WAYNE STATE UNIVERSITY
Walter P. Reuther Library, Archives of Labor and Urban Affairs

Digital Curation Program

POLICY

Purpose

The Digital Preservation Policy supports the mission of the Walter P. Reuther Library and should be viewed as a supplement to the Collection Policy. While reflective of established policies and procedures, it serves as the highest-level digital preservation policy document at the Reuther Library. Recognizing the unique challenges inherent in caring for and making digital formats accessible and the complex relationship between preservation and access of digital content, this policy explicitly states the Reuther Library's dedication to a comprehensive digital preservation program for both born-digital and digitized collection material. As digital information becomes increasingly a larger component of the historical record, providing a framework for ensuring sustainable access to its digital collections is critical to the future relevance of the Reuther Library. This policy will be routinely updated and may be altered as technologies and standards evolve affecting the ability to responsibly manage digital collections.

Mandate

The mandate for digital preservation at the Reuther Library is driven by mission, institutional policies, contractual responsibilities to donor institutions, commitment to stakeholders, and professional obligation. The Reuther Library will continue to acquire, preserve, and make accessible primary source material of enduring historical or cultural value regardless of format changes in information creation. Thus, the Reuther Library will follow a policy of active preservation that safeguards the authenticity and utility of digital resources entrusted to its care for the foreseeable future.

Objectives

The main goal of the Digital Preservation Policy is to ensure continued access to reliable and authentic digital records through a program that adheres to changing standards and best practices in order to promote sustained use of Reuther Library collections.

- Develop and maintain processes and systems to responsibly collect, manage, preserve, and make accessible digital material
- Adapt to the needs of our users in a cost-effective and conscientious manner as the technology for digital content creation and distribution evolves
- Explore collaborative opportunities to better utilize resources and avoid duplication of effort
- Establish and maintain OAIS (Open Archival Information System) compliancy and meet certification criteria as a trusted digital repository

Scope

This policy addresses preservation of:

- Unique born-digital material donated or deposited with the Reuther Library through the regular course of acquisition

- Digital material created by the Reuther Library or its designees through digitization of analog collections for which the Reuther Library is the primary custodian

Particular priority will be given to material that exists in digital form only, due to danger of obsolescence or loss.

When possible, preservation decisions will be made at the time of creation or acquisition of the digital materials. In accordance with criteria by which all collections, regardless of format, are evaluated, such as intrinsic value and condition, preservation decisions will also take into account the feasibility of long-term access, in consultation with the appropriate information technology experts.

Challenges

This policy acknowledges that there are challenges inherent in a digital preservation program that must be addressed through that program. Most notably:

- Creating and maintaining adherence to submission standards
- Financial commitment necessary to ensure long-term viability of a preservation program
- Ever-increasing volume of digital materials
- Rapid evolution of technologies used to capture, store, and represent digital materials
- Continuing education of staff involved with digital material
- Time needed to maintain and describe the complex relationships between components of digital objects
- Administrative complexities in ensuring cost-effective and timely action to implement preservation strategies

Operating Principles

The Reuther Library Digital Preservation Program will:

- Comply with Open Archival Information System (OAIS) Reference Model standard
- Perform self-assessments to insure adherence to requirements of a trusted digital repository (as defined by the 2002 RLG/OCLC report)
- Commit to an interoperable, scalable digital archives with appropriate storage management capabilities
- Establish minimum metadata requirements for all files to support preservation strategies and archival obligations, such as fixity, provenance, authenticity, and discoverability
- Manage necessary technology, appropriately and cost-effectively

- Follow current best practices and monitor new developments in the field to ensure interoperability and relevance
- Work with records creators to integrate preservations strategies in lifecycle management at the earliest stage possible
- Take action even if no perfect solution exists, on what ever scale is practical and ethical, understanding and documenting given limitations

Roles and Responsibilities

The Reuther Library and larger University Library System are committed to financial support, ensuring an adequate level of funding and resources necessary to sustain the digital preservation program. The Reuther Library accepts primary responsibility for the long-term preservation of and access to its digital collections. The Collections Team and Director periodically evaluate the Digital Preservation Policy and other high-level documents as well as programmatic plans and progress. The Digital Information Technology Team creates and maintains the procedures for management of the lifecycle of digital objects. Collection Archivists and the Digital Resources Specialist apply policy and procedure to active digital stewardship. ULS Discovery Services collaborates on digital repository design and software maintenance. The University IT Department provides server space and input on system security.

Selection and Acquisition

The Digital Preservation Decision Flowchart (see appendix) details the rationale and processes governing criteria necessary for developing and retaining digital records of enduring value.

