Managing Files

- Metadata Schemas
- Naming Files
- Checksums
- Structure
- Location
- Legacy Files
- Tools
- 3rd party Vendors
Metadata

• Metadata - “Data describing the context, content and structure of records and their management through time”

  ISO 15489-1:2001 Information and documentation - Records management -- Part 1: General

• A metadata *standard* or *schema* is a set group of elements that have been standardized for a particular field.
Metadata

Types of metadata

• Administrative
  – Used in managing and administering information resources
  – Examples: rights, location, acquisition information

• Descriptive
  – Used to describe or identify information resources
  – Examples: cataloging records, finding aids, relationships between objects

• Preservation
  – Related to the preservation management of information resources
  – Examples: record of data migration, description of physical condition

• Technical
  – Related to how a system functions or metadata behave
  – Examples: hardware/software identification, digitization information

• Use
  – Related to the level and type of use of information resources
  – Examples: exhibit records, content use and re-use data
Metadata Schemas

Lots to choose from!

Seeing Standards: A Visualization of the Metadata Universe

http://www.dlib.indiana.edu/~jenlrile/metadatamap/
Metadata Schemas

What to use?

– Find what works for your institution
– *Using a standard schema increases interoperability and possibilities for reuse.*
Metadata Schemas

What to use?

Common standards for libraries, museums and archives:
- Dublin Core
- MARC
- Encoded Archival Description (EAD)
- Cataloging Cultural Objects (CCO)
- Metadata Encoding and Transmission Standard (METS)
- Metadata Object Description Schema (MODS)
- Preservation Metadata: Implementation Strategies (PREMIS)

Common standards for museums
- Cataloging Cultural Objects (CCO)
- Standard ProcEdures for CollecTions Recording Used in Museums (SPECTRUM)
- Categories for the Description of Works of Art (CDWA)
# Metadata Schemas Crosswalks

<table>
<thead>
<tr>
<th>CDWA</th>
<th>CCO</th>
<th>CONA</th>
<th>CIDOC CRM [coming soon]</th>
<th>LIDO</th>
<th>CDWA Lite</th>
<th>VRA Core</th>
<th>MARC/AACR [PDF coming soon]</th>
<th>MODS</th>
<th>Dublin Core</th>
<th>DACS</th>
<th>EAD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## 1. OBJECT/ WORK (core)

### 1.1. Catalog Level (core)
- **CDWA**: Creator_Other_Displays.catalog_level
- **CCO**: <edwaite: recordType>
- **CONA**: <vra: works> or <vra: collection>
- **CIDOC CRM [coming soon]**: 
- **LIDO**: 655 Genre/Form
- **CDWA Lite**: 300a Physical Description - Extent
- **VRA Core**: <genre>
- **MARC/ AACR [PDF coming soon]**: Level attribute
- **MODS**: 1 Levels of Description
- **Dublin Core**: Type
- **DACS**: 3.1 Scope and Content
- **EAD**: 

### 1.2. Object/Work Type (core)
- **CDWA**: PType Role_Rels. subject_id and ptype_role_id [link to AAT]
- **CCO**: <edwaite: objectWorkType>
- **CONA**: <vra: worktype> in <vra: work> or <vra: collection>
- **CIDOC CRM [coming soon]**: 
- **LIDO**: 655 Genre - Form
- **CDWA Lite**: <genre>
- **VRA Core**: Type
- **MARC/ AACR [PDF coming soon]**: 3.1 Scope and Content
- **MODS**: 
- **Dublin Core**: 
- **DACS**: <controlaccess> <genreform> (in <archdesc>)
- **EAD**: 

### 1.4. Components/Parts
- **CDWA**: archival or other description in Descriptive Note Subject scope or (counts only) Dimensions Display Physical Characteristics, display_dimensions; indexing of counts is recorded in Dimensions_Indexing
- **CCO**: 
- **CONA**: 
- **CIDOC CRM [coming soon]**: 
- **LIDO**: 
- **CDWA Lite**: 
- **VRA Core**: 300a Physical Description - Extent
- **MARC/ AACR [PDF coming soon]**: <genre>
- **MODS**: Format Extent
- **Dublin Core**: 3.1 Scope and Content
- **DACS**: 
- **EAD**: 5.4 Accruals

