Processing Digital Records for Access and Preservation
What To Do and When To Do It!

Chris Prom
@chrisprom
prom@illinois.edu
Overview

- How to Form an Archival Information Packet and Tools To Do So
- Establishing Processing Workflows and Planning for Success
- Actually Processing the Records (Nuts and Bolts)

Credits: Portions of this workshop drawn from Digital POWRR and from SAA Course “Arranging and Describing Digital Records”
## Processing for Preservation and Access

### Long term access (Preservation)

- **Purpose**: ensure long-term access
- **Focus**: current & future users
- Relies on **proven (reliable)** technologies to preserve digital objects across generations of technology
- **Accumulates** metadata over the life cycle to trace preserved content
- Preservation systems **create** new versions of digital objects for access to deliver as needs change over time

### Short term access

- **Purpose**: provide content to users now
- **Focus**: current
- Relies on **cutting edge** technologies to provide best and fastest access at a point in time
- **Selects** metadata needed to use and understand content
- Access systems **deliver** objects with user-oriented services
Icebreaker Time!

NDSA Levels of Preservation

Where can my organization place its chips?

- Small Groups – Where do you think you fit in?
- All Together – Poll of who is where!
<table>
<thead>
<tr>
<th>NDSA</th>
<th>Table 1: Version 1 of the Levels of Digital Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level 1 (Protect your data)</td>
</tr>
<tr>
<td></td>
<td>- Two complete copies that are not collocated</td>
</tr>
<tr>
<td></td>
<td>- For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system</td>
</tr>
<tr>
<td></td>
<td>File Fixity and Data Integrity</td>
</tr>
<tr>
<td></td>
<td>- Create fixity info if it wasn’t provided with the content</td>
</tr>
<tr>
<td></td>
<td>- Virus-check high risk content</td>
</tr>
<tr>
<td></td>
<td>Information Security</td>
</tr>
<tr>
<td></td>
<td>- Restrict who has those authorizations to individual files</td>
</tr>
<tr>
<td></td>
<td>Metadata</td>
</tr>
<tr>
<td></td>
<td>- Ensure backup and non-collocation of inventory</td>
</tr>
<tr>
<td></td>
<td>File Formats</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[http://www.digitalpreservation.gov/ndsaint/activities/levels.html](http://www.digitalpreservation.gov/ndsaint/activities/levels.html)
Session One:
The AIP . . . And How to Form One
Processing Requirements I

Basic Policy Framework (i.e. a Program Statement)

- level of preservation
- Partners
- Documentation/Collection Foci
- Guiding Values/Commitments
- Acquisition Methods
- Processing Commitments
- Management/Storage
- Preservation Actions
- Access/Use
Processing Requirements II

- Hardware Infrastructure
  - Capture mechanisms
  - Working space
  - Processing workstation
  - Adequate storage space, preferably disk based and redundant (TRAC Guidelines; [www.crl.edu/PDF/trac.pdf](http://www.crl.edu/PDF/trac.pdf))
Hardware Infrastructure

Working Space (internal)

Processing Workstation

STORAGE SPACE (external), ideally 3 copies with preservation monitoring
Processing Requirements III

- Software Infrastructure
  - Capture Tools
  - Appraisal Tools
  - Arrangement Tools
  - Description Tools
  - Monitoring/Preservation Tools
  - Access Tools
Processing Requirements IV

Accessioning procedures must:

- Transfer/copy materials
- Provide ownership, legal custody, right to manage
- Gather contextual information
- Identify conservation Issues
  - Virus checking; format analysis
- Achieve basic control
  - Stabilize records
  - Identifies formats, extent, structure and ‘fixity’ information
  - Achieve rudimentary intellectual control
- **DO NO HARM | DO NO HARM | DO NO HARM**
Processing Requirements VI

Processing workflow that:

- Defines processing plans/commitments*
  - Establish intellectual arrangement and relationships
  - Develop access plan/transformations

- Provides you with tools to:
  - Arrange records
  - Describe records at whatever level you choose
  - Transfer them to storage, discovery, access infrastructure

- Preserves materials and makes them accessible.