Access and Use

The Reuther Library's Collection Policy defines users of and access to its collections, regardless of format. Restrictions on digital content will be determined by donor agreements. The expectation is that digital acquisitions will be preserved long-term and made renderable to users wherever access and use restrictions do not apply and limitations of current technologies employed by the Reuther Library do not hinder retrieval.

PROCEDURES

Purpose and Audience

The Procedures section outlines the Reuther Library's high-level digital practices and workflows based on the archival community's current (2015 May) best practices. These high-level procedures will instruct the selection of software tools and resulting workflow creation. This section is designed for the archivists who will create the workflows, and as a reference for the archivists selecting software tools.

Roles and Responsibilities

The donor (sometimes also the content creator) of the materials is responsible for the following actions:

- Transferring material (original data and metadata) to archives via the Submission Information Packet (SIP)
- Signing the submission and donor agreements
- Ongoing consultation about the digital materials (if necessary)

The archivists performing digitization tasks (referred to as the **digitization archivists**) including Collection Archivists, the Technical and Metadata Archivist, and the Field Archivist are responsible for:

- Performing the following archival methodologies: selection and acquisition, appraisal, transfer, process, and access as appropriate on materials in the Submission Information Packet (SIP) and Archival Information Packet (AIP)
- Creating, updating, and evaluating the workflows followed by the digitization archivists

The archivists performing preservation tasks (referred to as the **preservation archivists**) including the Digital Resources Specialist, are responsible for performing the following digitization program responsibilities:

- Managing the lifecycle of digital objects and providing data preservation treatments as needed on materials in the Submission Information Packet (SIP), Archival Information Packet (AIP), and Dissemination Information Packet (DIP)
- Providing documentation about supported formats and preservation plans
- Determining associated costs
- Managing the digitization program

Digital Preservation Strategies

Archivists at the Reuther Library strategically perform digital preservation measures at both the broader program level and more specifically at the collection level.

At the program level, archivists focus on internal and external approaches to preservation:

- Donor Relations: **Digitization archivists** establish good relationships with donors and the producers of digital archival materials, learning as much as possible about the context of the digital materials.
- Monitor Technologies and Risks: **Preservation archivist** stays aware of and makes recommendations on preservation and storage innovations.
    - Preservation strategies (migration, emulation, etc. and specific procedures)
    - Storage Considerations (longevity, capacity, scalability, redundancy, security/susceptibility to threats, sustainability)
- Staff Training: Staff, supported by the director and University Library System, should actively seek and pursue professional development in areas that support the digital preservation program.

At the collection level, archivists focus on approaches
- Conversion and Migration: Every 2 years the **preservation archivist** evaluates material for format conversion or migration needs, and if needed, conducts the appropriate procedure according to current archival best practices.
- Metadata: **Digitization archivists** determine appropriate metadata per object type, adding descriptive, administrative, and preservation metadata to the digital archival materials.

Services and Equipment

The Reuther Library's successful management of digital assets depends upon ensuring that the digital assets retain the following:

1. Authenticity: Digital objects are what they claim to be
2. Integrity: Digital objects are the same as when the Reuther Library received them
3. Significant Properties: Digital objects are (and remain) linked to their metadata, conveying content information, provenance, etc.
4. Enduring Access: Digital material is preserved in such a way that it remains accessible to users.

The Reuther Library meets the above management goals by providing the following basic services for digital materials:

- Transfer and ingest: The Reuther Library works with the producer/donor of the materials, ensuring the digital materials are securely and authentically relocated (physically or virtually) to the Reuther Library.
- Backups: After the Submission Information Packets (SIPs) are transferred or Archival Information Packets (AIPs) are generated, it is backed up, ensuring there is both an access and a preservation (i.e., permanent dark archive) copy.
- Data Encryption: The Reuther Library can encrypt data if the donor has security concerns.
- File Format Conversion: For some digital materials that are in file formats that are obsolete, the Reuther Library can normalize the materials into a more modern file format, making the material renderable and viewable. The

preservation copy (containing the materials in their original format) will be retained too.

- Virus checks: An initial post-transfer virus check ensures that viruses have not infected the digital materials. If the materials are infected, steps must be taken to clear the virus, or the producer (or donor) may need to be contacted and further action taken.
- Checksum/Fixity checks: The initial checksum establishes the file's baseline, and subsequent fixity checks verify that the file's authenticity and integrity. All copies of the digital materials, the preservation copy (i.e., the materials as accessioned), and the processed materials (both online and nearline, meaning access only within the Reuther Library) will be verified with fixity checks.
- Access: After completed processing, the digital archival materials are accessible via the Fedora Commons Digital Object Management system (DOM).
- Migration: The digital materials will be evaluated every two years to determine if migration is needed to ensure accessibility.

