MSU TDR Digital Preservation Program

The MSU Trusted Digital Preservation Program contains information on the Administrative Policy Framework, Digital Preservation Policies and Procedures, Technology Infrastructure, and investigation and applications of best practices and standards of OAIS and TRAC.

The Administrative Policy Framework contains the following sections:

1. Purpose
2. Mandate
3. Objectives
4. Scope
5. Challenges and Incentives
6. Operating Principles
7. Roles and Responsibilities
8. Cooperation and Collaboration
9. Selection and Acquisition for Preservation
10. Access and Use

The table of contents for the entire MSU TDR Digital Preservation Program is below.

Administrative Policy Framework

This Digital Preservation Policy for the Trusted Digital Repository of the Michigan State University Archives & Historical Collections (UAHC) follows digital preservation standards as defined in OCLC’s Trusted Digital Repositories: Attributes and Responsibilities. The attributes of a trusted digital repository include:

- **Open Archival Information System (OAIS) compliance**
  Develop digital preservation policies, repository, and strategies in accordance with the Open Archival Information System (OAIS) Reference Model.

- **Administrative responsibility**
  Accept responsibility for the long-term maintenance of digital resources on behalf of depositors and for the benefit of current and future users.

- **Organizational viability**
  Establish an organizational system that supports not only long-term viability of the repository, but also the digital information for which it has responsibility.

- **Financial sustainability**
  Demonstrate fiscal responsibility and sustainability.

- **Technological and procedural suitability**
  Develop policies, practices, and performance that can be audited and measured.

- **Systems security**
  Ensure the ongoing management, access, and security of deposited materials.

- **Procedural accountability**
  Dependably carry out the repository’s long-term responsibilities to depositors and users openly and explicitly.

Purpose

The Michigan State University Archives & Historical Collections (UAHC), in keeping with its mission and historical mandate, is responsible for collecting, preserving, and providing long-term access to the inactive records of all units in the university that have legal, administrative, fiscal, or historical value, including those in digital format. The Digital Preservation Policy supports this mission and is UAHC’s highest level digital preservation policy document. The Policy makes explicit UAHC’s commitment to preserving the university’s digital records through a comprehensive digital preservation program and contains references to other relevant MSU and UAHC policies and procedures. The audience includes UAHC staff, MSU administrators, other digital content contributors, donors, and researchers (students, faculty, alumni, and members of the public).

Mandate
The mandate for digital preservation at UAHC includes the following:

- **Institutional responsibility and scholarly commitment:** UAHC has a mission and historical mandate to collect, preserve, and provide access to the institutional records, scholarly records, and historical collections in its holdings, including digital material. The digital preservation program also supports the broader scholarly mission of Michigan State University. By following institutional policy in regards to records management, UAHC ensures legal compliance.

- **Grants:** In 2009, UAHC received grant funding from the National Historical Publications and Records Commission (NHPRC) for a pilot project to develop a preservation environment, including a repository, for digital records of archival value. Future grant funding may require the commitment to ensuring the preservation and accessibility of archival digital records.

- **Contractual obligations:** UAHC may contract with donors to ensure that digital material is preserved and accessible.

**Objectives**

The digital preservation function of UAHC addresses the following objectives:

- Maintain a comprehensive digital preservation program for the digital records created or collected by UAHC.
- Enable uninterrupted access to digital content over time as technology for digital content evolves.
- Collaborate with campus partners and regional and national institutions to make the best use of resources and avoid duplication of effort.
- Comply with and contribute to the development of the standards and best practices of the digital preservation community.

**Scope**

UAHC accepts responsibility for approving, ingesting, creating metadata for, preserving, and providing access to archival digital content submitted by university administrators and other depositors, as well as digitized material created internally or external to the archives, in accordance with the UAHC Collecting Policy.

**Challenges**

UAHC faces several digital preservation challenges.

- **Technological change:** UAHC, like any organization engaged in digital preservation, needs to be responsive to continually changing technology. As information technology hardware and software evolves, new digital content types, new capabilities, and new preservation challenges emerge and existing digital content faces the risk of obsolescence.
- **New digital content:** UAHC’s existing digital preservation policies, procedures, and practices may need to be revised or re-engineered to encompass new digital content.
- **Training and awareness:** All of UAHC’s staff, and some of IT Services’ Content and Collaboration staff, contribute directly and indirectly to the digital preservation function. Most staff members, however, do not have digital preservation as an explicit or significant portion of their responsibilities. UAHC is committed to providing appropriate training for and raising awareness about digital preservation issues and developments for internal staff and for the broader community of digital records producers, archivists, and users.
- **Staffing**
  - Difficulty finding a dedicated programmer. This challenge has been reduced now that UAHC is under the same leadership at IT Services Content and Collaboration.
  - Staff consistency magnified due working with different departments and teams. Staffing a bit more stable now that UAHC is under the same leadership at IT Services Content and Collaboration. IT Staff relationships are more consistent.
  - Difficulty coming to a shared understanding of goals and activities needed for digital preservation.
- **Funding:** Lack of recurring funds for digital preservation. UAHC has physical space for collections storage but no funding line to acquire, develop and support digital storage environment for electronic records.

**Operating Principles**

The UAHC digital preservation function operates in accordance with an established set of principles. UAHC will strive to:

- Comply with OAIS and other digital preservation standards and practices
- Ensure that content remains readable and understandable
- Participate in the development and adoption of digital preservation community standards, practices, and solutions
- Develop a reliable, scalable, sustainable, and auditable digital preservation repository
- Manage the hardware, software, and storage media in accordance with environmental standards, quality control specifications, and security requirements
Roles and Responsibilities

UAHC’s parent department, the Content and Collaboration Department of MSU IT Services, accepts responsibility for preserving UAHC’s digital assets. Within the Content and Collaboration Department, the Director, Assistant Director of UAHC, Electronic Records Archivists, other archival staff, and other Content and Collaboration assistant directors and staff all contribute to the management of the digital preservation function and the lifecycle of digital content at UAHC. MSU departments and other donors have the responsibility of adhering to submission guidelines. Currently the director of Content and Collaboration evaluates high-level policy documents and reviews programmatic plans and progress. In addition, the director is proposing that the campus-wide University Collections Committee and Information Technology Council assume responsibility for digital preservation of the university’s collections and assets.

Refer to Roles and Responsibilities for more detailed information.

Cooperation and Collaboration

UAHC acknowledges digital preservation as a shared community responsibility, and as such has long-standing and emerging relationships with similarly committed units at Michigan State University, in the Committee on Institutional Cooperation (CIC), and other archives, libraries, and museums worldwide.

Selection and Acquisition for Preservation

The UAHC Collecting Policy sets forth the priorities and criteria for acquiring archival material, including digital content. Selection and Acquisition Policies provides general information on content selection. Transfer Requirements and Deposit Guidelines provides submission information. A digital transmittal form including details about the submitted content must be provided with the content. In some cases, submission information specific to individual units may be articulated in submission agreements tailored to that unit’s digital content and preservation expectations.

Access and Use

The designated community of UAHC’s digital preservation program, as described by OAIS, includes MSU administrators, faculty, students, and alumni, as well as non-MSU donors and members of the public conducting archival research. Most of the digital records held in the repository are freely available, although embargoes and other restrictions may apply to some material. Information protected under the Federal Family Educational Rights and Privacy Act (FERPA) or the Health Insurance Portability and Accountability Act (HIPAA) cannot be made available in the repository.

Michigan State University holds copyright on most of the content in the repository. Researchers who wish to cite digital records from the repository must follow UAHC’s citation guidelines.

UAHC will provide access to digital records through a public website. Refer to Access and Use Policies for more information.

Digital Preservation Policies and Procedures

UAHC adheres to the following policies, strategies, and procedures in regards to the MSU digital preservation program:

Roles and Responsibilities

In handling MSU’s archival digital material, main roles and responsibilities are divided between the collection owner (Michigan State University department or other donor) that is submitting materials to UAHC.

Responsibilities of collection owners

1. Intellectual property rights: Most submitted material will be university records, and thus the property of Michigan State University. UAHC will make preservation copies of the material as needed to long-term access and availability. If material is submitted by a private individual for the historical collections, the collection owner must ensure all proper permissions associated with the deposited content are fully established. This includes the content’s subsequent preservation treatment, such as copying.

2. Transmittal form: A signed transmittal form must be submitted with digital records being transferred to UAHC. This form will provide details on materials being deposited and technical information about the systems and technologies used to create them. In some cases, a formal submission agreement may be made between the depositor and UAHC. Refer to Transfer and Deposit Guidelines for more information.

3. File transfer: Electronic files must be transferred per the procedures referenced in Transfer and Deposit Guidelines.
Responsibilities of UAHC

1. **Stewardship**: Manage digital objects throughout all phases of the objects’ life cycle, including:
   a. **Assessment/appraisal phase**: UAHC appraises the digital content to determine its suitability for deposit in the digital repository.
   b. **Deposit phase**: Digital objects are validated on submission.
   c. **Ingest phase**: The ingest workflow includes the creation of descriptive metadata; automated creation of administrative, technical, and preservation metadata; and possible migration to an acceptable archival format.
   d. **Archive and preservation phase**: Archival storage system will perform regular integrity checks to ensure the usability of digital objects over time. Digital objects may be migrated to new formats over time.
   e. **Access phase**: Access copies of the digital objects will be created and made available to researchers and the public. As acceptable access formats change over time, new access copies may be created in the most desirable formats.

2. **Reliability**: Provide services as agreed to in all agreements with collection owners and as articulated in these polices.

3. **Documentation**: Maintain current documentation of supported formats and the preservation action plan for each supported format. (See Appendix B.)

4. **Preservation**: Provide preservation treatments to digital objects that are as lossless as required given UAHC’s resources and current knowledge.

5. **Repository services**: Provide secure and redundant storage, backup, data management, fixity checks, and periodic refreshment by copying data to new storage media.

6. **Access**: Provide the public with access to digital objects and appropriate descriptive metadata, with the exception of restricted material.

7. **Sustainability**: Professionally manage the digital preservation program for long-term administrative, financial, and technical viability.

Digital Assets
UAHC adheres to the following policies regarding digital assets:

**Digital Assets: Quality Creation and Benchmarking**

UAHC is committed to providing long-term archival storage and access to all deposited content by applying best practices for data management and digital preservation while acknowledging the complexities involved in preserving digital information. UAHC commits to preserving content in the form it is originally deposited if that form is in an acceptable format. (See Appendix B.) UAHC will preserve the content, structure, and functionality of the files through normalization/migration to newer formats or other preservation strategies, where feasible. UAHC will provide basic services, including secure storage, backup, management and fixity checks.

Currently, UAHC can provide full preservation support for specific file formats only. (See Appendix B.) Evaluation criteria include prevalence of format, availability of tools for migration, and availability of local resources to take specific preservation actions. UAHC’s ability to provide full preservation support for more formats is likely to grow over time as more tools and techniques are developed and staff/resources are allocated to digital preservation support.

Full preservation support is currently provided only for formats that are both publicly documented and widely used, providing UAHC with a high degree of confidence in our preservation commitment, making it more likely that tools will exist or be developed to undertake preservation actions, and that those actions will result in an understood and controlled migration. Content may also be normalized—that is, transformed to another stable format—to provide additional assurance that functionality is preserved. If possible, all content will be preserved as originally deposited to ensure availability of the original bitstream.

**Digital Assets: Selection and Acquisition Policies**

The mission of UAHC is to collect and preserve the official records of Michigan State University as well as historical collections that support faculty and student research and classroom instruction. Many of these records are now created in digital form, and the mission extends to include these born-digital electronic records and in some cases digitized copies of paper, audio, and video originals.

Records that fall under the collection policy include:

- Official correspondence, including e-mail
- Annual reports
- Policy and procedure statements
- Speeches and presentations
- Meeting minutes and agendas
- Course syllabi and curricula
- Photographs
- MSU publications
- Recorded performances and events

For detailed information on selection and acquisition procedures, refer to Selection and Acquisition.

**Digital Assets: Transfer Requirements and Deposit Guidelines**

In order to transfer digital objects, a submission agreement must be established between UAHC and the transferring department or depositor. A transmittal form may serve this function, or a more formal submission agreement may be put in place.