* Processing procedures must formalize, repurpose, or build upon good accessioning actions | Bad accessioning = bad A and D
Processing Requirements Summary

Essential Tasks: Accessioning
- Transfer/copy
- Secure legal custody
- Gather contextual info
- ID conservation issues
- Stabilize records
  - Identify formats, extent, structure
  - Record checksums
  - Achieve rudimentary intellectual control (prepare SIP)

Essential Tasks: Processing
- Develop processing plans
  - Establish ‘intellectual arrangement’
  - Note relationships
- ‘Physically’ arrange records
  - Identify and record preservation and access issues
  - Organize records
- Describe records
- Transfer to storage, discovery, access infrastructure
Good Arrangement and Description Depend upon:

- Inventorying
- Transfer method
- Storage ‘solution’ (however imperfect)
- Descriptive system
- Preservation policies and planning
- Access tools
  — see Ben Goldman, *Bridging the Gap*
“Access drives preservation”
— Brewster Kahle, UNESCO Memory of World Conference, 2012.
Discuss with Partner

What software tools do you think you could or should use to survey, appraise, arrange, and describe digital records?

Given your resources, where could or should you store processed copies of the records?
Archival Information Packet

- Provide overview of OAIS Reference Model’s AIP
- How it relates to archival arrangement and description
  - Making sense of OAIS concepts
  - Map them to traditional archival tasks/functions
Where Arrangement and Description ‘Sit’ in OAIS terms

A and D initiated during accession and processing

Descriptive records provided to users

Archives provided to users
In OAIS terms, archival processing = the action taken that transforms records from the state in which they are received from producer to the state in which they are stored and (potentially) disseminated:

- Generation of Submission Information Packet (SIP) during accessioning
- Transformation into Archival Information Packet (AIP)
- Requires:
  - Generation of preservation metadata
  - Generation of descriptive metadata
  - Arrangement and preparation of dissemination copies
  - Storage and monitoring of AIP
Archival Information Package


Record: [http://www2.archivists.org/glossary/terms/r/record](http://www2.archivists.org/glossary/terms/r/record)
**AIP Part One: Content Information**

**Record:** – 2. Data or information that has been fixed on some medium; that has content, context, and structure; and that is used as an extension of human memory or to demonstrate accountability.

---

**Diagram:**

- **AIP**
  - **Content Information**
    - Content Data Object
    - Representation Information
  - **Preservation Description Information**
    - Reference
    - Provenance
    - Context
    - Fixity
  - **Packaging Information**
  - **Descriptive Information**

---

Record: [http://www2.archivists.org/glossary/terms/r/record](http://www2.archivists.org/glossary/terms/r/record)
Content Information I

Content Data Object

- Focus of preservation **and** the object of an archival descriptive record
- Can be simple or complex
- At any level of description
- First responsibility: to preserve bitstream
- May or may not be ‘original’ files
- Typically, metadata about the object is NOT stored with file.
Any content data object is just a discrete sequence of bits

“FILE” or other content data object
Content Data Object Examples

- Single file (photograph, email, etc)
- Series of **homogeneous** files, e.g. photographs, documents, spreadsheets
- Assemblage of **heterogeneous** files, organized into folder, *relating to a specific* function, activity, or records creator.

**Key Point**: The AIP can be formed at any level of arrangement or description (in archival terms).
Content Information II

- Representation Information
  - Must be preserved over time
  - Information necessary to render and make the content data visible
  - Simple example: extension on end of a file
    - .pdf
    - .doc
    - .wpd
  - Better example: actual file type based on information from header
AIP Part Two: Preservation Description Information (PDI)

AIP

Content Information

- Content Data Object
- Representation Information

Preservation Description Information

- Reference
- Provenance
- Context
- Fixity

Packaging Information

Descriptive Information
Preservation Description Information (PDI)

- Must also be preserved over time.
- Documents:
  - Information about the object(s) preserved
  - The preservation environment and practices
  - System management data
- Shows the trustworthiness of repository
- AKA Preservation Metadata
Fixity Information

- Fixed = unchanged over time
- Tracked with “Checksums”
- Why bother: used to validate the authenticity and integrity of a content object (for example, a file or files);
- Typically stored in a manifest (list or xml file)
- **Cannot** be easily tracked for file names, timestamps, file locations, etc.
Checksum Example
Preservation Description Information (PDI)

- **Context Information**
  - Relationships to other information objects
  - Either internal (e.g. files relationship to each other; logical arrangement on disk)
  - or to external objects
  - Can be represented via:
    - System of arrangement note (DACS 3.2)
    - Nested levels in finding aid
    - Directory structure on file system
    - in tree (xml or otherwise)
Preservation Description Information (PDI)

