Introduction to Digital Vocabularies

...what you need to look for and ask about...

(There will also be additional background information for site visits)
Using this Software

- Microphone
- Raising your hand
- Green ✓/ Red X
- Laughing / Clapping
- Stepping out
- Text chat
- Audio
- Exiting
Agenda

• Vocabulary of digitization
  – Common misconceptions
  – Delivery of files
  – Definitions and terms
• Digitization best practices
  – Imaging
  – Audio
  – Video
  – Metadata
• Storage and preservation of digital files
• Where to find appropriate resources
• Report writing process
Characteristics of a Well-Integrated Preservation Program

Preservation is:

– Incorporated into policies
– Part of the institution’s emergency management plan
– A component of public relations activities
– Incorporated into user education and staff training
A Preservation Policy

- Tied to institutional mission
- Indicates scope of institution’s commitment
- Clarifies what aspects of a collection are being preserved and why
- Outlines institution’s preservation strategies for all materials
- Outlines which standards/guidelines are being followed
Elements of Preservation: Proactive Activities

- Environmental controls
- Pest management
- Disaster preparedness
- Security
- Collections processing, management and housing
- Staff/user education
Components of Digital Collection Development Policies

- Introduction and Purpose
- Audience
- Scope of the Digital Collection
- Collection Strengths
- Selection Criteria
- Needs of the Materials
- Digitization Process
- Intellectual control and intellectual Property Rights
- Collaboration
- Disposition and Digital Preservation
Resources for Digital Collection Development Policies


  [http://www.mln.lib.ma.us/about/digitalcolldev.htm](http://www.mln.lib.ma.us/about/digitalcolldev.htm)

- University of Vermont Libraries Center for Digital Initiatives. CDI Collection Development Policy. 
  [http://cdi.uvm.edu/about/colldev](http://cdi.uvm.edu/about/colldev)
The NISO Framework of Guidance for Building Good Digital Collections has three purposes:

- To provide an overview of some of the major components and activities involved in creating good digital collections
- To identify existing resources that support the development of sound local practices for creating and managing good digital collections
- To encourage community participation in the ongoing development of best practices for digital collection building
Principles of Good Collections

• A good digital collection is created according to an explicit collection development policy.
• Collections should be described so that a user can discover characteristics of the collection.
• A good collection is curated.
• A good collection is broadly available and avoids unnecessary impediments to use.
• A good collection respects intellectual property rights.
• A good collection has mechanisms to supply usage data.
• A good collection is interoperable.
• A good collection is sustainable over time.
Digital Content

- Digital Objects, curriculum materials, social content—Content in the largest sense...
Definitions: Digitization—there is no “L”

- The process of creating digital files by scanning or otherwise converting analog materials. The resulting digital copy, or digital surrogate, would then be classed as digital material and then subject to the same broad challenges involved in preserving access to it, as "born digital" materials.
Definitions: Born Digital

- Refers to digital material that does not have an analog equivalent either as the originating source or as a result of conversion to analog form. This term has been used to differentiate them from:
  1) Digital materials which have been created as a result of converting analog originals; and
  2) Digital materials, which may have originated from a digital source but have been printed to paper, e.g. some electronic records.

Definitions: Digital Resources

• A broad term encompassing digital surrogates created as a result of converting analog materials to digital form (digitization,) and “born digital” resources for which there has never been and is never intended to be an analog equivalent, and digital records.
Definitions: Reformatted Digital Collections

• Reformatting is the copying of information content from one storage medium to a different storage medium (media reformatting) or converting from one file format to a different file format (file reformatting.)
“Scanning” to mean reformatting...

- Often you will hear “scanning” to mean digitization...
So what about Systems? Why are people so confused?