Acceptable methods of transferring electronic materials to UAHC include both electronic and storage device delivery.
For detailed information on current transfer and deposit procedures, refer to Transfer and Deposit.

Digital Assets: Access and Use Policies

Most of the digital records held by UAHC are freely available for use. Embargoes and other access restrictions may apply to some materials.

Michigan State University holds copyright on most of the content in the repository. Researchers who wish to cite digital records from the repository must follow UAHC's citation guidelines.

Student education records are protected under the Federal Family Educational Rights and Privacy Act (FERPA), and personal medical information is protected under the Health Insurance Portability and Accountability Act (HIPAA). Information protected under FERPA or HIPAA cannot be made available in the repository.

Historical course and student data from Office of the Registrar databases may be accessed at http://spartanarchive.msu.edu.

A public access portal to other digital content deposited in the MSU trusted digital repository is in development. Until the website is available, researchers may request information that may be found in the archived files by contacting UAHC 1-517-355-2330 or archives@msu.edu. UAHC staff will provide copies of the requested files that are available for use.

Digital Preservation Strategies

UAHC employs the following long-term digital preservation strategies for content in the MSU Trusted Digital Repository:

- Widely-used preservation formats
- Secure storage
- Replication to at least one other storage area
- Regular fixity checks
- Technology upgrades as needed
- Migration and other preservation actions as needed

Widely-used preservation formats

To ensure long-term preservation of archived digital content in the MSU Trusted Digital Repository, UAHC is committed to using well-documented, widely-used preservation formats. Refer to Appendix C for a current list of preservation formats.

Secure Storage

Access to storage is controlled through account permissions. Physical and network security practices are put in place for the storage devices. Preservation copies of digital objects are sent to Archival Storage and access copies are sent to Access Storage. Preservation copies may be viewed only by staff with authentication privileges. Permissions do not allow for edits on the preservation copies by privileged staff. Access copies will be available to view and download once the access systems are put in place for records in Access Storage.

Redundant Storage

MSU TDR digital content in Archival Storage is replicated periodically to a Dark Archive with the same technology configuration. UAHC is planning to implement either cloud-based or distributed storage as a third, offsite storage alternative.

Regular fixity checks

UAHC is committed to ensuring the integrity and authenticity of digital content in the MSU Trusted Digital Repository through the active and ongoing use of cryptographic hash functions to verify that the content has not been altered or corrupted. Unique checksums are assigned to digital files before submission or during ingest, and fixity checks to ensure the checksums haven't changed are performed regularly on stored digital content.

Technology upgrades

Technology systems will be monitored. Software and hardware will be kept secure with regular upgrades and patches. Technology watches and regular reviews of on repository software and processing tools will be conducted in order to stay current with developments in the field of archival technologies. Archivists will work with technology professional to keep technology systems current.

Migrations and other preservation actions as needed

The archivists will monitor recommended file formats for preservation and access. If a recommended format changes, content in that format will be migrated to the new recommended format.

Digital Archive Operations

Technologies, procedures, and processes used in the MSU Trusted Digital Repository will likely change as improved technologies become
available. Changes may be made to the Repository without the consent of depositors, but must maintain the current level of service.

Currently, archival operations follow different processing methodologies depending on the content.

- Spartan Archive Repository Procedures and Processes— for databases and PDFs from the MSU Office of the Registrar
- Trusted Digital Repository Procedures— for most non-database digital content

**Spartan Archive Repository Procedures and Processes**

In late 2009, the Michigan State University Archives & Historical Collections (UAHC) received a grant from the National Historical Publications and Records Commission (NHPRC) to build and test an archival solution for the long-term access and preservation of four large, born-digital records series in database format produced by MSU’s Office of the Registrar (RO).

The record series for the proof-of-concept “Spartan Archive” project included Academic Programs, Course Descriptions, Schedule of Courses, and the Student Directory. Historically, these record series had been printed periodically as serial publications; all four are now only available digitally.

Procedures and automated processes for processing and providing access to these RO records are as follows:

**Appraisal**

Interviews, surveys, and examination of paper records

Spartan Archive project staff first interviewed RO staff to learn about the business processes for creating the databases. The data for each record series originates in databases created using MSU’s proprietary Student Information System (SIS) database management system. As they are relational databases, they can be updated at any time; however, after certain points in the semester they are fairly set. Academic Programs and Course Descriptions are also posted to the RO’s website as PDFs during the summer, after the end of each school year.

To determine the information from the databases that would be of interest to researchers, informal surveys were conducted with researchers, UAHC reference staff, and MSU Libraries reference staff. The published data in the original paper serial publications and current online versions on the RO website were also examined.

Database field review

With RO staff, UAHC’s Spartan Archive project staff reviewed the available data fields for each database. Many of the data fields are solely for internal RO use, and some student information, such as student ID numbers, is protected under the Family Educational Rights and Privacy Act (FERPA). The fields considered to be of interest and legally available to researchers were noted.

Data dictionaries

Data dictionaries for each record series were created based on the data fields chosen during the appraisal process.

- Academic Programs Data Dictionary
- Course Descriptions Data Dictionary
- Schedule of Courses Data Dictionary
- Student Directory Data Dictionary

**Database Preparation**

Data held in databases may be converted to XML, and XML markup is in text format. As plain text is a preservation format, the Spartan Archive project team chose to extract the desired data from the databases to create XML files for preservation.

XML Schema

Based on the data fields chosen in the appraisal process and the resulting data dictionaries, XML schemas for extracting the required data from the RO databases were developed.

- Academic Programs XML Schema
- Course Descriptions XML Schema
- Schedule of Courses XML Schema
- Student Directory XML Schema

Creation of XML and checksum files

At the end of each semester, the RO runs a Visual Basic executable program for each desired record series. The executable gets information from the relevant database using standard Structured Query Language, and writes it to a text file following the corresponding XML schema. A checksum file for each XML file is then created. Checksums are used to verifying the fixity of files, that is, that they have not been altered or corrupted.
The XML file plus its checksum file comprise a Submission Information Package (SIP) as defined by the OAIS model.

XML files for the record series are created at the end of each semester as follows:

<table>
<thead>
<tr>
<th>Semester</th>
<th>Record Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring</td>
<td>Course Descriptions, Schedule of Courses, Student Directory</td>
</tr>
<tr>
<td>Summer</td>
<td>Academic Programs (past academic year), Course Descriptions, Schedule of Courses</td>
</tr>
<tr>
<td>Fall</td>
<td>Course Descriptions, Schedule of Courses, Student Directory</td>
</tr>
</tbody>
</table>

Instructions for creating the XML and checksum files are as follows:

**Academic Programs**
- Go to S:\Share-RO\Information Services\DataRequestJobs\Archive\AcademicPrograms\
- Open the VB project and change catalog year to the correct year
- Run the program - will sometimes run faster if you direct to your own hard drive location
- Find the file and run the checksum for the file (see checksum info below)

**Course Descriptions**
- Go to S:\Share-RO\Information Services\DataRequestJobs\Archive\CourseDescriptions
- Open the VB project and change Term Code to the correct term code
- Run the program - will sometimes run faster if you direct to your own hard drive location
- Find the file and run the checksum for the file (see checksum info below)

**Schedule of Courses**
- Go to S:\Share-RO\Information Services\DataRequestJobs\Archive\SOC_archives
- Open the VB project and change Term Code to the correct term code
- Run the program - will sometimes run faster if you direct to your own hard drive location
- Find the file and run the checksum for the file (see checksum info below)

**Student Directory**
- Go to S:\Share-RO\Information Services\DataRequestJobs\Archive\Student_Directory
- Open the VB project and change Term_Code to the correct term code
- Run the program - will sometimes run faster if you direct to your own hard drive location
- Find the file and run the checksum for the file (see checksum info below)

**Checksum Instructions**
- Copy files to a different, easy location, such as c:\users\yourid\documents\archive if they aren’t already there.
- Make sure that a copy of fciv.exe is in the folder. Then, in a cmd prompt, type “fciv term.xml > term.xml.sum”

**Descriptive Metadata**

Each semester, or academic year in the case of Academic Programs, will have the following metadata associated with it. This metadata is either extracted from the file or automatically generated during the ingest process. The metadata is included in the Fedora record for that period, and select metadata is included in the public access interface.

<table>
<thead>
<tr>
<th>Element</th>
<th>Qualifier</th>
<th>Scheme</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>collectionname</td>
<td></td>
<td></td>
<td>Academic Programs, 2012-2013</td>
</tr>
<tr>
<td>creator</td>
<td></td>
<td>NACO</td>
<td>Michigan State University, Office of the Registrar</td>
</tr>
<tr>
<td>publisher</td>
<td>NACO</td>
<td>Michigan State University</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td>subject</td>
<td>academic programs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>description</td>
<td>Annual directory of descriptions of academic programs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>date (issued)</td>
<td>W3CDTF</td>
<td>2013-08-26</td>
<td></td>
</tr>
<tr>
<td>date (created)</td>
<td>W3CDTF</td>
<td>2013-08-26</td>
<td></td>
</tr>
<tr>
<td>date (acquisitioned)</td>
<td>W3CDTF</td>
<td>2014-01-20</td>
<td></td>
</tr>
<tr>
<td>date (accessioned)</td>
<td>W3CDTF</td>
<td>2014-01-20</td>
<td></td>
</tr>
<tr>
<td>language</td>
<td>iso639-3</td>
<td>eng</td>
<td></td>
</tr>
<tr>
<td>type</td>
<td>DCMIType</td>
<td>dataset</td>
<td></td>
</tr>
<tr>
<td>format</td>
<td>IMT</td>
<td>text/xml</td>
<td></td>
</tr>
<tr>
<td>format (extent)</td>
<td></td>
<td>13955672</td>
<td></td>
</tr>
<tr>
<td>coverage (spatial)</td>
<td>LCNAF</td>
<td>East Lansing (Mich.)</td>
<td></td>
</tr>
<tr>
<td>coverage (temporal)</td>
<td></td>
<td>2012-2013</td>
<td></td>
</tr>
<tr>
<td>rights</td>
<td></td>
<td>Use of this public domain resource is unrestricted.</td>
<td></td>
</tr>
<tr>
<td>relation (is part of)</td>
<td></td>
<td>msu-uahc:UA6.7-AcademicPrograms</td>
<td></td>
</tr>
</tbody>
</table>

### Transfer/Submission

After the RO creates XML and checksum files for the scheduled record series, the files must be deposited in an authenticated area in MSU's NetFiles/AFS dropbox.

Transfer steps are as follows:

- Log in to [NetFiles/AFS](https://netfiles.msu.edu) with Net ID and password.
- Navigate to [https://netfiles.msu.edu/?path=/afs/\unit/spartanarchive/dropoff/RegistrarsOffice](https://netfiles.msu.edu/?path=/afs/\unit/spartanarchive/dropoff/RegistrarsOffice)
- Upload files.
- The system validates XML and verifies checksum.
- An e-mail notification is automatically sent to the archival administrator alerting them that the files have been transferred.
- As a courtesy, the RO also sends the archival administrator an e-mail to let them know that files have been uploaded.

### Ingesting a Records Collection: Procedures

The page on Ingesting a Records Collection from the Technical Documentation is included here.

When an office drops off a new submission in its dropoff area, the Spartan Archive administrator will receive an email notification that the file has been deposited and moved to quarantine. To begin the ingest process, the administrator should follow these steps:

1. Open the [Spartan Archive Current Submission console page](https://netfiles.msu.edu/?path=/afs/\unit/spartanarchive/dropoff/RegistrarsOffice) (authorized login required). The new files will have a status of In Quarantine.
2. To approve the ingest, click on the line with the new file. This will enable the Actions for Submission box on the right.
3. Open Archivists’ Toolkit. To find the next available Accession Number, double click on the word Accessions on the left menu. (See below). When the list of accessions appears, type A.YYYY where YYYY is the current year in the filter field at the top right. This will bring up a list of all of the accessions for the year. The new accession will be A.YYYY.XXXX where XXXX is the next available sequential number. In the example below, the new accession number would be A.2012.00117. **Do not create a new accession at this time.**

4. On the Current Submissions page in the Actions for Submission box, click the Accept for Ingest button.
5. Enter the new accession number in the field provided.
6. Click the Accept button.
7. The status for the file should change to Approved for Ingest. (See below.)
8. In the Actions for Submission box, left click (command click on a Macintosh) the AT Import File link to download the atimport.xml file to your system.
9. Save the file to a desired location.