- **Provenance**
  - History of materials; chain of custody
    - Typically record in DACS 5.1 - 5.4 (custodial history, source of acquisitions, appraisal note, accruals)
  - Preservation actions such as migrations
    - No DACS element specifically for this, but should use a Note element (DACS 7)
    - Can store as a PREMIS event (covered tomorrow)
Preservation Description Information (PDI)

- Reference Information*
  - Unique ID (DACS 2.1)
  - Be mindful of sharing context
    - Local, National, Global
  - Ties content object to descriptive and other metadata
  - Can be recorded at any appropriate level of description

*Often considered the first and most important element of PDI.
NOTE: Descriptive Information is functionally part of an AIP; however it is typically NOT STORED with it.
AIP Part Four: Packaging Information

Content Information
- Content Data Object
- Representation Information

Preservation Description Information
- Reference
- Provenance
- Context
- Fixity

Packaging Information

Descriptive Information
Packaging Information

- Ties the Content Information and PDI together
- Generally takes place at a system design and file system level.
- Example using file system and archival catalog:
  - *Content information* stored in folder with a particular unique ID.
    - One possibility: the folder holds files and xml file containing PDI
  - Descriptive metadata and some administrative preservation in catalog record with same unique ID.
Simple Example: Packaging Information

Package ID: ER-0001

AIP

Content Information
- Content Data Object
- Representation Information

Preservation Description Information
- Reference
- Provenance
- Context
- Fixity

Packaging Information

Linked via ID or system relationship

Descriptive Information

Record in descriptive system: ER-0001
The AIP

BEFORE you start processing records

- The AIP is a logical package
  - Representation Information, PDI and Descriptive information (metadata) may or may not be stored with the content information
- Important point is that everything must be packaged together somehow
- It’s up to you: it is an implementation decision that should be undertaken at a system design level.
Simple AIP using archival descriptive record and file system storage

Catalog Record: ER-0001

ID link forms implicit package

Description in Archival Information Management System

File System Hierarchy
AIP Structure is not Prescriptive

- Internal structure—many options, many systems
  - But need to account for all the parts of the packet as just described.
- Relationship of AIPs to parent resource and each other might be:
  - One AIP per accession/collection
  - Multiple AIP’s per collection with links
  - Other options
Basic Tools and Making an AIP
Three basic types are needed
- PDI generation tools (make a manifest)
- Archival descriptive tools (describe content)
- Data packaging tools (tie it together)

With these, you can:
- Meet Functional Requirements, and
- Generate Archival Information Packages that hold records in a way that maintains authenticity and integrity
Let’s Talk About Tools….

There are front-end/processing tools like…..

Archivematica
Curator’s Workbench
Data Accessioner
BitCurator

And there are back-end storage/preservation services like…..

MetaArchive
DuraCloud
Amazon Glacier
Fixity
Internet Archive

There are even some services that will pretty much do it all like…..

Preservica
Dsase Direct (uses DuraCloud)
ArchivesDIRECT

*Tools/Services in RED were tested in-depth by POWRR

Note: Yes, there are also CMS’s, IR software, ....ugh. However, these are outside the scope of this workshop!

Technical skill available + amount of annual funding devoted to DP = range of tools you will be considering
# POWRR Grid

<table>
<thead>
<tr>
<th>Tool</th>
<th>Ingest</th>
<th>Processing</th>
<th>Access</th>
<th>Storage</th>
<th>Maintenance</th>
<th>Other</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>DataAccessioner</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Free</td>
</tr>
<tr>
<td>BitCurator</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Free</td>
</tr>
<tr>
<td>Archivematica</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Free</td>
</tr>
<tr>
<td>Fixity</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DuraCloud</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>Varies</td>
</tr>
<tr>
<td>MetaArchive</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>Varies</td>
</tr>
<tr>
<td>ArchivesDirect</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Varies</td>
</tr>
<tr>
<td>Preservica</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>Varies</td>
</tr>
<tr>
<td>Glacier (Amazon)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Varies</td>
</tr>
<tr>
<td>Internet Archive</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>Free</td>
</tr>
</tbody>
</table>
Location where accession will be copied TO

Location where files will be copied FROM

Record Basic DC Metadata

Click this button to move files, run analysis, and generate XML file
But . . .

What if you don’t have access to any of these?