## 2. CLASSIFICATION (core)

### 2.1. Classification Term (core)
- **CDWA**: Class_Rels. subject_id and class_id
- **CCO**: <edwaite: classification>
- **CONA**: 
- **CIDOC CRM [coming soon]**: 
- **LIDO**: QSD 066 "Other classification number"
- **CDWA Lite**: <classification>
- **VRA Core**: Subject (classification schema)
- **MARC/ AACR [PDF coming soon]**: 
- **MODS**: 
- **Dublin Core**: 
- **DACS**: 
- **EAD**: 

## 3. TITLES OR NAMES (core)

### 3.1. Title Text (core)
- **CDWA**: Term.term
- **CCO**: 
- **CONA**: <edwaite:title>
- **CIDOC CRM [coming soon]**: 
- **LIDO**: <vra: titles in <vra: works> or <vra: collection>
- **CDWA Lite**: 
- **VRA Core**: 245a Title and Title Related Information
- **MARC/ AACR [PDF coming soon]**: <title>
- **MODS**: Title
- **Dublin Core**: 2.3 Title
- **DACS**: 
- **EAD**: <titleproper> (in <archdesc>)<untitleref> (in <archdesc>)
Metadata Schemas
Define what works for your institution

<table>
<thead>
<tr>
<th>Data Element</th>
<th>FieldName in Access database</th>
<th>Dublin Core</th>
<th>Definition</th>
<th>Value (example)</th>
<th>Required</th>
<th>Repeatable</th>
<th>Public view?</th>
<th>Has default val?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit Depth</td>
<td>ScanBitDepth</td>
<td>Format</td>
<td>Pixel depth.</td>
<td>Bit</td>
<td>Mandatory</td>
<td>No</td>
<td>Optional</td>
<td>No</td>
</tr>
<tr>
<td>Box Number</td>
<td>BoxNumber</td>
<td>Relation-IsPartOf</td>
<td>The number of the box source is stored in.</td>
<td>1</td>
<td>If Available</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Box Title</td>
<td>BoxTitle</td>
<td>Relation-IsPartOf</td>
<td>The title of the box source is stored in.</td>
<td>Correspondence</td>
<td>&lt;Not used in current design&gt;</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Capture Details</td>
<td>CaptureDetails</td>
<td>Format</td>
<td>Name scanner software, including version information; give scanner settings, gamma correction, and other relevant details pertaining to scanning.</td>
<td>Adobe Photoshop CS5.1</td>
<td>Mandatory</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Capture Device</td>
<td>CaptureDevice</td>
<td>Format</td>
<td>Indicates make and model of digital camera or scanner.</td>
<td>Epson 880XL [Flattened scanner]</td>
<td>Mandatory</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Collection Name</td>
<td>CollectionName</td>
<td>Relation-IsPartOf</td>
<td>Main entry/title of the collection (or accession).</td>
<td>Robert J. Dole House of Representatives Papers, 1960-1969</td>
<td>Mandatory</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Collection Number</td>
<td>CollectionNumber</td>
<td>Relation-IsPartOf</td>
<td>A unique sequential number assigned to the collection (or accession) - yet to be determined.</td>
<td>If Available</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Collection Abbreviation</td>
<td>CollectionAbbrev</td>
<td>Relation-IsPartOf</td>
<td>abbreviation used to identify the collection. Primarily useful for filing.</td>
<td>house</td>
<td>Mandatory</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Color Space</td>
<td>ScanColorSpace</td>
<td>Format</td>
<td>Determines the color model used to display and print images, i.e. RGB, CMYK, grayscale.</td>
<td>RGB</td>
<td>Mandatory</td>
<td>No</td>
<td>Optional</td>
<td>No</td>
</tr>
<tr>
<td>Contributor</td>
<td>Contributor</td>
<td>Contributors</td>
<td>Additional entities responsible for the creation of the item.</td>
<td>Sen. George McGovern</td>
<td>If Available</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Compression</td>
<td>ScanCompression</td>
<td>Format</td>
<td>Indicate whether or not the file has been compressed, and if it has, identifies the level and method of compression.</td>
<td>None</td>
<td>Mandatory</td>
<td>No</td>
<td>No</td>
<td>Yes = &quot;None&quot;</td>
</tr>
<tr>
<td>Condition of Item</td>
<td>ItemCondition</td>
<td>Description</td>
<td>Identifies condition of original item being digitized.</td>
<td>Slight crease mark on right</td>
<td>Optional</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Date</td>
<td>ItemDate</td>
<td>Date-Created</td>
<td>Date of the item</td>
<td>1907-03-09</td>
<td>Mandatory</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Description</td>
<td>ItemDescription</td>
<td>Description</td>
<td>Description of the item and its content</td>
<td>Meeting of the Committee on Agriculture, August 1972.</td>
<td>Mandatory</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Document Type</td>
<td>DocumentType</td>
<td>Description</td>
<td>Document type of the original</td>
<td>Press release, speech, letter, etc.</td>
<td>Mandatory</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>DPI</td>
<td>ScanDPI</td>
<td>Format</td>
<td>DPI of the master scanned image</td>
<td>600 dpi</td>
<td>Mandatory</td>
<td>No</td>
<td>Optional</td>
<td>No</td>
</tr>
<tr>
<td>Folder Title</td>
<td>FolderTitle</td>
<td>Relation-IsPartOf</td>
<td>The title of the folder source is stored in.</td>
<td>Issue Mail, March 1961</td>
<td>Mandatory</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Folder Number</td>
<td>FolderNumber</td>
<td>Relation-IsPartOf</td>
<td>The number of the folder source is stored in.</td>
<td>37</td>
<td>If Available</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Format</td>
<td>Format</td>
<td>Format-Medium</td>
<td>The physical or digital manifestation of the item. Format may be used to determine the software, hardware or other equipment needed to display or operate the item. Value to be selected from:</td>
<td>tif</td>
<td>Mandatory</td>
<td>No</td>
<td>Optional</td>
<td>No</td>
</tr>
</tbody>
</table>
Metadata