To perform the above services, the Reuther Library establishes a technical infrastructure using the equipment and providing the support detailed below:

- Forensic workstation: This workstation must be non-networked, so that the digital materials do not potentially contaminate (or become contaminated by) networked digital content. The workstation will be used for the initial:
    - Transfer or capture of materials
    - Virus scan
    - Fixity check

  After the successful completion of these procedures, the materials may be copied to another folder for processing which can be performed on a **digitization archivist's** own workstation.
- Hosted services: Servers to store materials (currently the Reuther has the digital hold (31.5 TB, 77% free) with NAS backup (19TB free)
- IT Support: Assistance and support for the technical infrastructure is provided by the Reuther Library's Digital Initiatives and Technologies Team (DITT).
- Storage space (internal): To ensure that digital materials are stored in the interim time while archivists are processing.
- Secure storage space (external): To ensure that digital materials are securely stored, at least one copy of each preservation copy (Submission Information Packet (SIP) and processed materials (Archival Information Packet (AIP) are held in offsite storage (digital archive) in case of an onsite disaster.
- Staff training: Ongoing support for staff as new tools and software are added to the technical infrastructure and digital preservation workflows.
- Tools, Software, and Licenses:
    - The Reuther Library is implementing ArchivesSpace as its new archival management system.
    - Fedora Commons is in the early stages of adoption as the digital object management system (this is where the Archival Information Packets

(AIPs), and when needed, Dissemination Information Packets (DIPs), are stored and accessed by users).
- More tools, software and licenses will be added to this list as they are selected.

Archival Methodologies Workflow

Digitization archivists perform the high-level workflow detailed below. The workflow will be used to inform the selection of software tools to carry out tasks. Once tools are chosen, the specific procedures will be created based on the workflow below to include specific steps and tasks.

1. Selection and Acquisition (see Selection and Acquisition Flowchart)
    a. Determine if the digital material should be acquired by Reuther (see Digital Preservation Decision Flowchart)
        i. Review Reuther Collection and Digital Preservation policies (creation context, record value, collecting scope, etc.)
        ii. Perform technical appraisal (custody, authenticity, completeness, volume, preservability, accessibility, reference frequency, cost)
        iii. Conduct donor interview (creation practices, technical environment, primary object of preservation, dependencies, identification schema, essential metadata, intellectual property or copyright issues, sensitive personal or organizational information, potential)
    b. Discuss transfer with donor (SIP requirements, transfer/capture method, physical storage method – if applicable)
    c. Establish Submission Agreement (legally binding document, sets the terms and establishes how the digital materials will be transferred to or captured by the Reuther Library) with donor
    d. Donor and Reuther Library representative sign the Deed of Gift and Submission Agreement

2. Transfer (Transfer Flowchart)
    a. Transfer or directly capture digital materials
    b. Backup transferred digital materials
    c. Contact donor and inform whether or not the capture or transfer was successful
    d. If capture or transfer was unsuccessful, discuss with donor, and try again.
    e. Dispose of any physical media (only necessary if physical capture method was used)

3. Accession (Accession Flowchart) (Reuther now has legal custody)
    a. Review content and condition of digital material
    b. Create case file (i.e., establish initial intellectual control)
        i. Create accession record (includes the manifest/file inventory/directory listing) in ArchivesSpace

ii. The accession number also serves as the persistent identifier, the unique ID for the digital collection (this ties the Preservation Description Information (PDI) with the content)
iii. Record and capture supporting documentation (includes the transfer paperwork/surveys, deed of gift, correspondence, etc.)

c. Stabilize the digital records: perform validation and conservation assessment on preservation copy (i.e., records as accessioned)
    i. Run virus check
    ii. Obtain (and record) checksums by running fixity check (recommend using MD5)
    iii. Identify and record information about the digital records: file formats, extent, structure, etc.
d. Ingest and create the initial Archival Information Packet (AIP) (these are the "working" files) from the Submission Information Packet (SIP)
    i. Copy the files from the Submission Information Packet (SIP) into a new folder. This new folder is now the working folder, which will be processed.
    ii. Perform basic ingest actions: fixity check, identify and record file formats, extent, structure, etc. (required of all accessions)
    iii. Perform advanced ingest actions: scan for private information, normalization to ensure access (if doing, automation and data extraction, integration of tools) (preferred)
    iv. Upload to preservation repository (dark archive) and access systems

4. Process (Process Flowchart) (Archival Information Packets (AIPs) can be made of multiple accessions)
    a. Develop processing plan
        i. Determine the intellectual arrangement
        ii. Understand and record relationships between files and folders
    b. Physically arrange records if necessary (if doing this, create a copy to work with, and rerun fixity check)
        i. Identify and record preservation and access issues
        ii. Organize records (See Reuther Manual for appropriate arrangement level information)
        iii. Remove any private/confidential information, create online and nearline (if needed) versions
    c. Describe records by recording and generating metadata
        i. Descriptive Metadata: Provides information about the materials making them findable, useable, and accessible. Include event metadata (also included in the manifest and log)
        ii. Administrative Metadata: Includes the custodial history, use and access rights, etc.
        iii. Preservation Metadata