Do you really need them???
DIY Archival Information Packet

Catalog Record for ER-0001

- Content Information
  - Data object(s)
  - Rep. Information
- Preservation Description Information
  - Reference
  - Fixity
  - Context
  - Provenance

folder structure:
- ER-0001
  - pres_desc_info
  - documentation
  - manifests
  - preservation_files_locked
  - processed_access_copies
  - nearline_files
  - online_files
AIP Structure in File System

- **Unique ID, link to Descriptive Record**
- **Checksums, File Characterization Data**
- **Files as accessioned**
- **Records of system actions, e.g. migration, renaming, etc.**
- **Deeds of Gift, transfer correspondence, etc; can be attached to descriptive record in some archival CMS applications**
- **Files after processing—on web server or in storage ready for dissemination**
Exercise One: Creating Preservation Metadata and Gaining Basic Control

- Use the Leon Dash folder in your “Workshop/1_Unprocessed_Storage” folder
  - Establish AIP structure in the Working Space
  - Copy/move files into your “preservation” folders
  - Use software to generate a manifest for preservation including checksums
  - Put the manifest file in “pres_desc_info” folder
  - Files are now stabilized and you have ‘just enough’ preservation metadata to accession and process them.

- If time, repeat with the UC2B Folder
Step One:

A: Establish AIP Structure

B: Move Files into Preservation Locked Folder
**Step Two:** Generate Preservation Metadata and Save to Manifests Folder

1. **Add Folder to analyze**
2. **Select Items for which you want PDI (typically all of them)**
3. **Save as CSV to Manifests Folder in your AIP**

---

### HashMyFiles

<table>
<thead>
<tr>
<th>Filename</th>
<th>MD5</th>
<th>SHA1</th>
<th>CRC32</th>
<th>SHA-256</th>
</tr>
</thead>
<tbody>
<tr>
<td>DashConvoFINAL.doc</td>
<td>36256094365418e07c5fd6bda1a59506</td>
<td>8c7f3b6b642d59802e2e089db3e6fb5895ae0</td>
<td>0a196c44</td>
<td></td>
</tr>
<tr>
<td>DashConvoFINAL.pdf</td>
<td>c123b2627a19889354b9f364e8f19c9</td>
<td>66bc3f926f3f19bba3ed2ccbd69958</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leon Dash Transcript.pdf</td>
<td>0a0464676842eb6983d8369842</td>
<td>4d8ab3d5ebf84eb2b07097c03c408468051374350a</td>
<td>0a45fd9c1b</td>
<td></td>
</tr>
<tr>
<td>ReadMe.txt</td>
<td>7e65ad6318441ec2ad7d9348c9493e0134d</td>
<td>755f7c71c72b8fd06c6ad2c7d29d8e63b969ed0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WinterMichael...-- 3-27...</td>
<td>032584a110235135d636a8d9c95998b4</td>
<td>032584a110235135d636a8d9c95998b4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WinterMichael...-- 3-27...</td>
<td>3c54b4ae1c126e6530b3b7ec15f1277f</td>
<td>125a19ed6690dca8e9eef3eaf38c4531d5</td>
<td>7c451564</td>
<td></td>
</tr>
<tr>
<td>BettsHenryB...-- 8-14-0...</td>
<td>11049907b0d8d674f954b35c39b2f7</td>
<td>d06938569c00d7c123f64e4b3e00360f7</td>
<td>3e7b0a58</td>
<td></td>
</tr>
<tr>
<td>BettsHenry...B-- 8-14-0...</td>
<td>8031a2139b8fa299b262a80152f1b010</td>
<td>00b2c3e2c1702d6a4f6f6368317152c593d43</td>
<td>65a5ce70</td>
<td></td>
</tr>
<tr>
<td>LeonSteurevaldwON...</td>
<td>148df363282a4e62be4678194900c2</td>
<td>1f815f7c71d54091a7d0a817dc115b2aa1cfa6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LeonSteurevaldwON...</td>
<td>362827576d47a2770809090cc51538</td>
<td>96fd477a3f2d8e2b0f9e355cd6c761964b</td>
<td>376e2110</td>
<td></td>
</tr>
<tr>
<td>Nugent Interview Not.</td>
<td>c86a3b9d8a7f7a07d7e3a73907a4aa4</td>
<td>314f4db3a870cecb29dd9b0214f818e6f080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nugent Interview Not.</td>
<td>71a7e6b358191563a3ce68d38f8eae</td>
<td>2f2b877bab91aa83e94f18a62011b34f99797</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatriciaSteurevaldwON...</td>
<td>148df363282a4e62be4678194900c2</td>
<td>3f815f7c71d54091a7d0a817dc115b2aa1cfa6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatriciaSteurevaldwON...</td>
<td>362827576d47a2770809090cc51538</td>
<td>96fd477a3f2d8e2b0f9e355cd6c761964b</td>
<td>376e2110</td>
<td></td>
</tr>
<tr>
<td>RobertsZona...-- 9-25-0...</td>
<td>3dc0c5ba9af1d8e8e2d68bb83b3e8b43f</td>
<td>801f8d68e1e82868684b7eb3e4b0e86</td>
<td>0a196c44</td>
<td></td>
</tr>
<tr>
<td>RobertsZona...-- 9-25-0...</td>
<td>1db8a94faca3a51ff4af52ad0f72b08</td>
<td>65a5ce70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>StallmanWilliamE--1...</td>
<td>6a236d45636efbced9308ac17b25c8</td>
<td>53822657a2b7375fe0af6b3f2d0e629e5c2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>StallmanWilliamE--1...</td>
<td>065252276a2b7375fe0af6b3f2d0e629e5c2</td>
<td>e53822657a2b7375fe0af6b3f2d0e629e5c2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SteuerwaldPatriciaNu...</td>
<td>1b0d662810f42a60f7989799b49c0b5</td>
<td>e08724a2f79e691da780f717c9a11c591e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SteuerwaldPatriciaNu...</td>
<td>74ac0f3b1d033c2ec700e6fe1997e</td>
<td>1675bccc050073b083e5799933726de66e0c19</td>
<td>148769e9</td>
<td></td>
</tr>
<tr>
<td>DriscoJean Outside - S...</td>
<td>999853bc02e9925992979e6d9e5a97b8d</td>
<td>05a2409b996d6992b175b5992596d9e5a97b8d</td>
<td>148769e9</td>
<td></td>
</tr>
<tr>
<td>DriscoJean Outside - S...</td>
<td>0d024d1a4c90d9686271c7f2bd3bf87</td>
<td>05a2409b996d6992b175b5992596d9e5a97b8d</td>
<td>148769e9</td>
<td></td>
</tr>
<tr>
<td>DriscoJean School - S...</td>
<td>4cae820e8b906ba4bca5761081b39815</td>
<td>05a2409b996d6992b175b5992596d9e5a97b8d</td>
<td>148769e9</td>
<td></td>
</tr>
</tbody>
</table>