How do we store it?

• Can be stored separately from the object, or can be embedded into the digital object.

• Most management software has a place to manage metadata, but use this with caution (you want to be sure you can get it back out!)

• Excel or Access are also good choices
  – Flexible
  – Can be converted to other formats with relative ease

• Put it all in one place!!!
Metadata

Sometimes it’s a good idea to embed some descriptive metadata!
Naming your Files

<table>
<thead>
<tr>
<th>Name</th>
<th>Date Created</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSC_0057_01.JPG</td>
<td>24/04/2012 09:34 AM</td>
</tr>
<tr>
<td>DSC_0057.JPG</td>
<td>10/01/2012 10:22 AM</td>
</tr>
<tr>
<td>DSC_0058_01.JPG</td>
<td>24/04/2012 09:34 AM</td>
</tr>
<tr>
<td>DSC_0058.JPG</td>
<td>10/01/2012 10:22 AM</td>
</tr>
<tr>
<td>DSC_0059_01.JPG</td>
<td>24/04/2012 09:35 AM</td>
</tr>
<tr>
<td>DSC_0059_02.JPG</td>
<td>17/08/2012 12:49 PM</td>
</tr>
<tr>
<td>DSC_0059.JPG</td>
<td>10/01/2012 10:22 AM</td>
</tr>
<tr>
<td>DSC_0060_01.JPG</td>
<td>17/08/2012 12:49 PM</td>
</tr>
<tr>
<td>DSC_0063_01.JPG</td>
<td>10/01/2012 10:24 AM</td>
</tr>
<tr>
<td>DSC_0063.JPG</td>
<td>10/01/2012 10:23 AM</td>
</tr>
<tr>
<td>DSC_0065_01.JPG</td>
<td>25/04/2012 08:31 AM</td>
</tr>
<tr>
<td>DSC_0065_02.JPG</td>
<td>17/08/2012 12:50 PM</td>
</tr>
<tr>
<td>DSC_0065_03.JPG</td>
<td>17/08/2012 12:50 PM</td>
</tr>
<tr>
<td>PB290002.JPG</td>
<td>10/01/2012 10:23 AM</td>
</tr>
<tr>
<td>PB290282.JPG</td>
<td>10/01/2012 10:23 AM</td>
</tr>
<tr>
<td>PB290283.JPG</td>
<td>25/04/2012 07:59 AM</td>
</tr>
<tr>
<td>PB290284.JPG</td>
<td>25/04/2012 08:31 AM</td>
</tr>
<tr>
<td>PB290285.JPG</td>
<td>17/08/2012 12:44 PM</td>
</tr>
</tbody>
</table>