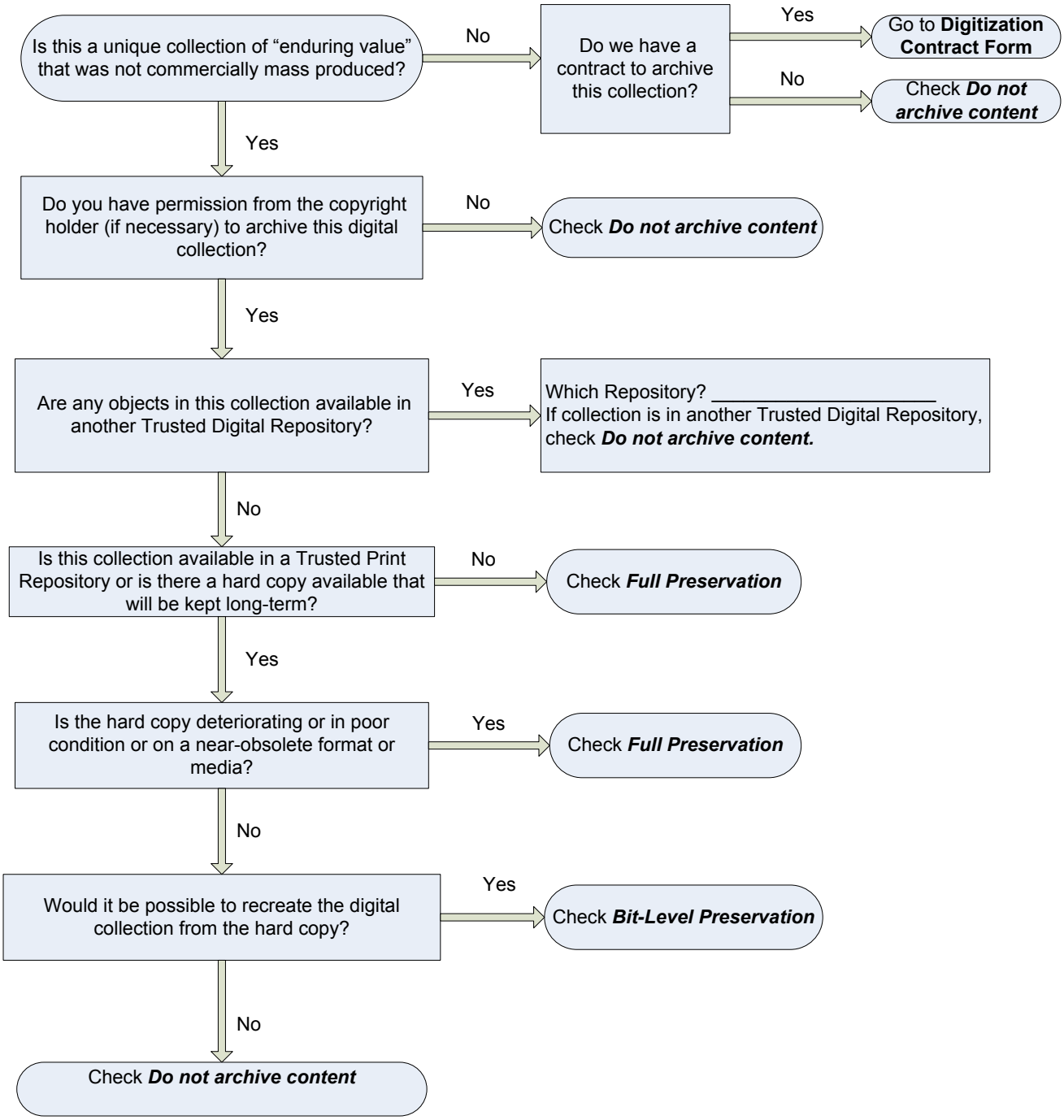
1. Technical Metadata: Documents and supports formats, access rights, etc., records changes to records, supports future actions, etc. Must include:
    a. Environment: What is needed to render, view, and use the digital materials
    b. Authenticity: Is the file what it purports to be
    c. Formats: What type of information is in the file and how is it encoded
    d. Rights: How is access controlled, and who has authority to use the file
2. Structural Metadata: How the files relate to one another/to files outside the Archival Information Packet (AIP). Includes how the records are organized, and how parts (files, folders) relate to the whole (folders, entire collection).