10. In Archivists’ Toolkit, click on Import/Import Accessions (XML) menu item (at the top left).
11. Navigate to the directory in which you saved the atimport.xml file.
12. Select the file and click the Import button. A dialog box will appear to confirm the import.
   a. If there are no errors, you are done.
   b. If there is an error message stating that this is a duplicate accession, the accession number may be in use. Get a new, available number and repeat the approval process. As long as this is done on the same day, there should be no problems.
   c. If there is any other error, investigate and correct any problems found.
   d. As a last resort, the atimport.xml file may be printed and the accession manually created.

13. If you do not want to approve the file, select either Pend or Reject in the Actions for Submission box. Pending serves to remind you to investigate.
14. If the status of the file in Current Submissions is "In FTP," there was a problem with it. Click on the file and note the events listed in the box below. They should show any problems. Technologists supporting SpartanArchive can follow the relevant procedures in the Potential Problems document to resolve any issues.
15. The ingest process should automatically complete overnight, showing a status of "Added to Database, IngestComplete" in the submission console. If the status has not updated or if it shows as anything other than "Added to Database, IngestComplete," contact the programmer. Also, the administrative archivist will receive notifications about the success or failure of the ingest.

16. On verification of a successful ingest:
   a. Open the new accession record in Archivists' Toolkit.
   b. Click on the "Acknowledgements, Restrictions & Processing" tab.
   c. Under Processing Status, check the "Processed" box and add the date of the ingest. Save.
   d. Open the Resources record for the accession's Record Group.
   e. If an Electronic Resources series and a file for this type of content already exists, click on the file and update the end date if necessary. Save.
   f. Add an Electronic Resources series and/or file for the content type if necessary, following the instructions for creating an Electronic Records series and linking digital objects on the Post-Ingest Procedures page.

Ingesting a Records Collection: Automated Processes

Ingesting a records collection kicks off several automated processes integral in the creation of preservation and access copies of the submitted XML file.

- **Metadata extraction**
  Descriptive metadata is extracted and generated from the XML file.

- **Creation of Fedora object (record)**
  A new Fedora object (record) containing the metadata and referring to the file is created. Together, the metadata stored in Fedora and the XML file comprise the Archival Information Package (AIP), as defined in OAIS.

- **Preservation copy sent to Archival Storage**
  The XML file is copied to its final location in Archival Storage.

- **Database load (for access)**
  Data is loaded into the Spartan Archive database for access. When delivered in response to a researcher request, data from the database is the Dissemination Information Package (DIP), as defined by OAIS.

For detailed technical information on the preservation functionality of the Spartan Archive, refer to the Spartan Archive technical documentation and procedures on the public Spartan Archive project deliverables page.

Ingesting PDFs: Procedures

The page on Ingesting PDFs from the Technical Documentation is included here.

The administrator should follow these steps to ingest PDFs into Fedora:

1. Go to: http://spartanarchive.msu.edu/PDFPuller/ (authorized login required).
2. Click on the button for the content to be ingested, such as Academic Programs or Course Descriptions. (See below.)
3. Enter the specific name of the link on the website from which the PDF is being pulled.
   a. For Academic Programs, the name of the link is in the format 20xx-20xy, such as 2012-2013.
      (See http://www.reg.msu.edu/UC/PYearIndex.asp.)
   b. For Course Descriptions, the name of the link is in the format 20xx-xy, such as 2012-13.
      (See http://www.reg.msu.edu/UC/DescYearIndex.asp.)
4. Click the Submit button.

Which collection are you pulling for?

- [ ] Academic Programs
- [ ] Course Descriptions

What period are you pulling for? [ ]

Click on the button below to download the PDF files to records/UA.6.7-RegistrarsOffice/

Submit

5. The next page will display the PDF/group of PDFs selected on the previous page.
6. Identify an Accession Number to be used for the ingest, following the method described in Ingesting a Records Collection, step 3.
7. Enter the Accession Number in the box provided. (See below.)
8. Click the Submit button. After a few moments, a page will display indicating a successful ingest. The PDF will be ingested into the Fedora repository and a link to it will be available in the public access interface.

Verify that this is the group of PDFs you wanted. If so enter the PID of the Fedora object that you want to add these to and click on submit.

Submit

9. Create an accession record for the ingest in Archivists' Toolkit.
   a. In the left menu, click on "Accessions."
   b. On the first page, fill in the Accession Number, accession date, title, extent, and date range fields. Link to the appropriate Resource (Record Group). Save.
   c. Go to the Accession Notes tab. Fill in the Retention Rules (Permanent) and Description fields. Save.
   d. Go to the User Defined Fields tab. Fill in the Acquisition Date field with the date of the ingest. Save.
   e. Go to the Acknowledgements, Restrictions & Processing tab. In Processing Status, click the "Processed" box. Fill in the Processed Date field with the date of the ingest. Save.

PDFs are stored on enterprise storage and backed up with an enterprise backup system. The future plan is to move the PDFs to Archival Storage as AIPS and are also available for access as DIPs.

Preservation
Database Records

An XML file submitted by the RO is considered a preservation copy of the database for that semester or academic year, in the case of Academic Programs. XML markup is text-based, and text is a preservation format.

The XML files will be stored and preserved in Archival Storage. Currently, they reside on a system with daily enterprise backups. Metadata extracted and generated from each XML file during ingest is stored as a record in the Fedora repository. Fedora maintains referential integrity between the XML files and their repository records.

PDFs

Fedora records are created for the ingested PDFs of Academic Programs and Course Descriptions. The PDFs will be stored and preserved in Archival Storage. Currently, they reside on a system with daily enterprise backups.

For detailed technical information on the preservation functionality of the Spartan Archive, refer to the Spartan Archive technical documentation and procedures on the public Spartan Archive project deliverables page.

Access

The public access website for the Spartan Archive Repository is located at http://spartanarchive.msu.edu.

Academic Programs

To access Academic Program information, researchers may click on the academic year and format desired (database or PDF). For database access, they may search on specific terms or browse through a detailed table of contents. The interface for the table of contents maintains the original structure of the document. (See below.)

Course Descriptions

To access Course Descriptions information, researchers may click on the semester and format desired (database or PDF). They also have the option to search on the contents of the entire database of all semesters since 1970.

The database may be searched on a number of terms, including subject, course, department, and college. Multiple terms may be used in a single search.

Returned search results will include data from the search terms above as well as interdepartmental information; credit information; semesters offered, grading option, restrictions, and recommended background information; and prerequisite, corequisite, and associated courses.

Search results may be downloaded as CSV files for import into a desired spreadsheet or other format.

Course Schedules
To access Course Schedules information, researchers may click on the semester desired. They also have the option to search on the contents of the entire database of all semesters since Fall 2013.

The database may be searched on a number of terms, including subject, college, course title, class meeting days, classroom locations, instructors, and credit hours. Multiple terms may be used in a single search.

Search results may be downloaded as CSV files for import into a desired spreadsheet or other format.

Student Directory

To access Student Directory information, researchers may click on the semester desired. They also have the option to search on the contents of the entire database of all semesters since Fall 1994.

The database may be searched on a number of terms, including class levels (undergraduate or graduate), majors, classes, and local and permanent addresses. Multiple terms may be used in a single search.

Search results may be downloaded as CSV files for import into a desired spreadsheet or other format.

**Trusted Digital Repository Procedures**

Procedures for acquiring, transferring, processing, ingesting, and providing access to non-database, non-website records are detailed in the following:

For information on database processing and archiving procedures, refer to [Spartan Archive Repository Procedures and Processes](#).

Select MSU websites are archived using the Internet Archive's [Archive-It](#) service.

**Selection and Acquisition**

**General Digital Records**

If an MSU department or unit has not previously sent digital content to UAHC, or if there are questions about whether the content has historical value, the unit should contact UAHC at 517-355-2330 or archives@msu.edu and ask to be put in touch with an electronic records archivist. The archivist will conduct a review of the content that may include an onsite visit to the department. Content of historical value will be recommended for transfer to UAHC. For deposits of similar content in the future, the archivist will advise on how to select content of historical value for transfer.

If the unit has deposited similar content in the past, content should be transferred to UAHC based on appraisal guidelines provided by the archivist.

Refer to [Transfer and Deposit Guidelines](#) for instructions on how to complete a transfer.

**Database Records**

Database records present a special case in digital content selection. The archivist will need to examine the databases to determine fields of archival interest and to review the forms of the actual data in each database under consideration. UAHC technical staff will create XML schema for each database and work with your department to convert each databases to XML files that correspond to the schema. Details for these procedures may be found in [Spartan Archive repository Procedures and Processes](#).

**Transfer and Deposit**

**Transmittal Form**

In order to transfer digital material, a transmittal form must be completed and signed by the department chair or unit head.

This form serves as a record of what has been transferred to UAHC and is useful when retrieving records. Information provided on the form will be used to create descriptive metadata, which is information about digital objects that will make them more accessible to researchers. Technical information provided will help the archivists in determining preservation strategies.

Multiple record series must NOT be listed together on one transmittal/inventory form. A separate transmittal/inventory form must be filled out for each series.

Any records with student personal information, social security numbers, medical histories, or other personal information should be marked as Restricted on the transmittal/inventory form.

On completion, the transmittal form may be provided in hard copy or e-mailed to UAHC at archives@msu.edu. The office should keep one copy for reference. Upon review by UAHC staff, an accession (tracking) number will be provided. This number should be written on the office's copy of the transmittal/inventory form.

**Transfer Options**

Acceptable methods of transferring electronic materials to UAHC include both electronic and storage device delivery.
**Electronic Delivery**

Contact UAHC at archives@msu.edu or 1-517-355-2330 to determine the optimal means of electronic transfer.

The files to be transferred should include:

- Digital objects
- Checksums of digital objects, for use in verifying integrity (optional)

**Storage Device Delivery**

Electronic records stored on removable media and other physical devices may be delivered to UAHC on those physical devices. If the total volume of the transfer exceeds 100GB, delivery on a removable hard drive is recommended. Files should be packaged using a lossless compression algorithm before copying to a removable hard drive or other media to ensure authenticity.

Acceptable physical devices for the delivery of electronic records include:

- USB flash drive
- Hard drive
- Memory card
- CD or CD-ROM
- DVD or DVD-R/W
- Floppy disk
- Zip disk
- Jaz disk

Include a hard copy of the transmittal form with the package containing the physical device.

Deliver to:

University Archives & Historical Collections
888 Wilson Road, Room 101
Conrad Hall
Michigan State University
East Lansing, MI 48824

UAHC may provide pickup service for large volumes of digital records. Contact the UAHC office at archives@msu.edu or 1-517-355-2330 for more information.

**Pre-Ingest Processing: TDR Documentation**

Please see the pre-ingest manual at https://spartanarchive.files.wordpress.com/2011/04/pre-ingest-processing.pdf.

**Ingest (Archivematica): TDR Documentation**

The following is a guide through the steps to ingest files using Archivematica 0.10 via the web-based dashboard. Follow these steps after Pre-Ingest procedures.

- Dashboard
- Transfer
- Metadata Input
- Ingest
- Checking

**Dashboard**

1. Go to armatica.itservices.msu.edu.
2. Enter your username and password.
   a. This will bring up the Archivematica Dashboard.