FINAL
FINAL.FINAL
final.FOR REAL
FINAL.version 2
absolutely.FINAL
FINAL.2
FINAL.3
FINAL.3.01
FINAL.3.02
FINAL.working
Naming your Files

• Developing a file naming convention
  – Naming the files you create according to predefined and consistent guidelines in order to:
    • Help determine something more about their content (now and in the future)
    • Accurately and consistently identify different versions/drafts of a document
    • Increase findability
    • Reduce risk of accidental overwriting or misplaced files
Naming your Files

• A few guidelines for file naming
  – Unique filenames whenever possible
  – Descriptive filenames whenever it makes sense
  – Avoid spaces, periods and other “special” characters within your file name - &%#@()<>!
  – Develop a system for naming different versions or iterations of a file and follow it consistently
    • agenda.doc
      vs.
      agenda_staff_meeting_20141105.doc
    • letterhead.doc
      vs.
      smith_thankyou_letter_20100125.doc
• Document these guidelines and make sure everyone uses them!
Checksums

• What is a checksum?
  – A “digital fingerprint”
  – Unique to each file
  – A bit-value expressed as a text number
  – Changes significantly if the file is changed in even a small way
  – Used to verify the integrity of the file
Checksums

• What types of checksums are there? Lots!
  – Cryptographic hash functions
    • Secure functions that do not allow any recreation of data from the value
    • MD5, SHA-1
      – Obsolete/”vulnerable”, but still in use
    • SHA-256
      – The most common secure checksum function
  – Checksum algorithms
Checksums

How do I use checksums?

1. Create
2. Verify
3. Repeat Step 2
Structure

Make sure the file structure reflects how you or your organization works.
Location

- Removable Media
- Network Drives
- Personal Devices

- LOCKSS
  - 3-2-1
<table>
<thead>
<tr>
<th>Benefits</th>
<th>Risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>Network Connections Dependency</td>
</tr>
<tr>
<td>Invisibility</td>
<td>Device Compatibility</td>
</tr>
<tr>
<td>Security</td>
<td>Data Ownership</td>
</tr>
<tr>
<td>Automation</td>
<td>No Hard Drive</td>
</tr>
<tr>
<td>Accessibility</td>
<td>Security – interconnectedness</td>
</tr>
<tr>
<td>Syncing</td>
<td>Contract Fine Print</td>
</tr>
<tr>
<td>Sharing</td>
<td>Longevity of Provider</td>
</tr>
<tr>
<td>Collaboration</td>
<td></td>
</tr>
<tr>
<td>Protection</td>
<td></td>
</tr>
<tr>
<td>Recovery</td>
<td></td>
</tr>
</tbody>
</table>
Legacy files - Be Proactive

Windows can't open this file:
File: permissionforms.wb3

To open this file, Windows needs to know what program you want to use to open it. Windows can go online to look it up automatically, or you can manually select from a list of programs that are installed on your computer.

What do you want to do?

- Use the Web service to find the correct program
- Select a program from a list of installed programs

OK Cancel

Did You Know?

Dead Men Tell No Passwords

The man in charge of archiving and maintaining electronic copies of Norway’s most important historical documents is dead and so is access to those archives. So the director of the Norwegian cultural center sought support from hackers to crack the center’s password-protected database.

BY THE TIME YOU DISCOVER A WORN OUT DISK, IT’S PROBABLY TOO LATE.
3rd Party Vendors

I need help!!! What should I do?