Total: 116 file(s), 116 selected
Exercise Two: Creating Preservation Metadata and Gaining Basic Control

- Using the Leon Dash folder in your “Unprocessed Storage” folder
  - Establish AIP structure in the Working Space
  - Copy/move files into your “preservation” folders
  - Use software to generate a manifest for preservation including checksums
  - Put the manifest file in “pres_desc_info” folder
  - Files are now stabilized and you have ‘just enough’ preservation metadata to accession and process them.
Follow Up Discussion

With your partner, examine the output of HashMyFiles. List the types of metadata you see there (hint: map to the AIP elements!)
Session Two:

Establishing a Processing Workflow and Planning for Success
Actual Conversation, ca. 2004

“I’d like our institution to be the home for your papers.”

*gets handed flash drive*
Don’t Panic - Your Pre-Ingest Workflow
aka . . . Wrangling your digital stuff before you can get it under control and eventually into a shiny system

NOTE: This is NO ONE WAY TO DO THIS! … Everyone’s workflow is a little different!

Starting from scratch:

• Run accessioning tools (creates basic preservation metadata files in XML for you!)
  - Move everything to a stable carrier (like a network drive)
• Begin an Inventory Spreadsheet
• Make an Access Copy from your Master Copy
• Continue populating Inventory Spreadsheet (if needed)
• OPTIONAL: Keep original media

✓ Most of these will cost you more time than money
✓ Document what you do pre-ingest. For future you.
✓ Remember: Good enough is just fine. For now.

