

Proposed Outline of Digital Preservation Activities in the Near Future
UA Libraries

I. Policies and Guidelines

1. Review and assess the proposed policies documented on the wiki: http://intradev.lib.ua.edu/groups/Digital_Services/Preservation, including the high level policies: http://intradev.lib.ua.edu/groups/Digital_Services/HighLevelPolicies and format support: http://intradev.lib.ua.edu/groups/Digital_Services/Formats
2. Identify gaps and assign tasks to research and remediate those gaps
3. Review and assess remediated policies

II. Obtaining Rights Statements

1. Develop appropriate rights agreements to be signed by creators/donors to support digitization, access and delivery, and preservation actions
2. Develop and implement appropriate methods of gathering these documents, both for born-digital content and analog content to be digitized
3. Develop and implement appropriate tracking and storing mechanisms for these

III. Metadata

1. Provenance
 - a. We need to identify the necessary provenance information, and select a scheme and profile
 - b. We need to determine how to gather that information and encode it
 - c. We need to develop a workflow and storage procedures
2. Rights
 - a. We need to document at least 2 kinds of rights: those given us by the users (see II above) and the rights we have over the content, and possibly also the rights available to users.
 - b. For each type of rights, we need to identify the necessary information, select a scheme and profile
 - c. We need to determine how best to gather that information and encode it
 - d. We need to develop a workflow and storage procedures.
3. Technical
 - a. We need to develop profiles of AudioMD and MIX for our audio and image content.
 - b. We need to test multiple sets of samples from content digitized on different equipment and with different software with FITS and JHOVE, to determine the best method of extracting technical metadata automatically, and to identify problem areas

- c. We need to compare extracted metadata to selected profiles to identify fields which will need to be filled in manually or by other means
 - d. We need to develop workflows to capture that information
 - e. We need to develop scripts to automate extraction of metadata and to transform it into the selected scheme and profile, and store the content in correct directories, testing for missing information
 - f. We need to explore other possibly helpful initiatives (BatchBuilder, ACE, and Indiana ISO capture)
4. Administrative (other)
- a. PREMIS may inform the selection of administrative fields
 - b. Some are determined by database tracking needs. For each, we need to automate information capture and update database, as well as make periodic exports of database content to plain text for the archive
 - i. Type of format and version of format (must be tracked for migration purposes)
 - j. Type of metadata and version of metadata (same)
 - k. Documentation of anomalies in the collection or the metadata or processing
 - c. We need to capture operating system and file system metadata, store it in a profiled XML text file in the archive
 - d. We need to capture open source software (including an OS which will support them) and keep a copy in the archive
 - e. Possibly apply the BagIt protocol, to prepare content for possible inclusion in a larger storage repository
 - f. Provide a manifest at the top level which describes what our repository contains, the layout, and the location of necessary information for reconstruction
5. Descriptive
- a. We have already agreed upon MODS for descriptive metadata.
 - b. We need a policy decision on at what point the MODS is ready for inclusion in preservation.
 - c. Existing MODS need to be remedied to be valid, and possibly to include PURLs.
 - d. We need to publicly document our MODS profile and data dictionary, and include a copy in our archive.
6. Container
- a. We need a METS profile to contain metadata and links to content, and potentially links to updated metadata as well. This profile needs to be publicly posted as well as included in our archive.
 - b. Jody needs to complete the script that will create the METS from content and metadata once they are located in correct directories and named appropriately.
 - c. Once completed, we need to develop a workflow to automate the creation of all supporting metadata and the combination of it into the finalized METS file
 - d. and then script storage in the archive and linking into the LOCKSS manifest

- e. We also need policies on updates and alterations to stored, “finalized” metadata.

IV. LOCKSS

1. We need to be involved in a LOCKSS partnership capable of maintaining sufficient (at least 6) copies of our content in disparate locations
2. That partnership needs to be operating up to speed.