d. Finalized Archival Information Packet (AIP) includes:
    i. Folder containing preservation copy (files as accessioned)
    ii. Documentation
    iii. Manifests
    iv. Processed copies
    v. Explicit link to descriptive/preservation metadata
    vi. Explicit link to content.
e. Create discovery tool (finding aid) (see Reuther Processing Manual)
f. Transfer to storage, discovery, access infrastructure

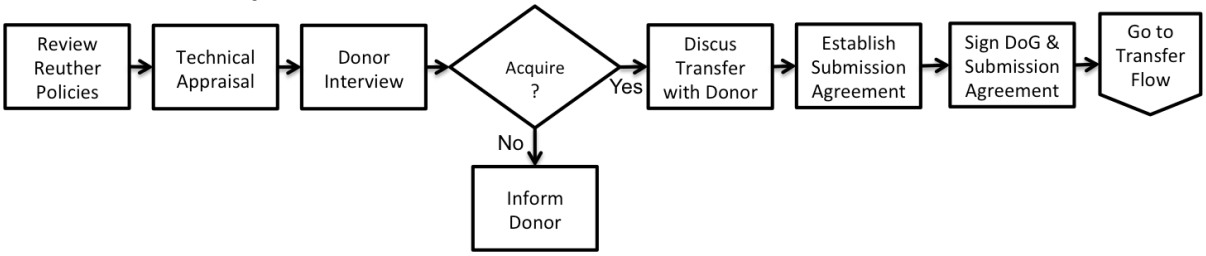
5. Access (Access Flowchart) (and Dissemination Information Packet (DIP) creation, if needed)
    a. Determine if it is necessary to create a Dissemination Information Packet (DIP) for access (i.e., Archival Information Packet (AIP) is too large for direct access), or if users can directly access the Archival Information Packet (AIP). If it is necessary to create a Dissemination Information Packet (DIP):
        i. Use the Archival Information Packet (AIP) to create the Dissemination Information Packet (DIP).
            1. Both the digital archival materials and the descriptive metadata are needed to make the content findable.
            2. May not contain the complete Preservation Description Information (PDI) from the Archival Information Packet (AIP) because not all metadata is needed for accessibility.
            3. Distinguish as the requested information
    b. Access to the digital files is via the Digital Object Management system (DOM) in Fedora Commons.

