of Australia, *Guidelines for the Preservation of Digital Heritage* (United Nations Educational, Scientific and Cultural Organization, 2003), <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf> (accessed June 21, 2010).
9. The Center for Research Libraries and Online Computer Library Center, *Trustworthy Repositories Audit and Certification: Criteria and Checklist*, version 1.0 (Chicago: CRL; Dublin, Ohio: OCLC, 2007), [www.crl.edu/sites/default/files/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf) (accessed June 2, 2010); Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-B-1 Blue Book issue 1 (Washington, D.C.: CCSDS, 2002), <http://public.ccsds.org/publications/archive/650x0b1.pdf> (accessed Sept. 13, 2010).
10. Anna Gold, "Data Curation and Libraries: Short-Term Developments, Long-Term Prospects" (Apr. 14, 2010), [http://digitalcommons.calpoly.edu/lib\\_dean/27](http://digitalcommons.calpoly.edu/lib_dean/27) (accessed Sept. 29, 2010).
11. Tyler Walters, "Data Curation Program Development in U.S. Universities: The Georgia Institute of Technology Example," *International Journal of Digital Curation* 3, no. 4 (2009): 83–92, [www.ijdc.net/index.php/ijdc/article/viewFile/136/153](http://www.ijdc.net/index.php/ijdc/article/viewFile/136/153) (accessed Dec. 23, 2010).
12. Stephen Abrams, Patricia Cruse, and John Kunze, "Preservation Is Not a Place," *International Journal of Digital Curation* 4, no. 1 (2009): 8–21, [www.ijdc.net/index.php/ijdc/article/viewFile/98/73](http://www.ijdc.net/index.php/ijdc/article/viewFile/98/73) (accessed Dec. 23, 2010).
13. *Ibid.*, 8.
14. Pauline Sinclair et al., "Are You Ready? Assessing Whether Organisations are Prepared for Digital Preservation," (paper presented at iPRES 2009: The Sixth International Conference on Preservation of Digital Objects, Oct. 5–6, 2009), [www.escholarship.org/uc/item/8dd2m5qw](http://www.escholarship.org/uc/item/8dd2m5qw) (accessed Sept. 13, 2010).
15. Ronald Yanosky, *Institutional Data Management in Higher Education*, Research Study no. 8 (Boulder, Colo.: EDUCAUSE Center for Applied Research, 2009), [www.educause.edu/Resources/nstitutionalDataManagementinH/191754](http://www.educause.edu/Resources/nstitutionalDataManagementinH/191754) (accessed Sept. 29, 2010).
16. *Ibid.*, 17.
17. University of Michigan, Engaging Communities—Fostering Internships for Preservation and Digital Curation, "About the Project," <http://preservation.cms.si.umich.edu> (accessed Dec. 23, 2010).
18. Anne R. Kenney et al., "Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems," (online tutorial and workshop, Cornell University Library, 2003, version 1); Inter-University Consortium for Political and Social Research (ICPSR), University of Michigan, 2007), [www.icpsr.umich.edu/dpm/dpm-eng/eng\\_index.html](http://www.icpsr.umich.edu/dpm/dpm-eng/eng_index.html) (accessed May 25, 2010).
19. MSU Digital Preservation Proposal—April 2009: Project: Preserving MSU's Digital Assets, <http://msudcp.archives.msu.edu/wp-content/uploads/2010/12/MSU-DPP-Proposal.pdf> (accessed Dec. 23, 2010.)
20. Richard Cox, "The Academic Archives of the Future," *EDUCAUSE Review* 43, no. 2 (2008), [www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume43/heAcademicArchivesoftheFuture/162683](http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume43/heAcademicArchivesoftheFuture/162683) (accessed June 3, 2010).
21. David Bearman, "An Indefensible Bastion: Archives as a Repository in the Electronic Age," *Archival Management of Electronic Records, Archives and Museum Informatics Technical Report* 13 (Pittsburgh: Archives & Museum Informatics, 1991): 17.
22. Michigan State University, "MSU Facts," 2010, [www.msu.edu/about/thisismsu/facts.html](http://www.msu.edu/about/thisismsu/facts.html) (accessed Oct. 6, 2010).
23. Dublin Core Metadata Initiative, "Dublin Core Metadata Element Set, Version 1.1," Oct. 11, 2010, <http://>