Is the vendor reliable?
Is their product compatible with the other tools you use?
How stable is the company?
What kind of support do they offer?
How widely used are they?
Will they provide training?
Can it grow with me?
DAM / Dig Pres systems

2.0 ACCESS

ACCESS

MANAGEMENT

PRESERVATION
DAM / Dig Pres systems

- Archivematica
- Preservica
- PastPerfect
- Omeka
- Archives Space
- Collective Access
- History Pin
- ArtBabble
- archivematica
- Preservica
DAM / Dig Pres systems

**PRESERVATION**

- **Ingest**
  - Checksums
  - Virus Scan
  - Dedupe
  - Unique ID

- **Processing**
  - Metadata harvest
  - Metadata creation
  - Rights management

- **Access**
  - Auto create access versions
  - Public interface

- **Storage**
  - Create open formats
  - Reliable long-term bit preservation
  - Redundancy
  - Exit strategy

- **Maintenance**
  - Format migration
  - Monitoring fixity
  - Auto recovery
DAM / Dig Pres systems

This does NOT indicate any endorsement or preference by NHPRC or KSHRAB
Trusted Digital Repository

- Repository whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future.

- Compliance with the *Reference Model for an Open Archival Information System (OAIS)*
- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technological and procedural suitability
- System security
- Procedural accountability
What does your institution need?

• Policies that make sense for your institution
• A dedicated person in charge of digital assets
• An inventory of digital assets and their importance for preservation.
• Documentation!
Policies & Forms

- Acquisition Policy
- Selection & Appraisal Policy
- Preservation Policy
- Deaccession Policy
- Destruction Policy
- Publication & Use Policy
- Access & Rights Policy
- Collections Management Policy

- Accession Form
- Deed of Gift
- Deaccession Form
- Appraisal Form
- Preservation Plan Form
- Destruction Form
- Publication Form
- Access Form
- Reproduction & Use Form
Questions?
Tools

• POWRR – (Preserving Digital Objects with Restricted Resources)
  – http://digitalpowrr.niu.edu/tool-grid/
  – An extensive list of digital preservation tools and brief descriptions of what they do
• AV Preserve
  – http://www.avpreserve.com/avpsresources/tools/
  – Fixity and other metadata/preservation tools
  – Includes a tutorial video on using Fixity
• Digital Preservation at the Library of Congress
  – http://www.digitalpreservation.gov/tools/
  – List of tools and services for digital preservation
• DataAccessioner – for migrating from removable media with built-in accession metadata
  – http://dataaccessioner.org/
• Checksum – easy checksum creator and manual checker
  – http://corz.org/windows/software/checksum/
• BagIt – a specification to package and inventory a collection of files, and to validate that the files' integrity was maintained through file transfer or over time.
  – https://github.com/LibraryOfCongress/bagit-java (files to download)
• BitCurator - Open source digital forensics tools and associated software libraries that have been modified and packaged for increased accessibility and functionality for collecting institutions.
  – http://www.bitcurator.net/
• TeraCopy - copy and move files at the maximum possible speed, with automated checksum verification
  – https://codesector.com/teracopy
• Fixity – automated checksum validation the documentation and regular review of stored files.
  – http://www.avpreserve.com/tools/fexit/
• ReNamer – powerful and flexible file renaming tool
  – http://www.den4b.com/?x=products&product=renamer
Resources

• Digital Preservation Toolkit (with specifics for museums)
  – http://www.rcip-chin.gc.ca/carrefour-du-savoir-knowledge-exchange/outils_preservation_numerique-
digital Preservation_toolkit-eng.jsp
  – Canadian Heritage Information Network

• Digital Preservation in a Box
  – http://dpoutreach.net/
  – National Digital Stewardship Alliance, 2014
  – The materials are geared towards a general audience who routinely create or manage digital information,
    but who may need a working knowledge of this area for digital preservation on the job or for training others
    on how to preserve digital resources.

• NEDCC
  – https://www.nedcc.org/free-resources/digital-preservation - lots of free resources for digital preservation

• Preservation Metadata (2nd Ed), 2013
  – Digital Preservation Coalition
  – http://dx.doi.org/10.7207/twr13-03

• Preservation and Sustainability Guidelines for digital assets in museums

• Upcoming Courses and Workshops
  – A calendar of mostly online trainings opportunities in digital preservation.
Special Thanks