APPENDIX

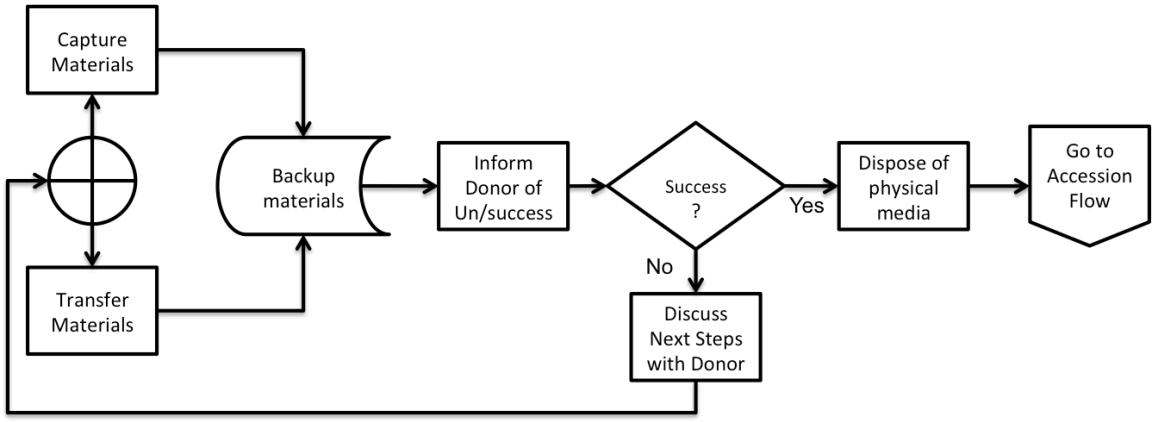
Digital Preservation Decision Flowchart



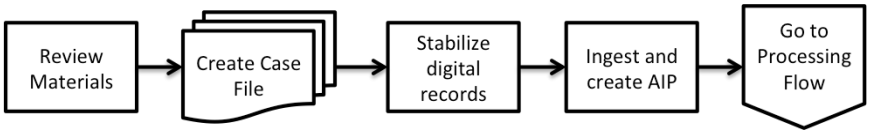
Selection and Acquisition Flowchart



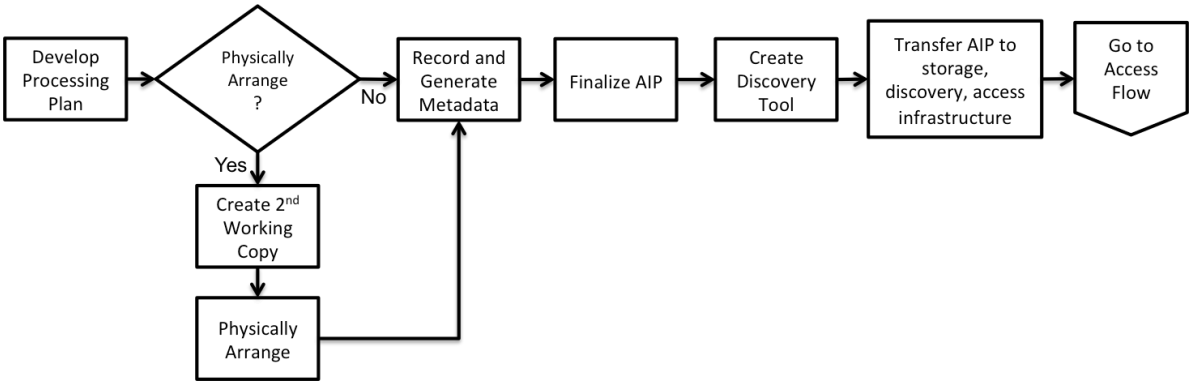
Transfer Flowchart



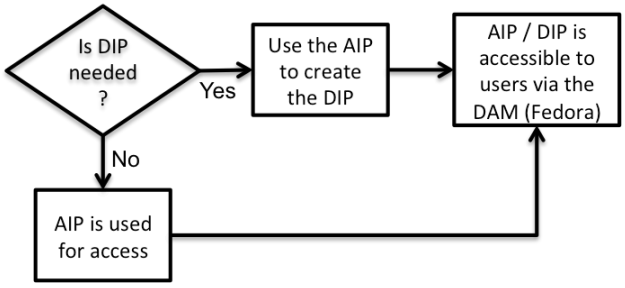
Accession Flowchart



Process Flowchart



Access Flowchart



Supported formats

The Reuther Library accepts the following file formats:

Audio
- avi
- mp3
- mpg
- wav

Excel
- xls
- xlsx

Images
- gif
- jpg
- png
- tiff

PDF
- pdf
- pdf/a

PowerPoint
- ppt
- pptx

Text
- doc
- docx
- rtf
- txt

Although the Reuther may ingest and preserve formats not on this list, long-term preservation cannot be ensured because of accessibility and migration issues. Content and file/folder sizes should be normalized into one of these listed file formats prior to preservation.

Acceptable Transfer and Physical Media Capture Methods

The Reuther Library uses the following transfer and physical media and format capture methods for digital material transfer and deposit:

Transfer Methods
- Dropbox
- SFTP
- WebDAV
- SSH
- Rsync

Capture Methods: Current Physical Media
- External hard drives
- Internal drives

Capture Methods: Obsolete Physical Media
- Zip
- FDDs
- CD-ROM

Although the Reuther Library may accept transfer and capture of materials using physical media and formats not on this list, doing so requires a conversation with Reuther staff.

Submission Information Packet (SIP) Levels

- SIP Level 1 (Minimum): Digital materials are on physical media, and transfer documentation is provided. No formal submission agreement no unique identifiers per file, does include collection and creator information, accession identifiers, donation agreements, and possibly access restrictions.
- SIP Level 2 (Preferred): Files are on physical media or transferred electronically with a manifest, fixity data, and full submission agreement. The producer uses software that identifies each individual file, its extent checksum, and relationship to the whole. The Reuther is able to validate this information after transfer (i.e., verify the checksum), and able to prepare unique identifiers. Descriptive metadata may be included, although it may not be standardized and mapping to the files is not necessarily automated.
- SIP Level 3 (Ideal): Producer transfers METS, including an XML file (content and metadata together in one file, includes unique identifiers per file, makes the package interoperable with other preservation systems).

Checklist for Submission and Archival Information Packets

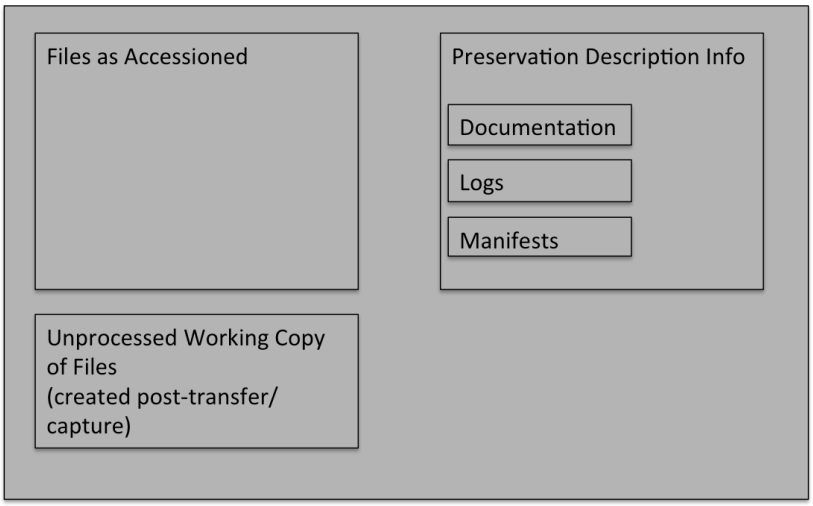
- Structure and contents for each SIP prior to transfer:
    - Unique ID (link to descriptive record)
        - Preservation Description information
            - Documentation (deed of gift, transfer correspondence, submission agreement, etc.)
            - Logs
            - Manifests
        - Preservation Locked (files as accessioned)