**Transfer**
1. In the box labeled “Transfer name,” enter the name of the transfer. This is usually the accession number, but will include the _XXX if there are multiple parts of the accession. The accession number should be entered in the form “UA.xxxxx” or “A.20xx.xxxx”.
2. In the box labeled “Accession no.,” enter the accession number again. This will not include the _XXX for multiple accession parts.
3. Click on the Browse button.
4. This brings up a popup window. Click on the Archivematica_ingest folder.
5. Click “Add” next to the folder labeled with the accession number for the accession to be ingested. If it is part of a larger accession, click the arrow next to the accession number’s folder, then choose the specific _XXX folder and click the “Add” next to that folder. The popup window will close. On the Dashboard, the string /home/archives/accessioned/Archivematica_ingest/ACCESSION_NUMBER will have appeared. Everything in this folder will be part of the transfer/ingest.
6. Click the green Start Transfer button.
7. After a few seconds, a bell icon and the name of the transfer will appear, indicating that an action is required. From the Action drop down menu on the right, select “Approve transfer.” Archivematica will run through a series of micro-services. This could take several minutes.
8. The bell icon will appear when the next action, creating a SIP from the transfer, must be taken. From the Action drop down, select “Create single SIP and continue processing.” This will end the Transfer phase of the accession.
On the top menu bar, click on “Ingest.”

The bell icon indicates another action to be taken, Normalization.

During the Transfer phase of ingests with disc images, the “Characterize and extract metadata” micro-service will indicate failure. This seems to be peculiar to running an ingest containing a disc image; presumably, one or more of the FITS tools does not recognize the .iso file format. Processing will continue, however, so ignore the failure.

**Metadata Input**

1. On the top menu bar, click on “Ingest.”
2. The bell icon indicates another action to be taken, Normalization.
3. Before taking action to normalize, descriptive metadata must be added to the submission. IMPORTANT: THIS IS THE ONLY POINT AT WHICH DESCRIPTIVE METADATA MAY BE ADDED TO THE ACCESSION. To get to the metadata entry form, click on the text document icon above the Action drop down.
4. This brings up a Submission Information Package page. Click the “Add” link under Metadata.

5. Enter descriptive metadata for the submission, following the “Descriptive Metadata for Use in Archivematica Processing” guidelines. Note that the Rights Statement that should be used for most content is available in a Microsoft Word file on the desktop of the E-Records Processing Workstation. This statement may be copied into the metadata form.
6. IMPORTANT: Click on Create button at the bottom of the page to save metadata.
7. If you wish to add or change metadata, click on the Edit link to return to the metadata form page. Click the Save button at the bottom of the form to save your changes.

Ingest

1. Return to the ingest process by clicking on “Ingest” on the top menu bar.
2. From the Action drop down menu, select “Normalize for preservation and access.”
3. When the bell icon appears, you have the option of reviewing the normalization. To do so, click on the “review” link. You may see the DIPs that have been created by clicking through to the preservationAndAccess folder and the folder within that.
4. Go back to Ingest.
5. From the Action drop down menu, select “Approve” to approve the normalization.
6. After a few minutes, the bell icon will appear. It will request action for “Upload DIP”. Do not take any action at this decision point.
7. When the second bell appears, you will now have the opportunity to review the AIP. Click on the “review” link and the storeAIP folder. The folder containing the AIP will be present under the name of the accession plus a checksum number. You may review the METS file and the contents of the AIP. Note that the AIP will likely need to be unpacked using 7zip.
8. Go back to Ingest.
9. Select “Store AIP” from the drop down menu.
10. The accession is complete when a green circle with a checkmark appears in the upper left hand corner of the processing flow list.

Checking

To ensure that the accession has gone through:

1. Check the AIP folder: In Archivematica, click on the Archival Storage tab. Click on folder of recent accession. When directed, open as
a 7zip file. Double-click on the folder in 7zip and then on the “Data” folder. Ingested digital objects may be found in the “Objects” folder.

a. Note that TIFFs, JPEGs, and METS files stored in the AIPsStore may not open directly from the 7zip interface. They may be extracted from 7zip or simply dragged to the desktop for viewing.

b. Alternatively, you may also view the AIPs in the AIPs directory on the Samba server (\samba.spartanarchive.msu.edu\aips). You can identify the correct folder by the date and time under Date modified. Even then, you will have to click through a number of folders to get to the accession, a 7zip file that will need to be unpacked. Within you will find the METS file, the objects, logs, and submission documentation.

2. Check DIP folder: Locate and double click on the DIPs directory on the Samba server (\samba.spartanarchive.msu.edu\dips). Double-click on folder of recent accession. Access copies of ingested digital objects may be found in the “Objects” folder.

### Processing Disc Image Accessions in Archivematica - Checking the AIP and DIP

Once the AIP has been stored, check on the AIP and DIP that have been generated. The DIP object should only contain the mp4 file. The AIP object should contain the disc image (.iso file), the .mp4 or .m4v file, and possibly an .mkv file (preservation copy as result of normalization of .mp4).

---

**Post-Ingest: TDR Documentation**

The following steps should be completed after ingest of an accession has been verified.

- **Quality Check**
- **Archivist Toolkit**
- **Final Steps**

#### Quality Check

1. Check the AIP/DIP for quality
   
   a. Look at the metadata.
   
   b. Open the files/play video.
   
   c. If you encounter errors, notify the archivists.

#### Archivist Toolkit

1. Log in into Archivist Toolkit.
2. Search in “Resources” for the record group identifier.

---

**Checking AIPs/DIPs**

Both are a separate drive in the directory. DIPs can be found easily, as the accession number is clearly visible at the beginning of the folder name. The four characters after the accession number will help to locate the AIP, along with the date the ingest was completed.

![Checking AIPs/DIPs](image)

2. If the quality is acceptable, write the accession number + first four letters/numbers Archivematica assigns to the AIP/DIP on the top of the metadata sheet.
   
   a. Example: UA.05465-9e55

3. This will be the digital object identifier for now.
a. You may have to click “List All” in the top bar to generate the list of resources before being able to search.
b. If resource does not exist for this group, or if there are no series listed, let an electronic records archivist know.

3. At the Resource level, change the end date to match the end date of the resource if necessary.

4. Within the resource for this group, scroll to the bottom of the series list on the left.
   a. If an “Electronic Resources” series exists, use that series.
   b. If an “Electronic Resources” series does not exist, create one.

   a. Select the last series (the yellow folders).

   ![Electronic Resources Series]

   “Select the last series”
   In the top example, there are no other series under the resource, so click the last bullet. In examples such as the one on the bottom, be careful to select the last series under the record group folder (in this case “Annual Reports”), and not something that would be underneath that series (“Unpublished,” for example). The dotted-line branching should help to identify the correct line.

   b. Click on “add sibling” and enter title as [Electronic Resources] and select [series] from drop-down.
   c. Select the “Electronic Resources” series just created on the left.
   d. Click on “add child” and enter digital object information.

6. Enter Digital Object Information.
   a. Select a file from drop-down menu.
   b. Title is: [Name of Digital Object from metadata sheet].
      i. Do not include date.
   c. Enter dates in date fields.
   d. Click “Save” button. The title should appear as an indented line branching from the “Electronic Resources” folder icon.

   ![Digital Object Information]

   e. Instances: Click on “Add Instance.”
      i. Add “Digital Object.”
         1. Select object type (i.e. moving images).
            a. Example: UA.05464-feeb

   ![Digital Object ID]

   3. Click “OK” button.
ii. Add Instance Type, such as “Moving Image” or “Photographic Materials.”
1. Add Container Type: object.
2. Add Indicator: accession number + first 4 digits of Archivematica number/letters.

<table>
<thead>
<tr>
<th>Container 1 Type</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container 1 Indicator</td>
<td>A.2013.0201-1d38</td>
</tr>
<tr>
<td>Container Barcode</td>
<td></td>
</tr>
</tbody>
</table>

3. Click “OK” button.
4. A window will pop up, showing that containers are loading. This may take some time depending on the content.
5. Click “Save” button.

7. Click “Manage Locations” button.
   a. Select the Object ingested in the Containers box.
   b. Search for “IX” and choose IX Storage.
   c. Click “Add location link.”
   d. Click “Done” button.
   e. Click “Save” button.

8. Add Notes/Subjects.
   a. Select the Record Group folder (topmost folder).
   b. Go to “Notes, etc.” tab.
   c. Double-click on “Scope and Content” and update as necessary.
      i. The electronic resource consists of...
   d. Go to “Names & Subjects” tab.
      i. Add names or subjects as appropriate.
      ii. These should be official Library of Congress names and subject headings.
      iii. Search within AT or See the LoC Authorities website, http://authorities.loc.gov/.
   e. Go to “Finding Aid Data” tab.
      i. Add revision date in month/year format (e.g., September 2013). If applicable, replace existing date with this update.
      ii. Add brief revision description plus initials (i.e. Added electronic resource LS). If applicable, replace existing description with this update.
      iii. In the Finding Aid Note box, add a revision entry to list of entries, e.g., “Revised September 2013.
      iv. Status should be “3_Ready_for_Cataloging.”

9. Save and exit Resources tab.

10. Open Accession Record (when finished with entire accession).
   a. Click on Acknowledgements, Restrictions, & Processing Tasks tab.
      i. If this ingest is part of a larger accession that has not been processed yet, go to the “Accession Notes” tab and add a Disposition Note stating that the digital files were processed, the date processed, and the initials of the archivist.
      ii. If this ingest includes the entire accession:
         1. Change processing status to: Processed.
         2. Check “Processed” box.
         3. Change processing date to: yyyy/mm/dd.
Final Steps

When all the above steps are completed:

1. Write “A Ingest Done, [initials], [date]” on the top of the metadata sheet
2. Move accession folder from Archivematica_Ingest and Processing in Workspace to the AI and P folders, respectively, in the Processed folder on the “Ingests” removable hard drive.
   a. Workspace -> Archivematica_Ingest -> [accession] move to Ingests -> Processed -> AI
   b. Workspace -> Processing -> [accession] move to Ingests -> Processed -> P
      i. NOTE: The Processing folder should only be moved to “P” if there is no removable media associated with accession. Check the AT records and/or any applicable physical locations. In most cases, if there is removable media, these files will be deleted rather than copied over. In some cases, if the processing of the materials was especially complicated, rather than losing this time and effort, the files will be copied to the P folder regardless of removable media. Ask an electronic records archivist if it is unclear which procedure should be followed. This only applies to the Processing and P folders.
3. Check files moved to Ingests. When satisfied that the move was complete, delete the folders for that accession from the Archivematica_Ingest folder and Processing folders, if applicable.
4. Send Ed an e-mail at buschedw@msu.edu with the accession number and resource/record group number so he can update the finding aid for the resource.
Technology infrastructure requirements and technologies used in the MSU Trusted Digital Repository will most likely change as improved technology becomes available. Changes may be made without the consent of depositors, but must maintain the current level of service. Information on current technologies, as well as those technologies currently in the pipeline for future use, may be found at the following links:

**Infrastructure and Workflow Diagrams**

- Spartan Archive Repository Workflow
- MSU Trusted Digital Repository Infrastructure and Workflows
- Archivematica Processing-Ingest Workflow
- MSU Trusted Digital Repository Architecture

Aspirational plans are expressed by dotted lines in some diagrams. The Spartan Archive Repository Workflow shows the RO AIPs being stored in Archival Storage; this is not currently the case, but current storage is on an enterprise system with enterprise-level back-ups.

**Technology Platforms**

The MSU Trusted Digital Repository utilizes a number of technology platforms, including:

**Desktop Workstations**

From any desktop Windows, Macintosh, or Unix workstation, an authenticated administrator may approve and ingest files into the MSU Trusted Digital Repository through one of two web-browser based dashboards: the Spartan Archive console (database content) or the Archivematica processing console.

UAHC has set up a Windows 7 PC dedicated to electronic records processing, known as the "E-Records Workstation." This PC is loaded with pre-ingest processing tools for use before following Archivematica ingest procedures for non-database content.

**Spartan Archive Console: Documentation**

The Spartan Archives Administrative Interface program (SAControl) serves three purposes. First it is the tool that archive administrators use to manage ingest of records based submissions. Second it is used to maintain configuration and data dictionary information for Spartan Archive collections. Finally it provides a means for viewing the record of activity in the archives system.

The Administrative Interface is implemented as a web application. A single Java Servlet is used to coordinate between the services requested by the administrator and the Java classes that service the requests. This is illustrated below:
The web pages returned are a combination of pure Java Server Pages (JSP), Asynchronous Javascript and XML (AJAX) and a combination of the two. In the diagram above the JSP pages returned are shown with a .jsp extension while AJAX data is shown as XML shorthand (e.g. <events>).