THIS IS TECHNICAL
Definitions: CMS

• A content management system (CMS) is a computer application used to create, edit, manage, and publish content in a consistently organized fashion.
Collections Management Software

PastPerfect 5.0: The best-selling museum collection management software. Affordable, comprehensive, and easy-to-use for organizations and collections of all sizes.
Definitions: DAM

- Digital asset management consists of tasks and decisions surrounding ingesting, annotating, cataloguing, storage, retrieval, and distribution of digital files.
- Digital asset management systems (DAMS) include computer software and hardware systems that aid in the process of digital asset management.
- The term "digital asset management" (DAM) also refers to the protocol for downloading, renaming, backing up, rating, grouping, archiving, optimizing, maintaining, thinning, and exporting files.
Software

• Digital Asset or Digital Content Management Software Systems
  – CONTENTdm (OCLC)
  – LUNA Insight
  – DigiTool (ExLibris)
  – Fedora/Islandora
  – D-Space
  – The Gallery System
  – Innovative Interfaces ContentPro
  – Collective Access
  – Extensis Portfolio
  – Others

• Collections Management Software
  – PastPerfect
  – Collections Space
  – ArchivesSpace
    • Archivists Toolkit
    • Archon
  – Argus
  – Others

• Website Software
  – Omeka (Open Source, Publishing tool)
  – Drupal (Open Source)
  – Others
What do we look for? Where can we point folks?

FILE TYPES AND BEST PRACTICES FOR CREATION
NISO Framework...

- Principle 1: A good object exists in a format that supports its intended current and future use.
- Principle 2: A good object is preservable.
- Principle 3: A good object is meaningful and useful outside of its local context.
- Principle 4: A good object will be named with a persistent, globally unique identifier that can be resolved to the current address of the object.
- Principle 5: A good object can be authenticated.
- Principle 6: A good object has associated metadata.
Besser Principles

• Capture at the highest resolution appropriate to the informational content of the originals
• Capture at an appropriate level of quality to avoid re-handling of the originals in the future—capture once
• Create and store a master file that can be used to produce derivative files and serve a variety of current and future user needs
• Use system components that are non-proprietary
• Use file formats and compression techniques that conform to (CHI) industry standards
Besser Principles

• Create backup copies of all files on a stable medium—Look for collections of CD’s/DVD’s ask about external drives and servers…
• Create meaningful metadata for files or collections
• Store media in an appropriate environment
• Monitor and recopy data as necessary
• Outline a migration strategy for transferring data across generations of technology
• Anticipate and plan for future technological developments
Guidelines for Source Type
<table>
<thead>
<tr>
<th>Photographs</th>
<th>Master</th>
<th>Access</th>
<th>Thumbnail</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Format</td>
<td>TIFF</td>
<td>JPEG</td>
<td>JPEG</td>
</tr>
<tr>
<td>Bit Depth</td>
<td>16 bit grayscale</td>
<td>8 bit grayscale</td>
<td>8 bit grayscale</td>
</tr>
<tr>
<td></td>
<td>48 bit color</td>
<td>8 bit grayscale</td>
<td>24 bit color</td>
</tr>
<tr>
<td>Spatial Resolution</td>
<td>400 to 800 PPI</td>
<td>150 to 200 PPI</td>
<td>144 PPI</td>
</tr>
<tr>
<td>Spatial Dimensions</td>
<td>4000 to 8000 pixels across the long dimension, depending on size of original, excluding mounts and borders</td>
<td>600 pixels across the long dimension</td>
<td>150 to 200 pixels across the long dimension</td>
</tr>
<tr>
<td>Maps</td>
<td>Master</td>
<td>Web</td>
<td>Thumbnail</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------</td>
<td>-------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>File Format</td>
<td>TIFF</td>
<td>JPEG</td>
<td>JPEG</td>
</tr>
<tr>
<td>Bit Depth</td>
<td>16 bit grayscale 48 bit color</td>
<td>8 bit grayscale 24 bit color</td>
<td>8 bit grayscale 24 bit color</td>
</tr>
<tr>
<td>Spatial Resolution</td>
<td>600 PPI 300 to 400 PPI for larger maps</td>
<td>150 to 200 PPI</td>
<td>144 PPI</td>
</tr>
<tr>
<td>Spatial Dimensions</td>
<td>6000 to 8000 pixels across the long dimension</td>
<td>1078 pixels across the long dimension</td>
<td>150 to 200 pixels across the long dimension</td>
</tr>
</tbody>
</table>
### Artwork/3-Dimensional Objects