- dublincore.org/documents/dces (accessed Dec. 23, 2010).
24. Visual Resources Association, "VRA CoreFour," [www.vraweb.org/projects/vracore4/](http://www.vraweb.org/projects/vracore4/) (accessed May 14, 2010); CCO Commons: Cataloging Cultural Objects, "Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images," [www.vraweb.org/ccoweb/cco/about.html](http://www.vraweb.org/ccoweb/cco/about.html) (accessed May 24, 2010).
25. Committee on Institutional Cooperation Chief Information Officers, "A Research Cyberinfrastructure Strategy for the CIC: Advice to the Provosts from the Chief Information Officers," 2010, [www.cic.net/Libraries/Technology/2010Report\\_-\\_6\\_21reduced.sflb.ashx](http://www.cic.net/Libraries/Technology/2010Report_-_6_21reduced.sflb.ashx) (accessed Sept. 16, 2010).
26. National Science Foundation, "Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans," Press Release 10-077, May 10, 2010, [www.nsf.gov/news/news\\_summ.jsp?cntn\\_id=116928&org=NSF](http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928&org=NSF) (accessed June 2, 2010).

## Appendix A. Michigan State University Digital Curation Planning Project 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 Publisher
- 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](mailto:lisa.schmidt@matrix.msu.edu).

### **Appendix B. 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 C. Michigan State University Digital Curation Planning Project One-on-One Interview Results

Unit and Scope	Content Management System, Digital Repository, and Other Software	Metadata
Confucius Institute at MSU	Alfresco content management system, to manage project workflow ( <a href="http://www.alfresco.com">www.alfresco.com</a> )	
Department of Art & Art History	Open-source, web-based Madison Digital Image Database system to manage, share, and organize digital images ( <a href="http://mdid.org/overview.htm">http://mdid.org/overview.htm</a> )	Image Resource Information System (IRIS) data standard, uses Visual Resources Association (VRA) Core, Cataloging Cultural Objects (CCO)
Department of Theatre	DOT:Media, a digital repository created using ResourceSpace digital asset management software ( <a href="http://www.resourcespace.org">www.resourcespace.org</a> );	Based on Dublin Core
MATRIX	In-house developed open-source KORA digital repository software ( <a href="http://www2.matrix.msu.edu/research/technology/kora">www2.matrix.msu.edu/research/technology/kora</a> )	Based on Dublin Core
MSU Extension/Agriculture and Natural Resources Technology Services	Knowledge Repository ( <a href="http://www.msue.msu.edu/portal">www.msue.msu.edu/portal</a> ), digital repository system custom-developed by Intrafinity ( <a href="http://www.intrafinity.com">www.intrafinity.com</a> )	Based on Dublin Core
National Superconducting Cyclotron Laboratory (NSCL)	In-house developed, open-source —NSCL Data Acquisition System nuclear physics data acquisition software ( <a href="http://sourceforge.net/projects/nscldaq">http://sourceforge.net/projects/nscldaq</a> ); —NSCL SpecTeL Histogramming System, an open-source C++-based analysis package for nuclear physics data ( <a href="http://sourceforge.net/projects/nsclspectcl">http://sourceforge.net/projects/nsclspectcl</a> )	
Physical Plant Division	—Oracle-based FAMIS enterprise facility management software suite to manage operations ( <a href="http://solutions.oracle.com/solutions/famis/famis">http://solutions.oracle.com/solutions/famis/famis</a> ) —InnoCielo Meridian Enterprise software for document management ( <a href="http://www.cyco.com/products/ice/">www.cyco.com/products/ice/</a> ) —Skire project management software to track vendor activity ( <a href="http://www.skire.com/">www.skire.com/</a> ) —Munsys spatial data management software to map underground utilities ( <a href="http://www.munsys.com/index.htm">www.munsys.com/index.htm</a> ) —InStep eDNA Real-Time Historian to measure utility usage ( <a href="http://www.instepsoftware.com/edna_overview.asp">www.instepsoftware.com/edna_overview.asp</a> )	Local controlled vocabulary for project transactions
Turfgrass Information Center of the MSU Libraries	Cuadra STAR content management system ( <a href="http://www.cuadra.com/products/products.html">www.cuadra.com/products/products.html</a> )	Customized indexing terms
University Relations	—Extensis Portfolio media management system for indexing photos and NetPublish Portfolio for public access ( <a href="http://www.extensis.com/en/home.jsp">www.extensis.com/en/home.jsp</a> ) —Zenfolio online delivery system ( <a href="http://www.zenfolio.com/">www.zenfolio.com/</a> )	