Yes, it can be as simple as creating a spreadsheet! . . . or a google form!
Pre-Ingest Inventory Spreadsheet Categories

These suggestions follow the recommended DPOE step “Identify” as locally defined by curator/archivist. Example at: [http://www.carli.illinois.edu/sites/files/digital_collections/documentation/digipres_identify.pdf](http://www.carli.illinois.edu/sites/files/digital_collections/documentation/digipres_identify.pdf)

- Category (digitization project; born digital; university archives)
- Title and Description
- Date(s) (date range of what’s IN there or date of creation if born digital)
- Location (CD, Jump drive, server location?)
- Extent (quantity: 48 journal issues; 106 images; 2 TB of video)
- Format (file formats: PDF, .Jpeg, Animated GIF, Wordstar2.0 file)

This is YOUR inventory... YOU get to decide if it needs additional fields, if some can be deleted, etc. You are the boss of this!

FILL OUT WHAT YOU CAN AS YOU WOULD WITH ANY NORMAL ACCESSION
Accessioning Actions:

- Capturing records in authentic fashion
- Gaining custody (physical and legal)
- Undertaking a conservation assessment
- Recording basic technical metadata ***
- Virus Check
- Gathering and recording basic descriptive information regarding **content and context of creation** for the accessioned files ***
- Begin forming archival information packet ***

*** Essential actions related to A & D
Preservation description information may include materials similar to that of the typical “acquisition case file,” such as:

- Transfer paperwork/surveys
- Deeds of Gift
- Correspondence relating to an acquisition
- Notes
- Anything else of relevance

These materials were put a ‘documentation’ folder (not necessarily provided to end user), since they may be essential to management of the materials.
Exercise Three: Accession and Prep to Process

■ Step One: Make Copy of content in Access folder (this will be your working copy moving forward)
■ Copy documentation into the documentation folder
■ Step Three: Accession the Records you just transferred to storage

[https://goo.gl/forms/itEmQGClF7Ve8huj2](https://goo.gl/forms/itEmQGClF7Ve8huj2)

○ Skip over processing plan and description sections (for now, we’ll come back to them)
Step One and Two:

Copy content to access folder (this will become your working copy)

B: Move Accession information into doc folder (this will become preservation metadata and you will use it in next step)
Step Three: Make an accession record entry

After you complete this page, just click thru following screens for now; you will be emailed a link to edit the form.
Arrangement and Description of Digital Records and Manuscripts: Key Messages

- Is collaborative
- Begins before records are received
- Implements same ‘business processes’ employed for analog materials
- Methods, procedures, techniques are different
- Is iterative
  - “Every inventory is preliminary”—Maynard Brichford
Arrangement and Description Workflow

- A & D tasks place at many stages in an archival processing workflow:
  - Pre-custodial work
  - Accessioning
  - Providing an Arrangement Scheme/Intellectual Arrangement
  - Arranging the Records
  - Describing Records (includes new types of data)
  - Creating Access
  
  * from Daines, *Processing Digital Records and Manuscripts*

- See sample processing workflow on handout
  - We will use a modified version of this workflow through the course.
Arrangement and Description: Six Steps to Success

- Conduct Research
- Survey Collection
- Create Plan
- Arrange Materials
- Describe Materials
- Provide Access
Step One

Conduct Research

Survey Collection

Create Plan

Arrange Materials

Describe Materials

Provide Access
First Step  Gather Contextual Information and Documentation

- Determine extent and record it
- List inclusive dates, dates created, last modified (range if necessary)
- Examine accession documentation
- Follow up with creator/donor as necessary re: bibliographical information, recordkeeping, functions of records, how generated, etc.
- Record your takeaways in case file (includes deed of gift, correspondence, survey forms, other forms, etc. becomes part of AIP)

Remember:
Use the working copy of files for collection analysis!
File Allocation Table Example (and a word about Dates)

MS/DOS Directory Entries

<table>
<thead>
<tr>
<th>FileName.Ext</th>
<th>Autoexec.bat</th>
<th>Scheduler.cc</th>
<th>DoomII.exe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date/Time</td>
<td>01Mar97/12:01:00</td>
<td>08Apr92/06:22:33</td>
<td>28May90/22:10:40</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start Block</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

File Access Table (FAT)

Step Two and Three

Conduct Research

Survey Collection

Create Plan

Arrange Materials

Describe Materials

Provide Access
Plan to Process

- Set strategy to appraise files
- Review files and record appropriate notes regarding structure, current arrangement and description, preservation issues and actions (see template processing plan in google form you’ll be using soon 😊)
- Note expected restrictions
- Propose an arrangement
- Determine appropriate Level of Description
File Browsing and Properties
Tree Size Pro
Your Arrangement Ideas?
Exercise Four: Develop a Processing Plan
Step One