<table>
<thead>
<tr>
<th>Master</th>
<th>Access</th>
<th>Thumbnail</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>File Format</strong></td>
<td>TIFF</td>
<td>JPEG</td>
</tr>
<tr>
<td><strong>Bit Depth</strong></td>
<td>48 bit color</td>
<td>24 bit color</td>
</tr>
<tr>
<td><strong>Spatial Resolution</strong></td>
<td>Device Maximum</td>
<td>300 PPI</td>
</tr>
<tr>
<td><strong>Spatial Dimensions</strong></td>
<td>100% of original</td>
<td>600 pixels across the long dimension</td>
</tr>
</tbody>
</table>

Materials may be captured as a RAW file and converted

http://collectioncare.auraria.edu
<table>
<thead>
<tr>
<th>Document Character - Original</th>
<th>Recommended Image Parameters</th>
<th>Alternative Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clean, high-contrast documents with printed type (e.g. laser printed or typeset)</td>
<td>1-bit bitonal mode or 8-bit grayscale - adjust scan resolution to produce a QI of 8 for smallest significant character or 1-bit bitonal mode - 600 ppi* for documents with smallest significant character of 1.0 mm or larger or 8-bit grayscale mode - 400 ppi for documents with smallest significant character of 1.0 mm or larger</td>
<td>1-bit bitonal mode - 300 ppi* for documents with smallest significant character of 2.0 mm or larger or 8-bit grayscale mode - 300 ppi for documents with smallest significant character of 1.5 mm or larger</td>
</tr>
<tr>
<td>*The 600 ppi 1-bit files can be produced via scanning or created/derived from 400 ppi, 8-bit grayscale images.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documents with poor legibility or diffuse characters (e.g. carbon copies, Thermofax/Verifax, etc.), handwritten annotations or other markings, low inherent contrast, staining, fading, halftone illustrations, or photographs</td>
<td>8-bit grayscale mode - adjust scan resolution to produce a QI of 8 for smallest significant character or 8-bit grayscale mode - 400 ppi for documents with smallest significant character of 1.0 mm or larger</td>
<td>8-bit grayscale mode - 300 ppi for documents with smallest significant character of 1.5 mm or larger</td>
</tr>
<tr>
<td>*The 300 ppi 1-bit files can be produced via scanning or created/derived from 300 ppi, 8-bit grayscale images.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documents as described for grayscale scanning and/or where color is important to the interpretation of the information or content, or desire to produce the most accurate representation</td>
<td>24-bit color mode - adjust scan resolution to produce a QI of 8 for smallest significant character or 24-bit RGB mode - 400 ppi for documents with smallest significant character of 1.0 mm or larger</td>
<td>24-bit RGB mode - 300 ppi for documents with smallest significant character of 1.5 mm or larger</td>
</tr>
<tr>
<td>*NOTE: Regardless of approach used, adjust scan resolution to produce a minimum pixel measurement across the long dimension of 4,000 lines for 8-bit files</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Source Types
Following are some recommendations for basic source types

- **Spoken language:** Virtually all human voices fall within the 50kHz-20kHz frequency range, so the recommended sampling rate for spoken language is 44.1 kHz
Source Types

Field Recordings: We are generally referring to spoken language recordings occurring in the field, so the recommended sampling rate is, again, 44.1 kHz. For field recordings that include music, or wild sounds such as insects, birds, or other natural sounds, the 96 kHz sampling rate should be considered.
Source Types

- *Musical Recordings*: Musical instruments produce a wide range of frequencies, some above the 22.05 kHz capabilities of the 44.1 kHz sampling rate. Although 44.1 kHz may be more than adequate for many musical recordings, in general we recommend the 96 kHz sampling rate, which will record frequencies up to 48 kHz, and provide additional information that may be useful during audio processing.
Recommendations for Video

Why migrate to digital media?