The flow of the PDF puller is very simple. First the administrator request a set of PDFs for a given collection and period. A second page is returned that contains links to each of the PDFs requested so that the administrator can test them and ensure that they are the ones desired. If so the administrator can enter an Accession Number (obtained from Archivists Toolkit) and press Submit Query. This causes the PDFs to be copied to the digital vault and also added to Fedora. A final page is then displayed that shows what was done.

The core of the application is the Java Servlet PDFPuller.java. This serves as a controller for the application. Requests to PDFPuller come as extended URLs similar to the following:

http://spartanarchive.msu.edu/SAControl/ControlSource?type=submissions

All requests start with http://spartanarchive.msu.edu/SAControl/ControlSource. The specific request is controlled by the type parameter. There are currently 5 web pages associated with the application each is associated with one or more specific requests to ControlSource as follows:

<table>
<thead>
<tr>
<th>Request Type</th>
<th>Page Returned</th>
</tr>
</thead>
<tbody>
<tr>
<td>submissions</td>
<td>Spartan Archives Current Submissions</td>
</tr>
<tr>
<td>allEvents</td>
<td>Spartan Archive Event Log</td>
</tr>
<tr>
<td>submissionHistory</td>
<td>Spartan Archives Submission History</td>
</tr>
<tr>
<td>offices</td>
<td>Spartan Archive Configuration Page</td>
</tr>
<tr>
<td>getDefinitions*</td>
<td>Spartan Archives Data Dictionary</td>
</tr>
</tbody>
</table>

*definitions cover a specific collection so its URL adds &collection=collection name

Within some of the pages is a set of AJAX calls. These are summarized below:

<table>
<thead>
<tr>
<th>Page</th>
<th>Request</th>
<th>Purpose/Returns</th>
</tr>
</thead>
</table>

The specific request is controlled by the type parameter. There are currently 5 web pages associated with the application each is associated with one or more specific requests to ControlSource as follows:
| Current Submissions | type=updateStatus&id=submissionID1&action=[Accept, Reject, Pend]&accessionNumber=accessionNumber&comment=comment | Tells the system what to do with the submission. Returns an xml file with the updated status, event and download location for the Archivist Toolkit import |
| Current Submissions | type=events&id=submissionID1 | Returns an xml file with all of the events associated with this submission |
| Configuration Page | type=collections&officeID=officeName | Returns an xml file with all of the collections associated with the office |
| Configuration Page | type=parameters&collectionID=fedora pid of collection | Returns an xml file with all of the default parameters associated with the collection |
| Configuration Page | type=updateParameters&collectionID=fedora pid of collection | Update the Fedora defaults datastream for the collection using a POST action. Returns an xml data stream with the new parameter values |
| Data Definitions | type=updateDefinitions | Updates the data dictionary using a POST action (the collection is a field in the form that is sent). Returns an xml data stream with the new definitions |

1. submissionID is the fileId from the submissionstatus table in the spartanarchive database

The Java backend for the Configuration and Data Dictionary pages make use of a set of utility classes named XXXXProcessor.java. These classes use JDom to convert to and from the xml strings stored in Fedora and the Java objects and collections that are used in the main process.

Individual Pages

**Current Submissions**

This page is initially built as shown below as via JSP. SubmissionsDataSource.java (via ControlSource.java) sets the submissions attribute which is used to populate the first box.

![Spartan Archives Current Submissions](image)

When an administrator selects a submission to determine its fate the all of the associated events appear in the box beneath (via the events AJAX call listed above) and the actions box on the right is activated as shown below:
The administrator can use the information in the events list to help determine what to do with the submission. If the administrator wishes to accept the submission he or she can click on Accept for Ingest. At this point the administrator must enter an Accession Number (Note an edit to enforce this should be added) and may enter comments. Clicking on Accept keys the updateStatus AJAX call that causes SAUpdates.java to update the status record for this submission to Accepted for Ingest and adds the Accession Number to the record. The program then returns an xml file with the new status information, the new event and the location of an Archivist Toolkit import file that can be downloaded for use in updating Archivist Toolkit. (Note this should be changed to return all of the events associated with the submission since doing it this way messes up the order.) This completes the administrators interaction with the ingest process.

The database status has been updated and will key the Ingest2 program, running each night, to complete the ingest of the submission. If the administrator choses to Pend the submission the only change will be that the submission is shown as Pended in the submissions list as a reminder to compete any investigation needed. It is intended (but not yet implemented) that choosing Reject and delete will lead to the ingest process deleting the record.

Web Page used: submissions.jsp
Database/Tables used: spartanarchive/submissionstatus, saevents
Javascript files: jquery.js, submissions.js
Java classes used:

<table>
<thead>
<tr>
<th>Class</th>
<th>Called via</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>ControlSource.java</td>
<td></td>
<td>All interactions</td>
</tr>
<tr>
<td>SubmissionsDataSource.java</td>
<td>type=submissions</td>
<td>Sets a request attribute to the list of submissions</td>
</tr>
<tr>
<td>EventDataSource.java</td>
<td>type=events</td>
<td>Returns an xml string with all of the events associated with this submission</td>
</tr>
<tr>
<td>SAUpdate.java</td>
<td>type=updateStatus</td>
<td>Updates the status record in the database. Returns an xml string with the new status, event and the location of the AT import file.</td>
</tr>
<tr>
<td>AccessionData.java</td>
<td>type=updateStatus</td>
<td>Collects Accession level data to build the AT import file</td>
</tr>
<tr>
<td>DOMCreate.java</td>
<td>type=updateStatus</td>
<td>Builds the AT import file.</td>
</tr>
<tr>
<td>NameData.java</td>
<td>type=updateStatus</td>
<td>Collects name level data to build the AT import file</td>
</tr>
<tr>
<td>SubjectData.java</td>
<td>type=updateStatus</td>
<td>Collects subject level data to build the AT import file</td>
</tr>
</tbody>
</table>
Submission History

This page is built as shown below as via JSP. SubmissionsDataSource.java (via ControlSource.java) sets the submissions attribute which is used to populate the page.

![Submission History Table]

Web Page used: submissionHistory.jsp
Database/Tables used: spartanarchive/submissionstatus
Javascript files: None
Java classes used:

<table>
<thead>
<tr>
<th>Class</th>
<th>Called via</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>ControlSource.java</td>
<td>All interactions</td>
<td></td>
</tr>
<tr>
<td>SubmissionsDataSource.java</td>
<td>type=submissions</td>
<td>Sets a request attribute to the list of submissions</td>
</tr>
</tbody>
</table>

Event Log

This page is built as shown below as via JSP. EventsDataSource.java (via ControlSource.java) sets the events attribute which is used to populate the page.
Web Page used: eventDisplay.jsp
Database/Tables used: spartanarchive/saevents
Javascript files: None
Java classes used:

<table>
<thead>
<tr>
<th>Class</th>
<th>Called via</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>ControlSource.java</td>
<td>All interactions</td>
<td>Sets a request attribute to the list of events</td>
</tr>
<tr>
<td>EventsDataSource.java</td>
<td>type=allEvents</td>
<td></td>
</tr>
</tbody>
</table>

**Configuration Page**

This page is initially built as shown below as via JSP. FedoraAccess.java (via ControlSource.java) retrieves the source data stream from the base Spartan Archives Fedora object (msu-uahc:SpartanArchive), then uses CollectionProcessor to parse the xml stream into a list of office names. It then sets the offices attribute to this list which is used to populate the first box.

When an administrator selects an office this causes an AJAX call (type=collections as listed above) to be made. As a result of this ControlSource.java calls FedoraAccess.java to retrieve the collection names for the office.
FedoraAccess does this by first retrieving the source xml data stream from the Spartan Archives base object (msu-uahc:SpartanArchive). This is passed to CollectionsProcessor.java which extracts the collection information into a List of CollectionFragments objects. (A CollectionFragment is a small subset of the collection information needed for this purpose.)

The xml string retrieved is used to populate the Collections box as shown below:

```
Spartan Archive
Configuration Page

Offices:
Office of the Registrar

Collections:
- Academic Programs
- Course Descriptions
- Course Schedules
- Student Directory

Parameters
Update Parameters
```

When the administrator chooses a collection this sets off a new AJAX call (type=parameters) which causes ControlSource.java to call FedoraAccess.java. FedoraAccess retrieves the defaults xml data stream from the collection defaults Fedora object (these are identified by a pid of the form msu-uahc:DDCollectionName). FedoraAccess uses DefaultsProcessor.java (in the FedoraDefaults library) to parse the file and return a list of the parameters. This returns a list of Parameters (also from FedoraDefaults. Parameters are composed of an entry and one or more values implemented as a Parameter class.) which is returned to the page as an xml string and used to populate the bottom box.

```
Spartan Archive
Configuration Page

Offices:
Office of the Registrar

Collections:
- Academic Programs
- Course Descriptions
- Course Schedules
- Student Directory

Parameters
Update Parameters

<table>
<thead>
<tr>
<th>Entry Name</th>
<th>Value</th>
<th>Scheme</th>
<th>Qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submission Name</td>
<td>publicCourses.xml</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collection Name</td>
<td>Description of Courses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collection Short Name</td>
<td>CourseDescriptions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>Annual Listing of course descriptions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Record Group</td>
<td>UA.6.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Record Group Description</td>
<td>Office of the Registrar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Record Group Short Name</td>
<td>RegistrarsOffice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publisher</td>
<td>Michigan State University</td>
<td>NACO</td>
<td></td>
</tr>
</tbody>
</table>
```

Add Office/Collection
New Office
New Collection
Remove Collection
Remove Office
Data Dictionary
At the same time the Update Parameters and Data Dictionary buttons are enabled. Because parameters can have multiple values clicking on one of the + signs will cause a new line to be added with the same entry name. Many parts of the Spartan Archive depend on the values of the entry name, so these cannot be changed or added. Deleting the contents of the Value, Scheme an Qualifier fields will effectively delete a parameter. (Note: an edit needs to be added that will require those critical parameters to be entered.)

Clicking on Update Parameters causes an AJAX call (type=updateParameter) to be issued. ControlSource calls FedoraAccess which calls DefaultsProcessor.java to convert the Parameters list to an xml string which is then used to modify the defaults data stream.

Clicking on the Data Dictionary button brings up the Data Dictionary page for this collection.

Web Page used: SAConfig.jsp
Database/Tables used: None
Javascript files: jquery.js
jquery.form.js
parameters.js
popup.js
Java classes used:

<table>
<thead>
<tr>
<th>Class</th>
<th>Called via</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>ControlSource.java</td>
<td>All interactions</td>
<td></td>
</tr>
<tr>
<td>FedoraAccess.java</td>
<td>type=offices</td>
<td>Retrieves list of office names from msu-uahc:SpartanArchive, assigns to office attribute</td>
</tr>
<tr>
<td></td>
<td>type=collections</td>
<td>Retrieves list of CollectionFragments from msu-uahc:SpartanArchive, returns xml string</td>
</tr>
<tr>
<td></td>
<td>type=parameters</td>
<td>Retrieves parameters from appropriate collections defaults object defaults data stream, returns xml string</td>
</tr>
<tr>
<td></td>
<td>type=updateParameters</td>
<td>Sends a POST request with contents of parameters box to update the value of the collection defaults object defaults data stream, returns xml string with updated values</td>
</tr>
<tr>
<td>CollectionProcessor.java</td>
<td>type=offices</td>
<td>Converts the xml data string from FedoraAccess to a list of office names</td>
</tr>
<tr>
<td></td>
<td>type=collections</td>
<td>Converts the xml data string from FedoraAccess to a list of CollectionFragments</td>
</tr>
<tr>
<td>CollectionFragment.java</td>
<td>type=collections</td>
<td>This is a data object representing a subset of the collections information</td>
</tr>
<tr>
<td>DefaultProcessor.java (from FedoraDefaults)</td>
<td>type=parameters</td>
<td>Converts the xml data string from FedoraAccess to a parameters</td>
</tr>
<tr>
<td></td>
<td>type=updateParameters</td>
<td>Converts the list of Parameters to an xml string for FedoraAccess to write to the appropriate collection defaults object</td>
</tr>
<tr>
<td>Parameter.java</td>
<td>type=parameters</td>
<td>Data object representing a value of a parameters entry</td>
</tr>
<tr>
<td></td>
<td>type=updateParameters</td>
<td>Data object representing a complete parameter consists of an entry name and list of Parameter with the possible values</td>
</tr>
</tbody>
</table>

Data Dictionary

This page is initially built as a JSP page. When ControlSource.java receives a getDefinitions request, it calls SADefinitions.java to set the request attribute dictionary. SADefinitions gets the data the FedoraAccess.java which extracts the dataDictionary data stream from the defaults collection object. This is converted to a DataDictionary object by DataDictionaryProcessor.java. The returned page looks like this:
An administrator can change either the overview material or any of the definitions. Similarly he or she can add or delete definitions. **Warning:** the query system is closely tied to the data definitions, they must exist with the same names and order as in the querytable table in the appropriate database. So these should only be added to or deleted with the greatest of care. When an administrator clicks on the Update Dictionary button the overview and list of entries are sent to ControlSource.java via a POST AJAX command. The data is sent on to SAUpdate.java and hence to FedoraAccess.java and DataDictionaryProcessor.java to update the Fedora object. When this is done the data is returned in an xml string to update the page with the new values.