Use file manager to browse through access copies folder and run a few tree size pro analyses of it.
Step Two:

Fill out Processing Plan section of Google Form (edit your existing record)
Session Three:

Actually Processing the Records (A &D Nuts and Bolts)
Step Four

Conduct Research

Survey Collection

Create Plan

Arrange Materials

Describe Materials

Provide Access
Arrangement and Description

- Takes place at each stage of archival processing workflow:
  - Pre-custodial intervention (survey collection)
  - **Accessioning** (capture not covered today)
    - Gathering Contextual Information
    - Performing Conservation Assessment
  - Providing an Arrangement Scheme/Intellectual Arrangement
  - Arranging the Records
  - Describing Records
  - Creating Access Tools (not covered today)

from Daines, *Processing Digital Records and Manuscripts*; see workbook for more detailed breakdown
Arrangement

- **SAA glossary:**
  - The process of organizing materials with respect to their provenance and original order, to protect their context and to achieve physical or intellectual control over the materials. – 2. The organization and sequence of items within a collection.

- **DACS:**
  - The process of identifying the logical groupings of materials within the whole as they were established by the creator, of constructing a new organization when the original ordering has been lost, or of establishing an order when one never existed.
Arrangement

AIMS* Project Definition

“At its most fundamental, arrangement in the digital world is the representation of relationships between items. The organization of material into a “folder” and “file” is representational only—the data of the digital items themselves are not organized this way on the physical hard disk or other storage medium.”

Implicit Structural Metadata/Arrangement
Step Four: Arranging the Records

- The point being... when you rearrange/rename the records, be sure to work with the copy of the files in your ‘processed’ folder.
- Watch date last modified and copy from Original_Locked as necessary
- Establish logical series order in case of disordered files, or no discernable order.

IMPT: When done, generate manifest for the “Processed Copy” and put with the other manifests in our AIP
Den4B Renamer

ReNamer Lite (non-commercial use only)

File  Settings  Presets  Language  Help

Add Files  Add Folders

Add  Remove  Up  Down

#  Rule  Statement

1  Remove  Remove all "--" (skip extension)

Files  Columns  Options  Export  Filters  Analyze

State

DashConvoFINAL.doc  DashConvoFINAL.doc
DashConvoFINAL.pdf  DashConvoFINAL.pdf
Leon Dash Transcription Accession April 19 2010.docx  Leon Dash Transcription Accession April 19 2010.docx
ReadMe.txt  ReadMe.txt
WinterMichael-- 3-27-09Archives..doc  WinterMichael 3-27-09Archives..doc
WinterMichael-- 3-27-09Archives.pdf  WinterMichael 3-27-09Archives.pdf
BettsHenryB-- 8-14-09Archives.doc  BettsHenryB 8-14-09Archives.doc
BettsHenryB-- 8-14-09Archives.pdf  BettsHenryB 8-14-09Archives.pdf
Leon Dash Transcription Accession April 7 2010.doc  Leon Dash Transcription Accession April 7 2010.doc
Leon Dash Transcription Accession April 7 2010.doc  Leon Dash Transcription Accession April 7 2010.doc
Nugent Interview Notes from PayneLindaJanuary082008Archives.doc  Nugent Interview Notes from PayneLindaJanuary082008Archives.doc
Nugent Interview Notes from PayneLindaJanuary082008Archives.pdf  Nugent Interview Notes from PayneLindaJanuary082008Archives.pdf
PatriciaSteuerwaldONE1-31-08.WAV  PatriciaSteuerwaldONE1-31-08.WAV
PatriciaSteuerwaldTWO1-31-08.WAV  PatriciaSteuerwaldTWO1-31-08.WAV
Batini-- 3-25-09Archives.doc  Batini-- 3-25-09Archives.doc

116 files (271.57 MB)
Keep in Mind the DACS Principles about Arrangement

- Records are Unique: organic nature, aggregations sharing common provenance
- Respect des Fonds, provenance, original order
- Arrangement means identifying groupings; order can be imposed as with physical records
- Arrangement to facilitate description and end-user access.
- Given ‘implicit’ metadata, can leverage file system as ‘finding aid’
Descriptive Metadata: Definitions

- “Information that characterizes the nature of an information artifact, record, archival resource, or library resource and supports its discovery”
  — Draft by SAA Dictionary Working Group, June 2015

- “Metadata, primarily on the intellectual content of an item, designed to allow resource discovery and assessment”
  — Lavoie/Garnter, *Preservation Metadata, 2nd edition*
DACS

- Creation of an accurate representation of the archival material by the process of capturing, collating, analyzing and organizing information that serves to identify archival material and to explain the context and records systems that produced, it as well the results of that process.
Descriptive Metadata

- Information regarding **content** of records
  - Formats, scope, subjects matter, etc.