- Initial setup structure and contents after copying the SIP and transitioning it to an AIP while working:
    - Unique ID (link to descriptive record)
        - Preservation Description information
            - Documentation (deed of gift, transfer correspondence, submission agreement, etc.)
            - Logs
            - Manifests (checksums, file characterization data)
        - Preservation Locked (files as accessioned)
        - Unprocessed Working (AIP in progress)

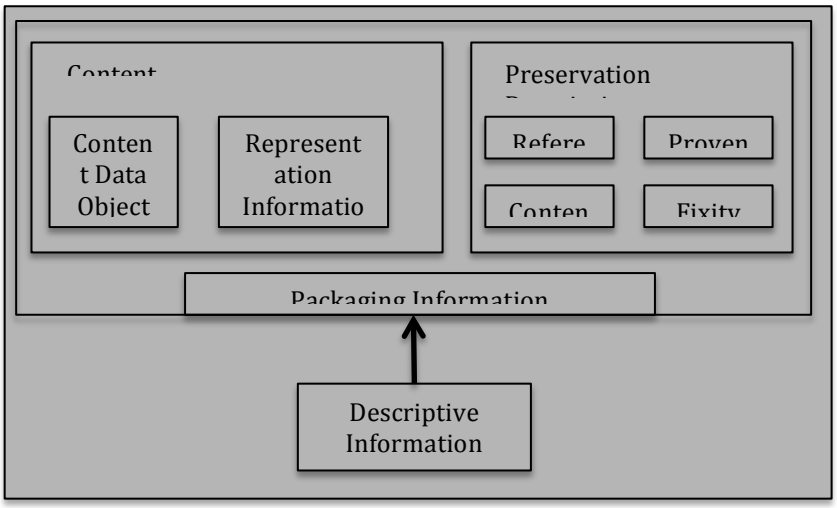
- Structure and contents for each AIP completed:
    - Unique ID (link to descriptive record)
        - Preservation Description Information
            - Documentation (deed of gift, transfer correspondence, submission agreement, etc.)
            - Logs
            - Manifests (checksums, file characterization data)
        - Preservation Locked (files as accessioned)

- Processed Access Copies (files after processing)
  - Nearline Files (Access copy, materials with potential privacy, rights, ethical concerns, available for access only in the Reuther Library)
  - Online Files (Access copy, all other materials)

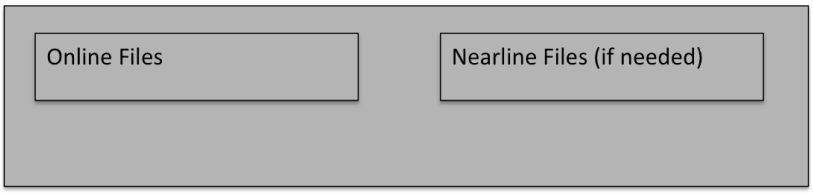
Submission Information Packet



Archival Information Packet



Dissemination Information Packet



Digital Preservation Support Policy[1]

This is intended for use by the **preservation archivist** and **digitization archivists** to assist with workflows, as well as in donor negotiations.

- Level 1 (Ideal): Preservation is the highest available at the Reuther, and ensures that the original digital object is viable, renderable, understandable and functional.

  Determining criteria: The file format is publically documented, widely adopted, supported and rendered by numerous software packages, has lossless data compression, and contains no embedded files or dynamic content.

- Level 2 (Preferred): Preservation ensures that the original digital object is viable, renderable, and understandable (although not necessarily functional).

  Determining criteria: The file format is publically documented and is either in a lossy data compression or is a version that has been deprecated. Or the file format is proprietary, widely adopted and migration tools will likely be available.

- Level 3 (Basic): Preservation only applies to original data object (i.e., bit stream).

  Determining criteria: May not be able to preserve long-term preservation due to external factors (e.g., file format is proprietary, scarcely adopted, or supported by few software formats) or internal factors (does not meet level 1 or 2 criteria)

Metadata Guidelines

Digitization archivists initially add metadata after transfer or capture and again during processing. **Digitization archivists** may subsequently edit metadata after an appropriate event (accrual, reprocess, correction, deaccession, etc.).