**Note:** For this page and the previous there needs to be a better way to let the administrator know that the update was completed successfully.

Web Page used: dataDictionary.jsp  
Database/Tables used: None  
Javascript files: jquery.js  
jquery.form.js  
definitions.js  
popup.js  
utility.js  
Java classes used:

<table>
<thead>
<tr>
<th>Class</th>
<th>Called via</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>ControlSource.java</td>
<td>All interactions</td>
<td>Calls FedoraAccess to return DataDictionary object. Sets a request attribute to this value.</td>
</tr>
<tr>
<td>SADefinitions.java</td>
<td>type = getDefinitions</td>
<td>Builds a new DataDictionary object from the updated values provided and uses DataDictionaryProcessor to convert it to an xml string that FedoraAccess can use to update the dataDictionary data stream</td>
</tr>
<tr>
<td></td>
<td>type=updateDefinitions</td>
<td></td>
</tr>
<tr>
<td>FedoraAccess.java</td>
<td>type = getDefinitions</td>
<td>Returns a DataDictionary object obtained by converting the dataDictionary data stream from the appropriate collections default object.</td>
</tr>
<tr>
<td></td>
<td>type=updateDefinitions</td>
<td>Uses the provided xml string to update the dataDictionary data stream in the appropriate collections default object</td>
</tr>
<tr>
<td>DataDictionaryProcessor.java</td>
<td>type=getDefinitions</td>
<td>Converts the xml data string from FedoraAccess to a DataDictionaryObject</td>
</tr>
<tr>
<td></td>
<td>type=updateDefinitions</td>
<td>Converts an overview and list of definitions to an xml string suitable for updating the fedora collections default object</td>
</tr>
<tr>
<td>DataDictionary.java</td>
<td>type=collects</td>
<td>This is a data object representing the overview and data definitions for a collection</td>
</tr>
</tbody>
</table>

**Storage and Storage Access**

The storage system is a combination of a dedicated storage system and virtual systems on a shared enterprise stack.

**Archival and Dark Storage: IX Storage**

Two **FreeNAS 4U IX Storage servers**, 35 TB each, are being used in the MSU Trusted Digital Repository. One is for Archival Storage, and the other is for redundant, limited-access Dark Archive. After ingest, Archival Information Packages (AIPs) containing original and preservation copies of digital objects, as well as metadata, are stored on the Archival Storage server. Dissemination Information Packages (DIPs) containing access copies of digital objects with metadata are stored on the IX in a different area. It can be possible to separate the two at a later time if necessary.

Archival Storage is regularly synchronized to the Dark Archive.

**Enterprise Storage for Proof-of-Concept Design**

Archives has a 2TB share of enterprise **NetApp storage**. This storage is replicated across two data servers. In addition, it is backed up nightly with **CommVault**. The 2TB storage is used for the Spartan Archive Proof-of-Concept system in addition to processing workspace. Access to proof-of-concept records and processing workspace have separate access to prevent human write-access to the records.

Access is through **Samba** on a **CentOS Linux** server.
Storage Access: Samba

On the desktop, archivists see the Samba as a file directory containing the following folders:

- **Workspace**: Used by archivists as a staging area for pre-ingest processing of accessioned files.
- **aips**: Archival Information Package (preservation copy) storage area on the Archival Storage server, for digital content ingested through Archivematica following the Trusted Digital Repository processing procedures. Mirrored and synched to the Dark Archive server, and viewable through the Samba.
- **dips**: Dissemination Information Package (access copy) storage area on the Archival Storage server, for digital content ingested through Archivematica following the Trusted Digital Repository processing procedures. Mirrored and synched to the Dark Archive server, and viewable through the Samba.
- **devaips**: Test Archival Information Package (preservation copy) storage area on the Archival Storage server, for testing digital ingest content ingested through Archivematica. Mirrored and synched to the Dark Archive server, and viewable through the Samba.
- **devdips**: Test Dissemination Information Package (access copy) storage area on the Archival Storage server, for testing digital content ingested through Archivematica. Mirrored and synched to the Dark Archive server, and viewable through the Samba.
- **records**: Archival Information Package (preservation copy) storage area for database records ingested through the Spartan Archive processing procedures.
- **warc**: Web Archival records from the Archive-It service. UAHC is downloading a copy of the captured websites (WARC format) and storing it in our preservation environment.

**Client Drop-off**

**AFS and NetFiles**

Andrew File System (AFS) space, in combination with the NetFiles user interface, is used for transferring files to the Spartan Archive repository. A Spartan Archive account in the MSU AFS has been created. UAHC may create accounts within the Spartan Archive account for each depositing unit. To date, an account has been created for the Office of the Registrar (RO).

When the RO uploads files to its Spartan Archive account, an alert is sent to the electronic records archivist. The files may then be approved and ingested into the repository through an automated process.

At this time, files from other offices uploaded to AFS would have to be manually downloaded by UAHC for processing and ingest.

**Spartan Archive Infrastructure**

Besides the storage and storage-access platforms already described in the section on Storage and Storage Access and Client Drop-off, Spartan Archive infrastructure includes an application/access server and database server.

**Application and Access Server**

A CentOS Linux server running the Open Java Platform OpenJDK and the Java servlet container Apache Tomcat is used to run both the Fedora Repository and other Spartan Archive custom software. This server is connected to the enterprise storage and client drop-off (AFS) storage.

**Database**

Another CentOS Linux server is utilized to run a MySQL open-source database. This database is used by the Fedora Software, Spartan Archive Access software, Archivist Toolkit.

**Software**

The Spartan Archive Repository and greater MSU Trusted Digital Repository make use of a number of software technologies, including:

**Fedora Repository Software**

The Spartan Archive Repository is based on Fedora repository software. Fedora open-source software is a Java-based architecture for storing, managing, and accessing digital content. During the ingest process, metadata for the digital content is created and stored in Fedora. PDFs ingested using PDF Puller are also stored in the Fedora-based repository and available for access through the public Spartan Archive website.

Information on the technical design of Fedora Commons for the Spartan Archive Repository may be found in the technical documentation. Refer to the Spartan Archive technical documentation and procedures on the public Spartan Archive project deliverables page.

UAHC plans to use Fedora as the repository software in the overall MSU Trusted Digital Repository.

**Spartan Archive Ingest Software**
Multiple custom-built Java tools were built and implemented into Spartan Archive to Ingest client submissions. Below are details of these custom-built programs.

**GetIngestFiles** is a Java batch program that discovers the new submissions, validates them for integrity, moves them to a quarantine area and sets a status marker to indicate that the submission is ready for submission. The program emails the administrator a notification of new ingests that are ready for processing with **SAControl**.

**Ingest2** is a Java batch program that copies the submission from the client drop-off area to enterprise storage, creates a Fedora Object for the submission and updates the collection and repository objects. At the end, Ingest2 updates the status marker.

Ingest is completed with a set of batch ingest programs that load the archival XML submissions to accessible databases for client search and access. These are:

- **AcadXMLToDatabase**: loads the Academic Programs Database
- **CDXMLToDatabase**: loads the Course Descriptions Database
- **SDXMLToDatabase**: loads the Student Directory Database
- **SOCXMLToDatabase**: loads the Schedule of Courses database
- **CreateLookupLists**: extracts data from the above databases to use as aids to researchers who are building queries.

**Spartan Archive Console (SAControl)**

The Spartan Archive Administration Interface Program, also known as SAControl or the Spartan Archive console, is a Java servlet that serves three purposes: for managing ingests of database content; for maintaining configuration and data dictionary information for Spartan Archive collections; and for viewing activity in the Spartan Archive system.

Detailed technical information about the console may be found at [SAControl Documentation](#).

**Spartan Archive Access Software**

The Spartan Archive Access system is a combination of web-accessible objects in the Fedora Repository Software and a custom-built Java servlet called **GeneralQuery**. **GeneralQuery** searches the access databases built with data from the Registrar Office XML file submissions from each record series and presents the results to the end-user.

**Spartan Archive Software Utility Programs and Libraries**

**Utility Programs**

Java utility programs and shell scripts were also created.

Utility programs include tools to build a database from a CSV file, validate submissions, integrity checking, and more. Some of the other programs use these programs.

Batch processes are managed by shell scripts run by a cron job.

**Libraries**

Libraries that are used:

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>fedora-client-0.1.6-with-dependencies.jar</td>
<td><a href="https://github.com/mediashelf/fedora-client">https://github.com/mediashelf/fedora-client</a></td>
<td>Provides programmatic access to fedora objects and data streams</td>
</tr>
<tr>
<td>jdom.jar</td>
<td><a href="http://www.jdom.org/">http://www.jdom.org/</a></td>
<td>Provides XML parsing support</td>
</tr>
<tr>
<td>log4j-1.2.16.jar</td>
<td><a href="http://logging.apache.org/log4j/">http://logging.apache.org/log4j/</a></td>
<td>Provides logging support</td>
</tr>
<tr>
<td>mysql-connector-java-5.1.16-bin.jar</td>
<td><a href="http://dev.mysql.com/downloads/connector/j">http://dev.mysql.com/downloads/connector/j</a></td>
<td>Provides programmatic access to the mysql database</td>
</tr>
</tbody>
</table>

**Archivists’ Toolkit**

UAHC uses **Archivists’ Toolkit** (AT), the open-source archival data management system, as its collection management system for all archival material. In addition to serving as a database of information on accessions and processed material, AT can be used to generate EAD-based finding aids.

The **Spartan Archive ingest process** creates a file that imports metadata into an AT accession record. On completion of the processing/ingest of digital material, the **AT collection record must be updated**.

As a charter member of the **ArchivesSpace** community, UAHC will migrate collection data from AT to ArchiveSpace by early 2015. Accessioning and ingest procedures will change as necessary to correspond with the functionality of the new system.
Archivematica Processing System

UAHC uses the open-source Archivematica digital preservation system software for processing and ingesting most file types, databases and websites being the exceptions. (Databases may be archived by following the operations outlined in Spartan Archive Repository Procedures and Processes. Select MSU websites are harvested and archived using the Archive-It service.)

Designed to maintain standards-based, long-term access to digital objects, Archivematica uses a micro-services design pattern to provide an integrated suite of software tools that allows users to process digital objects in compliance with the ISO-OAIS functional model. Users monitor and control the micro-services via a web-based dashboard. Archivematica uses METS, PREMIS, Dublin Core and other best practice metadata standards.

Depending on the file types, pre-ingest processing using other software tools may be required prior to ingest through Archivematica.

Software Processing Tools

A number of software processing tools are recommended for use when processing files for ingest into the MSU Trusted Digital Repository. For the latest list of tools refer to Useful Tools for Electronic Records Processing. This non-exhaustive list constantly evolves as new tools come available.

Information on how to use some of the tools may be found in Pre-Ingest Processing.