- Information regarding **context** of creation
  - Creator or collector of materials
  - Functions or activities that produced them
  - Relations to other records, creators, or functions

- Record descriptive metadata according to rules described in DACS and in companion metadata standards
Critical Points from Definitions

- Description is a representation of the records.
- Descriptive system must preserve information about the context in which records are created, used, accumulated, or accessed.
- “capturing, collating, analyzing and organizing”: needs specialized tools.
Step Five: Describing the Records

- Use appropriate level of description
- Use DACS as guide
- Keep access tools in mind when designing the AIP structure, examples:
  - DAMS: CONTENTdm, Luna, etc.
  - Institutional Repositories: DSpace, Fedora, Digital Commons
  - AV systems
  - File browsers
- Include access points in your descriptive record
DACS Principles in re: Description

- Description reflects Arrangement
- Rules apply regardless of form or medium
- Rules apply to both records of organizations and personal/family papers
DACS Principles, cont.

- May use a variety of levels
  - Levels of description correspond to levels of arrangement
  - Relationship of levels must be indicated (either implicitly via system or explicitly in arrangement note)
  - Information provided at each level must be appropriate to level

- Creators must also be described (as well as their functions/activities)
Good to Know about DACS

- Doesn’t require subject access points (but a good idea to create them)
- Scope and Contents is critical field
- Appraisal and Processing Notes: critical fields for preservation metadata/actions w Digital Records.
Access Points--aim to use at collection or other top level

- Access points often are structured versions of terms mentioned in scope/contents notes
  - Names
  - Places
  - Subjects
  - Documentary Forms
    - MIME not handled well in DACS, for E-records, must look to non-DACS extensions to control, e.g. digital object management.
  - Occupations
  - Functions
Depth of Description

- DACS Chapter 1 (p. 7-11)
- Minimum:
  - ID, Repository, Title, Date, Extent, Creator Name, Scope Content, Access Conditions, Languages and Scripts
- Value Added:
  - + Admin History/Bio, improved Scope, Access Points
- Optimal:
  - Whatever else you wish
  - Processing note critical for E-records; Appraisal also Important
Levels of Description

- **Single** - suitable where documentary forms, topic, arrangement or other characteristics are consistent
- **Multilevel** - best for complex records or hybrid collections

Discussion: Do digital records require multi-level description in finding aid?
Using DACS for Digital Records

Review handout of DACS “Cheat Sheet”
Exercise Five: Single-Level Description of Digital Records using DACS

We are almost done creating our AIP, we just need to describe it!

Devise a ‘optimal’ single-level description for the records you are working with

Note: normally this would be done in a descriptive system such as ArchivesSpace, AtoM, Preservica, or ArchivesDirect

Refer to DACS 'Cheat Sheet'

Use Google form.
Accessioning Register, Processing Plan, and Descriptive Record

Single-Level Descriptive Record

ID (same as folder name)

Your answer

Level of Description

- Collection
- Series
- File
- Item

UnitTitle

- Option 1
Step Six

Conduct Research

Survey Collection

Create Plan

Arrange Materials

Describe Materials

Provide Access
Sample DIY Repo (circa 2012)
Step Six: Move records into Preservation and Access Systems:

- When all A and D work is done:
  - Generate manifest for processed copies
  - Review information in your descriptive system
  - Deposit AIP in your storage infrastructure
  - Place processed copies in dissemination (access) systems
  - Link the AIP and dissemination copies to your descriptive record

- SIP ➔ AIP ➔ DIP  You are done ( . . . for now!!!!)
Exercise Six: Practicing Skills (Making and Describing AIPs)

- Do it all again for another set of records!
Closing Discussion

- One thing I learned . . .
- One thing Chris could improve the next time around . . .
Course Evaluations

- Please provide specific feedback and written comments
- Feel free to email me if you wish with suggestions in regard to workshop
Processing Digital Records for Access and Preservation
What To Do and When To Do It!

Chris Prom
@chrisprom
prom@illinois.edu