• Obsolescence of analog formats and equipment
• Film / tape degradation
• To create access copies
• Copies can be made without generational quality loss
What is compression?

• Compression makes files smaller and easier to use.
• In video, compression discards color information, resolution and, sometimes, sound quality.
• With audio, lossy compression reduces the amount of data in a recorded waveform for transmission.
• Dynamic range compression, also called audio level compression, is when the dynamic range is reduced.
Codecs

- Codec: Coder/Decoder – a computer program that create and/or play back digital video and/or audio
- Also can stand for “Compressor/Decompressor”
- As of December 10, 2013 recommendations for video have emerged in the cultural heritage community
Some Common Video Codecs

- WMV (Windows Media Viewer) – proprietary
- DivX – proprietary
- Sorenson (used in Quicktime) – proprietary
- Xvid – open source
Video File Formats

- File formats use the different codecs to play back video and audio
- Video file formats are containers
- Look for non-proprietary file formats when possible
Some Common Video File Formats

- Quicktime MOV – proprietary; common on Apple platforms; well supported
- AVI (Audio Video Interleaved) – proprietary; common on Windows platforms; well supported
- Flash Video – proprietary; common on online streaming sites like YouTube
Some Common Video File Formats

• MPEG-2 – standard devised by Motion Picture Experts Group; DVD encoding; compressed

• MPEG-4 – used online as downloadable video; very compressed
Example of Preservation Master File Requirements

• Wrapper: QuickTime (.mov)
• Video Stream:
  – Uncompressed, 10-bit 4:2:2 YUV
• Audio Stream: 48Khz/24-bit PCM
Motion JPEG 2000

- Lossless video compression
- Large file size
- There have been recommendations that say it can be used as an archival standard for preserving digital video
- Has space for metadata, recommendations are evolving the FADGI working group suggested including using this with an MXF wrapper
- Continue to monitor!
Resources

• Available for download

  – BCR’s CDP Digital Imaging Best Practices, Version 2.0
  – CDP Digital Audio Working Group
    Digital Audio Best Practices Version 2.1
  – CDP Western States Metadata Best Practices Version 2.1.1
• MODS User Guidelines Version 3
  – http://www.loc.gov/standards/mods/userguide
• Guidelines: Technical Guidelines for Digitizing Cultural Heritage Materials
• Digitizing Video for Long-Term Preservation: An RFP Guide and Template.
  – Barbara Goldsmith Preservation & Conservation Department, New York University Libraries, 2013. Funded by The Andrew W. Mellon Foundation. Contributors include: Paula De Stefano, Kimberly Tarr, Melitte Buchman, Peter Oleksik, Alice Moscoso, Ben Moskowitz
  – http://library.nyu.edu/preservation/VARRFP.pdf
Standards and Best Practices

METADATA
NISO Framework...A quick review

• A primary reason for building digital collections is to increase access to the resources held by the organization. Creating broadly accessible descriptive metadata is a way to maximize access by current users and attract new user communities.

• As cultural heritage institutions explore the metadata standards that are being adopted within their field, they will want to consider the interoperability issue early in their metadata implementation to ensure the greatest likelihood of interoperability.

• Give considerable thought to which controlled vocabularies and thesauri they should implement.
What is Metadata?

• Metadata is data that *facilitates the management, description, and preservation of a digital object or aggregation of digital objects.*

• The creation of metadata is governed by a *body of standards, best practices and schemas* that, when appropriately applied, work together to facilitate the management, description, and preservation of digital objects.
Metadata for Users

• **Descriptive metadata**: used for the indexing, discovery, and identification of the contents of a digital resource.