- Administrative Metadata: Case file information. This type of metadata is often updated by simply adding additional files as needed in the digital case file folder.
- Descriptive Metadata: Describes the content of the records and their context of creation. Record descriptive metadata via digital files and forms: accession record, transfer form, manifest, log(s), as well as in the discovery tool(s) (i.e., finding aid). Follow DACS and the Reuther Processing Manual for guidance on adding and editing descriptive metadata.
- Preservation Metadata: Structural and technical metadata
    - Structural Metadata: Records how the records are organized, how each file relates to the whole, and the file systems (allocation and access tables) track it. **Digitization archivists** need to record any updates to structural metadata (e.g., when processing indicate any changes made to the file organization or arrangement).

[1] Based on the University of Illinois at Urbana-Champaign IDEALS Digital Preservation Support Policy

- Technical Metadata: Records the environment, authenticity, format, and rights information about materials. Stored in the Preservation Description Information (PDI). Establish a file folder manifest for all copies of the material (i.e., preservation, and various access copies). **Digitization archivists** must record any updates to technical metadata.

The recommended standard for descriptive metadata is METS, generated with the assistance of software tools. Additionally, EAD can still be used for descriptive metadata as well as administrative and structural metadata as EAD can be wrapped in the METS framework. PREMIS XML schemas should be reviewed for the creation of preservation metadata.

Access and User Expectations

The Reuther Library strives to exceed user expectations wherever possible. In that spirit, the following guidelines will be followed pertaining to user access of digital materials:

- When not prohibitive by copyright or constrained by sensitive information, digital materials will be made accessible to users via the Digital Object Management system (DOM- Fedora Commons), on the web. For digital materials where copyright or sensitive information is a concern, nearline access will be provided, meaning users may access the content only by visiting the Reuther Library.
- Users will have the ability to copy, paste, and print content available via the Digital Object Management system (DOM – Fedora Commons)
- When possible content will be available in the original format, however, some content will be normalized into a different format than the original to enable preservation and access. Additionally, some normalized will be constrained to only be renderable and viewable rather than useable and interactable.

Impact on Existing Reuther Policy

The Reuther Library's Digital Preservation Policy Program impacts existing Reuther policy, the most significant of which is described below:

Selection and acquisition of archival materials is impacted as the Reuther Library actively seeks collections that are both whole and hybrid.

- Whole collections: The Reuther Library will accept collections entirely composed of digital materials.
- Partial collections: The Reuther Library will also accept collections partially compose of digital materials, i.e., hybrid collections that are partially paper and partially digital materials.

The concept of access shifts regarding digital materials, as all digital materials will have multiple copies to address access purposes as well as long-term preservation storage (dark archive) needs.

- Access copies: A copy of the digital materials as accessioned (pre-processed) will be preserved, although not accessible to users. The processed version of the

materials will be made accessible, generally online, although some materials with potential privacy, rights, or ethical concerns may be made available via nearline access (only available within the Reuther Library).
- Dark Archive: All Submission Information Packets (SIPs), Archival Information Packets (AIPs), and Dissemination Information Packets (DIPs) will be permanently backed up into a "dark archive" where they will remain permanently (with the exception of being updated periodically to ensure accessibility by migration).

Deaccessioning digital materials introduces additional concerns.

- Physical storage media used to donate or transfer digital materials will be deaccessioned and disposed of after the digital materials are successfully copied from the media onto the Reuther Library's storage devices.
- Digital materials are subject to the same deaccessioning policies as paper materials. See the Reuther Collection Policy for further details.
- Digital materials are subject to additional deaccessioning criteria, and may be deaccessioned based on: being in an unaccepted, non-normalizable, unreadable file format; infection with a virus the Reuther Library is unable to eradicate, containing sensitive or copyrighted material, etc.

Related Reuther Documents

- Reuther Processing Manual [Link]
- Reuther Library Collection Policy
- Donor Organization Digital Deposit Guidelines [To be created: Would be a tailored list to each of the Reuther's donor organizations, explaining what files should be retained in what formats, how transferred, etc.]
- Reuther Library Digital Preservation Security and Emergency Planning [To be created: Would list security measures, emergency planning, current platforms like the DOM and digital "dark" archive]

Terminology Glossary

AIP	Archival Information Packet
DIP	Dissemination Information Packet
MD5	MD5 Message-Digest Algorithm, cryptographic hash function used to ensure fixity but not security
METS	Metadata Encoding and Transmission Standard
OAIS	Open Archival Information System
PAIMAS	Producer-Archive Interface Methodology Abstract Standard
PDI	Preservation Description Information
SIP	Submission Information Packet

Reuther Digital Curation Workflow