Useful Tools for Electronic Records Processing: TDR Documentation

Useful Tools For Electronic Records Processing

Identified by Michigan State University Archives

<table>
<thead>
<tr>
<th>TOOL NAME</th>
<th>DESCRIPTION</th>
<th>VENDOR</th>
<th>SOURCE</th>
<th>COST</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-Zip</td>
<td>open source compression software</td>
<td></td>
<td><a href="http://www.7-zip.org/">http://www.7-zip.org/</a></td>
<td>Free</td>
</tr>
<tr>
<td>CD Recovery Toolbox</td>
<td>recover damaged files from different disk types: CD, DVD, HD DVD, Blu-Ray, etc.</td>
<td>OEMailRecovery.com</td>
<td><a href="http://www.oemailrecovery.com/cd_recovery.html">http://www.oemailrecovery.com/cd_recovery.html</a></td>
<td>Free</td>
</tr>
<tr>
<td>Cdx</td>
<td>convert tracks on a CD to standard computer sound files</td>
<td>Georgy Berdyshev</td>
<td><a href="http://cdex.sourceforge.net/">http://cdex.sourceforge.net/</a></td>
<td>Free</td>
</tr>
<tr>
<td>CUSpider</td>
<td>identify files that may contain confidential data</td>
<td>Columbia Univeristy</td>
<td><a href="http://www2.cit.cornell.edu/security/tools/spider-windows.html">http://www2.cit.cornell.edu/security/tools/spider-windows.html</a></td>
<td>Free</td>
</tr>
<tr>
<td>Disk Image and Browse</td>
<td>Browse and make an image copy of an inserted 5.25” floppy</td>
<td>Device Side Data</td>
<td><a href="http://www.dewiveside.com/fc5025.html">http://www.dewiveside.com/fc5025.html</a></td>
<td>Requires special hardware</td>
</tr>
<tr>
<td>Forensic Software Utility</td>
<td>Write blocker software</td>
<td>Wiebetech</td>
<td><a href="http://www.wiebetech.com/software/forensic_software_utility.php">http://www.wiebetech.com/software/forensic_software_utility.php</a></td>
<td>Requires special hardware</td>
</tr>
<tr>
<td>FTKImager Lite</td>
<td>Creates images of disks, CDs, DVDs, etc.</td>
<td>AccessData</td>
<td>no longer available?</td>
<td>Free</td>
</tr>
<tr>
<td>Handbrake</td>
<td>open-source, GPL-licensed, multiplatform, multithreaded video transcoder. Used for extracting video from DVDs.</td>
<td>Open</td>
<td><a href="http://handbrake.fr/">http://handbrake.fr/</a></td>
<td>Free</td>
</tr>
<tr>
<td>MaclinkPlus Deluxe</td>
<td>converts older Mac file formats</td>
<td>Dataviz</td>
<td>Not for Sale; try eBay</td>
<td></td>
</tr>
</tbody>
</table>
ISO Recorder 3.1
make images of the existing data CDs and DVDs
Alex Feinman
http://isorecorder.alexfeinman.com/W7.htm
Free

MagicISO
mounts disk images
MagicISO
http://www.magiciso.com
Free

pcdMagic
converts Kodak Photo CD image files to tiff format and preserves embedded metadata
Sandy McGuflog
https://sites.google.com/site/pcdmagicsite/
$79

PST Viewer
converts email message to PDF; exports messages and attachments to a designated folder
Encryptomatic LLC
http://www.pstviewer.com/
$70

Quick View Plus
view virtually any file or email attachment
Avanstar
http://www.avantstar.com/metro/home/Products/QuickViewPlusStandardEdition
$49

Remove Empty Directories (RED)
searches and deletes empty directories recursively below a given start folder
Source Forge
http://sourceforge.net/projects/rem-empty-dir/
Free

Renamer
file renaming tool
Denis Kozlov
http://www.den4b.com/?x=
downloads&product=renamer
Free

SearchMyFiles
search and duplicate finder
Nirsoft
http://www.nirsoft.net/utils/search_my_files.html
Free

TeraCopy
Copy program that preserves creation date.
Code Sector
http://codesector.com/teracopy
Free

ViewNX 2
Browsing, sorting, and quick editing for both still images and movie files. This includes Nikon NEF raw files.
Nikon
https://support.nikonusa.com/app/answers/detail/a_id/61#Anchor-1
Free

VisiPics
Finds same or similar image files with a visual presentation.
http://www.visipics.info/
Free

VLC Media Player
open source cross-platform multimedia player and framework
VideoLAN Organization
http://www.videolan.org/vlc/download-windows.html
Free

WinMerge
File and Directory compare
Source Forge
http://winmerge.org
Free

XnView
multimedia viewer, browser and converter (exif, iptc)
XnSoft
Free

Security and Backup

Security

Only authenticated users may deposit files in the Spartan Archive account on AFS/NetFiles.

Only authenticated users may access the Archivematica or the NetApp/Samba.

Storage systems can only connect to specific servers that have been granted access through Access Control Lists (ACLs). Access to these servers are behind firewall rules. Client-access is granted with Samba server controls that use both user-access controls and firewall protections. Archivists only have read access to the AIPs and DIPs in both the Spartan Archive and Trusted Digital Repository. Firewall protections also limit access to the database.

Backup
The E-Records workstation, NetApp/Samba server, Fedora database server, AT database server, and Spartan Archive MySQL database server are all included in the regular MSU backup to CommVault and offsite tape (30-day recovery, effectively 2 months).

The Fedora databases are dumped to a file twice a day and captured in the CommVault backup.

Content on the Archival Storage IX server is mirrored and periodically synched to the Dark Archive IX server.

Archivematica output goes to dedicated IX Storage.

**Risks and Emergency Planning**

**Programmatic Risks**

Recognized risks to the MSU Trusted Digital Repository include the following resource constraints and technology issues. With the help of MSU IT Services, UAHC will strive to mitigate these risks by maintaining the resources and technologies necessary for ongoing development and continuation of the repository.

**Resource constraints:**
- System resources
- Development resources
- Funding
- Hardware

**Technology issues:**
- Keeping up with upgrades/patches
- Many different software packages tools, most open source
  - Could be compatibility issues as software is upgraded and modified
- Workflow stoppages if servers aren’t available
  - Samba/NetApp
  - Fedora
  - IX Storage
  - Database server
  - Access server (when available)
  - TAPER/submission agreement home server (when available)

**Emergency Planning**

UAHC recognizes that unexpected incidents putting data at risk can occur through technology failures, human actions, and natural disasters. The following risks to the digital content in the MSU Trusted Digital Repository have been identified, along with plans to recover in case of disaster.

<table>
<thead>
<tr>
<th>Risk</th>
<th>Disaster Recovery Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss of AIPs</td>
<td>If lost on one IX, can recover from other IX</td>
</tr>
<tr>
<td>Loss of Fedora databases — metadata, relationships</td>
<td>—Included in regular MSU backup to CommVault and offsite tape (30-day recovery, effectively 2 months)</td>
</tr>
<tr>
<td></td>
<td>—Backed up twice a day on same server</td>
</tr>
<tr>
<td>Loss of DIPs</td>
<td>—If more than one copy, can recover from second copy</td>
</tr>
<tr>
<td></td>
<td>—Can recreate from AIPs</td>
</tr>
<tr>
<td></td>
<td>—Refer to DIP Disaster Recovery</td>
</tr>
<tr>
<td>Flood or other common disaster strikes Computer Center / Admin Building area</td>
<td>No current recovery strategy; plans for third offsite storage site, cloud or distributed</td>
</tr>
<tr>
<td>—Loss of both IX Storage systems and AIPs</td>
<td>Included in regular MSU backup to CommVault and offsite tape (30-day recovery, effectively 2 months)</td>
</tr>
<tr>
<td>—Loss of Samba (current pre-accessions) and Fedora</td>
<td></td>
</tr>
</tbody>
</table>
IT Roles and Responsibilities

SLA

The MSU University Archives and Historical Collections (UAHC) technologies are supported by IT Services Content and Collaboration (CNC). The description and agreement of this support needs to be formally documented in a Service Level Agreement (SLA). CNC is currently undertaking a project to build and revamp SLAs throughout all departments of IT Services. Both UAHC and CNC are under the same leadership.

Customer-level SLA

Each customer of IT Service will get an SLA package with attached SLA Addendums for each service included in the total customer service package.

Master agreement sections

1. Business Service Description.
2. Service Level Agreement Overview.
3. Agreement Period.
4. Services Included.
5. Maintenance.
6. Requesting Support
7. Priority of Support
8. Amendments or Termination.
9. Cost of Agreement.

Addendums sections

1. Service Description
2. IT Services Responsibilities
3. Customer Responsibilities
4. Assumptions
5. Service Support Hours
6. Requesting Support
7. Response and Resolution Times
8. Billing information
9. Price
10. Authorized Contact List

Suggestions for inclusion

Some details to possibly go in the SLA between UAHC and IT Services:

- frequency of integrity checks.
- responsibilities in regards to Tech Watch

Current IT Services Work to Address in SLA

- Tech watch assistance: IT Staff subscribe to some email lists of archival software currently in use.
- Update watch: IT staff subscribe to email lists regarding patches, updates, and upgrades.
- Virtual servers configured to do automatic checks of available software updates. Messages are sent to IT Staff.
- Updates applied during regular maintenance window. IT Staff developed internal testing checklist to perform after maintenance.
- Use of a Quality Assurance (QA) and Development (dev) systems for release management.
- Rudimentary change management: IT Staff and UAHC staff regularly communicate.
- Dedicated storage has a support plan in place with vendor. Enterprise storage systems from central IT have procedures and trained staff.
- The storage platform dedicated to UAHC is a similar device to other storage systems used and managed by staff who are experienced with updating and migrating storage.
- The virtual servers are on a virtual stack that is managed by staff who are experienced with updating and migrating resources.

Future Repository

The Spartan Archive Repository for database content and use of Archivematica for ingest and AIP and DIP creation are only the beginning of the development of the MSU Trusted Digital Repository system. Functional requirements for the greater system may be found at the following links:

- Submission Functions/Requirements
- Administration Functions/Requirements
- Data Management Functions/Requirements
Access Functions/Requirements

Technologies under exploration for future use in the MSU Trusted Digital Repository include:

**ArchivesSpace**

In late 2014/early 2015, UAHC will adopt the ArchivesSpace open-source archives information management application as a replacement for the Archives’ Toolkit (AT) collection management system. Content will be migrated from AT to ArchivesSpace at that time, and Spartan Archive Repository and MSU Trusted Digital Repository processing procedures will be updated to work with ArchivesSpace.

Designed by archivists, ArchivesSpace has a web browser-based interface for describing, managing, and providing access to archives, manuscripts, and digital objects. ArchivesSpace is considered a next-generation application, the result of a merger of AT and Archon collection management technologies.

The touted primary added value for the first release is the integration of AT’s collection management and metadata authoring features with Archon's web access features. ArchivesSpace also “offers additional modules for rights management, collection management, recording events (collection management actions), and recording core technical metadata, as well improved workflows and a more robust technical architecture and support infrastructure” and is expected to be more intuitive to learn and easier to maintain.

Michigan State University is a charter member of the ArchivesSpace community.

**Tufts Accessioning Program for Electronic Records (TAPER)**

UAHC has adapted the Tufts Accessioning Program for Electronic Records (TAPER) for possible use as an automated submission solution in the MSU Trusted Digital Repository. Developed under an NHPRC grant, the Perl-based submission application would allow campus units to submit digital files and transmittal information via a web browser-based form interface.

The decision to explore the use of TAPER code in developing an automated submission process followed from the documentation of Submission Functions/Requirements and an implementation recommendation.

**Islandora**

UAHC is exploring the open-source Islandora software framework as an access system for the MSU Trusted Digital Repository. Built on the PHP and Python scripting languages and the Java programming language, Islandora utilizes Drupal content management software, Fedora repository software, and Apache Solr search technology for the management and discovery of digital content.

Islandora has the potential to integrate with the Archivematica processing solution, and its functionality matches most of the needs outlined in the Access Functions/Requirements document.

**OAIS and TRAC**

**OAIS**

The Spartan Archive project model is based on the principles of best practices and emerging standards and intends to use the OAIS model as a conceptual framework. Development of the program is intended as a proof-of-concept model with practical, real-life solutions to address the archival preservation and access needs faced by most U.S. higher education institutions.