• **Structural metadata**: information used to display and navigate digital resources; also includes information on internal organization of the digital resource. For example it might include information such as the structural divisions of a resource (i.e. chapters in a book) or sub-object relationships (such as individual diary entries in a diary section).
Metadata for Creators cont.

• **Technical Metadata:**
  Records information about the digital file itself, such as: format, size, encoding, etc.

• **Rights Metadata:**
  Records information about intellectual property and copyrights.

---

http://collectioncare.auraria.edu
Metadata for Creators

- Administrative metadata: Represents the management information for the digital object, which may include information needed to access and display the resource, as well as rights management information.
Metadata for Creators cont.

• Preservation Metadata:
Records information about the provenance of a digital object throughout it’s lifetime, including activities to help administrators manage and record information about preservation actions.

  – It helps maintain the provenance of the digital object and track preservation activities throughout the lifetime of a digital object.
What does this mean and what is the confusion here?

DIGITAL PRESERVATION
Changing world…and more

http://collectioncare.auraria.edu
Backup vs. digital preservation

‘Disaster recovery strategies and backup systems are not sufficient to ensure survival and access to authentic digital resources over time. A backup is short-term data recovery solution following loss or corruption and is fundamentally different to an electronic preservation archive.’

» JISC. Digital Preservation: Continued Access to authentic digital assets. (November, 2006)
Digital Preservation

- Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time.
  - Prepared by the ALCTS Preservation and Reformatting Section, Working Group on Defining Digital Preservation
  - ALA Annual Conference, Washington, D.C., June 24, 2007
  - http://www.ala.org/alcts/resources/preserv/defdigpres0408

http://collectioncare.auraria.edu
Assumptions

- Digital preservation is more challenging and complex than preservation of analog objects
  - Encourage sites to not delay doing something just because they don’t feel we have all the answers...we have a lot of good answers already
  - With standards in place for content creation and metadata you are already ahead of the game
- Digital preservation is more than just a technical preservation strategy
  - “THE” one perfect solution doesn’t exist
- Digital preservation needs to be integrated into an organization’s culture
- Current financial environment isn’t a reason to delay action
Organizational Challenges

• Technical expertise
• New partnerships
  – Content creators: what is their role?
  – Collaborate on preservation programs
• Understand your scope
  – Who are we preserving for?
  – What should we preserve?
  – How will we preserve it?
  – How do we assure it’s accessible in the long term?
The Elements of a Good Digital Plan: Policies that Define

- Mission and vision for digital collections and that tie back to collection policies and institutional mission statements
- Collections development and selection criteria
- Capture guidelines
- Rights management
- Metadata guidelines
- Quality control procedures
- Delivery methods
- Digital preservation approach
Digital Preservation Tools and Systems

• Archivematica
• LOCKSS
• OCLC’s Digital Archive
• Preservica
• Rosetta
• Others...
Report Writing
Thoughts? Remarks?
Hold the Date

The next Peer-to-Peer Webinar

- *Understanding the Role of Temperature, Relative Humidity and Dew Point in*
  - *Creating a Sustainable Preservation Environment*

- Date: February 13th: 11:00 – 1:00

- Guest Presenter:
  - Jeremy Linden, Preservation Environment Specialist from Image Permanence Institute (IPI)

Homework: Please watch the following YouTube before February 13th!

http://www.youtube.com/watch?v=8S-xRILPAM8
Acknowledgements

• This project is made possible by a grant from the U.S. Institute of Museum and Library Services (IMLS). The Colorado Connecting to Collections initiative is managed by the Center for Colorado & the West at Auraria Library as a collaborative partnership with the Colorado Wyoming Association of Museums (CWAM), the Society of Rocky Mountain Archivists (SRMA), the Colorado State Library (CSL) and History Colorado

• The views, findings, conclusions or recommendations expressed in this program do not necessarily represent those of the Institute of Museum and Library Services
Instructor

Leigh A. Grinstead
Digital Initiatives Consultant
404-520-8615
leigh.grinstead@lyrasis.org