**Preservation Planning**

Preservation Planning activities included analyzing the content, context and structure of each records series; engaging with the user community to understand their current and future needs; identifying individual data elements to be archived for each records series; determining preservation metadata requirements for each records series and the entire collection; and determining search, navigation, and access requirements for each records series.

**Archival Administration**

**Spartan Archive:** Archival Administration consists of technology to control security and access management to processes and storage. The SA Console is used by archivists to review and approve client-submitted records. Validation software validates the client’s submissions. Integrity checking is performed.

**TDR:** Access to processing workspace and Archivematica ingest software is controlled via authentication. Electronic records archivists have superuser permissions in Archivematica, allowing for manual management or deletion of AIPs and DIPs if necessary.

**Metadata Management**
Spartan Archive: The combination of Archivist Toolkit, Spartan Archive Console, and the Fedora repository allow for the application and use of metadata.

TDR: Combination of Archivist Toolkit, Archivematica, and eventually integration with Fedora repository.

Archival Storage

See information regarding storage and storage access.

Archival Access

Spartan Archive: See information regarding client access.

TDR: See plans at Islandora page.

Preservation Environment

Spartan Archive: Archivists cannot edit AIPs through storage access. DIPs are available through the client access software. All data is on enterprise storage with enterprise backup for full recovery.

TDR: There is a separation between AIPs and DIPs. Access to the Preservation Environment (AIPs) is limited. The Preservation Environment will not be available for routine search and retrieval of archived records series. However, in the event of loss of the Current Archive, processes will be developed to use content saved in the Preservation Environment to restore the data. This will ensure that the information can be fully restored if necessary. In addition to the AIPs and DIPs location, there is also a dark archive replicating the storage.

TRAC Checklist

Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist

TRAC Checklist for Spartan Archive (Needs updating, Update is planned.)

On the onset of the project, the TRAC criteria was used in the planning of the Spartan Archive proposal, and continues to shape the working practices and design of Spartan Archive and the larger Trusted Digital Repository (TDR). Subsets of the TRAC criteria apply to different elements of the Spartan Archive project. For example, Subset A. Organizational Infrastructure, will be used to evaluate the Spartan Archive project as a whole and identify gaps in the program such as governance structure, policies, procedures, staff expertise or operational support. The TRAC criteria’s Subset B. Digital Object Management is directly applicable to the workflow of the Spartan Archive project and will allow the project team to audit the various components of the complex system. Analysis of Spartan Archive in comparison to Subset C. Technologies, Technical Infrastructure & Security will ensure the project team is creating a secure and trusted technical infrastructure to archive the institution’s electronic business records. The overall intent of using the TRAC criteria and checklist as an evaluation tool is twofold – 1) to allow the project team to identify weaknesses and plan ahead to address gaps and 2) to utilize an established standard to assess the project and communicate its successes and challenges to the larger archival community. Cynthia: The TRAC checklist will be reviewed on a schedule. (yearly? bi-yearly?)

Appendices

Appendix A: Glossary

This glossary of digital preservation terminology as used by the Michigan State University Digital Preservation Program is based on definitions found in the following documentation:

- California Digital Library Glossary (CDL)
- Digital Preservation Management Workshop, Terms and Concepts (DPW)
- Reference Model for an Open Archival Information System (OAIS), Definitions (OAIS)
- Trusted Digital Repositories: Attributes and Responsibilities (TDR)
- Tufts University Ingest Guide for University Records, Submission Agreement (Tufts)

Access Functional Entity: The OAIS functional entity that provides the services and functions that support Consumers in determining the existence, description, location, and availability of information stored in the OAIS, and allowing Consumers to request and receive information products. (OAIS)

Access Rights Information: In OAIS, the information that identifies the access restrictions pertaining to the digital object, including the legal framework, licensing terms, and access control. It contains the access and distribution conditions stated within the Submission Agreement, related
to both preservation (by the OAIS) and final usage (by the Consumer). It also includes the specifications for the application of rights enforcement measures. Rights information may be considered preservation metadata. (OAIS, DPW)

**Administration Functional Entity:** The OAIS functional entity that contains the services and functions needed to control the operation of the OAIS on a day-to-day basis. (OAIS)

**Archival Information Package (AIP):** An information package consisting of the digital object and the associated Preservation Description Information (PDI), which is preserved within an OAIS. (OAIS)

**Archival Storage Functional Entity:** The OAIS functional entity that provides the services and functions for the storage, maintenance and retrieval of AIPs. Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfill orders. (OAIS)

**Consumer:** The role played by those persons, or client systems, that interact with OAIS services to find preserved information of interest and to access that information in detail. (OAIS)

**Context Information:** In OAIS, the information that documents the relationships of the digital object to its environment. This includes why it was created and how it relates to other digital objects. Context information may be considered preservation metadata. (OAIS, DPW)

**Dark Archive:** An archive that is inaccessible to the public. It is typically used for the preservation of content that is accessible elsewhere. (CDL)

**Data Dictionary:** A formal repository of terms used to describe data. (OAIS)

**Data Management Functional Entity:** The OAIS functional entity that provides the services and functions for populating, maintaining, and accessing both Descriptive Information which identifies and documents archive holdings and administrative data used to manage the archive. (OAIS)

**Descriptive Metadata:** Metadata used for the discovery and interpretation of a digital object. (CDL)

**Designated Community:** An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. (OAIS)

**Digital Object:** The digital content (file or entity) to be preserved.

**Digital Preservation:** The managed activities necessary for ensuring the long-term retention and usability of digital objects. (CDL)

**Dissemination Information Package (DIP):** An information package, derived from one or more AIPs, received by the Consumer in response to a request to the OAIS. (OAIS)

**Fixity Information:** In OAIS, the information which documents the mechanisms that ensure that the digital object has not been altered in an undocumented manner. Technologies such as checksums and digital signatures are used to verify that a digital object retains its fixity, which helps maintain the object's authenticity and integrity. Fixity information may be considered preservation metadata. (OAIS, DPW)

**Ingest Functional Entity:** The OAIS functional entity that provides the services and functions to accept Submission Information Packages (SIPs) from Producers (or from internal elements under Administration control) and prepare the contents for storage and management within the archive. (OAIS)

**Long Term:** A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community, on the information being held in an OAIS. This period extends into the indefinite future. (OAIS)

**Long Term Preservation:** The act of maintaining information, Independently Understandable by a Designated Community, and with evidence supporting its Authenticity, over the Long Term. (OAIS)

**Management:** The role played by those who set overall OAIS policy as one component in a broader policy domain. (OAIS)

**Metadata:** Data that describes or characterizes a digital object, whether internal or external to the object itself. It is often subdivided into categories such as descriptive, administrative, and structural. Preservation metadata is not so much a separate category as an amalgam of other types—including descriptive, technical, structural, and rights metadata—as virtually any metadata element can be seen as having value for preservation purposes. (DPW)

**Migration:** To copy data, or convert data, from one technology to another, whether hardware or software, preserving the essential characteristics of the data. This includes the conversion of data to avoid obsolescence of the physical storage medium as well as the encoding and format of the data. (DPW)

**Normalization:** Within an archival repository, all digital objects of a particular type (e.g., color images, structured text) are converted into a single chosen file format that is thought to embody the best overall compromise amongst characteristics such as functionality, longevity, and preservability. (DPW)
Open Archival Information System (OAIS): An Archive, consisting of an organization, which may be part of a larger organization, of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. (OAIS)

Preservation Description Information (PDI): In OAIS, the information necessary for adequate preservation of the digital object and which can be categorized as Provenance, Reference, Fixity, Context, and Access Rights Information. (OAIS)

Preservation Planning Functional Entity: The OAIS functional entity that provides the services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete. (OAIS)

Producer: The role played by those persons or client systems that provide the information to be preserved. (OAIS)

Provenance Information: In OAIS, the information that documents the history of the digital object. This information tells the origin or source of the digital object, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. Provenance information may be considered preservation metadata. (OAIS, DPW)

Reference Information: In OAIS, the information that is used as an identifier for the digital object. It also includes identifiers that allow outside systems to refer unambiguously to a particular digital object. (OAIS)

Refreshment: To copy digital information from one long-term storage medium to another of the same type, with no change in the bitstream. (DPW)

Replication: Copying digital content to multiple storage locations with the intention to enhance its longevity while maintaining authenticity and integrity. (DPW)

Submission Agreement: The agreement reached between an OAIS and the Producer that specifies the format/contents of submitted material and any other arrangements needed for submission of digital content. It provides both entities with a document describing their responsibilities and grants the repository the right to electronically store, convert, and copy digital assets for preservation purposes. (OAIS, CDL, Tufts)

Submission Information Package (SIP): An information package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs. (OAIS)

Trusted Digital Repository: Repository with the mission to provide reliable, long-term access to managed digital resources to its designated community, now and in the future. (TDR)

Appendix B: Supported Digital Formats

The MSU Digital Repository commits to preserving most digital content in the form in which it is originally deposited and, for some formats, will preserve the content, structure and functionality of the files through normalization, migration, or other preservation strategies.

Acceptable file formats for preservation include:

<table>
<thead>
<tr>
<th>File Type</th>
<th>File Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital images and photographs</td>
<td>TIFF, JPG, JP2, PCT, PNG, PSD, BMP, GIF, TGA, AI, EPS, SVG</td>
</tr>
<tr>
<td>Spreadsheet files</td>
<td>XLS, XLSX</td>
</tr>
<tr>
<td>Web records</td>
<td>HTML*</td>
</tr>
<tr>
<td>E-mail, including attachments</td>
<td>PST, MSG</td>
</tr>
<tr>
<td>Audio files</td>
<td>AC3, AIFF, MP3, WAV, WMA</td>
</tr>
<tr>
<td>Video files</td>
<td>AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV</td>
</tr>
<tr>
<td>Text files</td>
<td>TXT, DOC, DOCX, ODF, ODT, WPD, RTF, PDF</td>
</tr>
<tr>
<td>Presentation files</td>
<td>PPT, PPTX</td>
</tr>
</tbody>
</table>

- Limited support for self-contained HTML content. Since 2012, many Michigan State University websites have been archived using the Archive-It technology.
Appendix C: Format Policies

Preservation and access format policies used by the MSU Trusted Digital Repository are based on the Archivematica format policies.

<table>
<thead>
<tr>
<th>Media Type</th>
<th>File Format(s)</th>
<th>Preservation format(s)</th>
<th>Access format(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>AC3, AIFF, MP3, WAV, WMA</td>
<td>WAVE (LPCM)</td>
<td>MP3</td>
</tr>
<tr>
<td>Office Open XML</td>
<td>DOCX, PPTX, XLSX</td>
<td>Original format</td>
<td>PDF for PPTX</td>
</tr>
<tr>
<td>Plain text</td>
<td>TXT</td>
<td>Original format</td>
<td>Original format</td>
</tr>
<tr>
<td>Portable Document Format</td>
<td>PDF</td>
<td>PDF/A</td>
<td>Original format</td>
</tr>
<tr>
<td>Presentation files</td>
<td>PPT</td>
<td>Original format</td>
<td>PDF</td>
</tr>
<tr>
<td>Raster images</td>
<td>BMP, GIF, JPG, JP2*, PCT, PNT*, PSD, TIFF, TGA</td>
<td>Uncompressed TIFF</td>
<td>JPEG</td>
</tr>
<tr>
<td>Spreadsheets</td>
<td>XLS</td>
<td>Original format</td>
<td>Original format</td>
</tr>
<tr>
<td>Vector images</td>
<td>AI, EPS, SVG</td>
<td>SVG</td>
<td>Original format</td>
</tr>
<tr>
<td>Video</td>
<td>AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV</td>
<td>FFV1/LPCM in MKV</td>
<td></td>
</tr>
<tr>
<td>Websites**</td>
<td>HTML</td>
<td>WARC</td>
<td>NA</td>
</tr>
</tbody>
</table>

* PNG and JPEG2000 are not normalized to a preservation format.

** MSU websites are harvested and archived as WARC files using the Archive-It service.

References


