Building Digital Preservation Pathways at NEDCC
Challenger:
Understanding Our
Clients and
Collaborators...

...and Ourselves
Making Choices

Analog / Physical

Born Digital

Digital Capture

Digital Preservation Actions

“Digital Mortgage”

Digital Objects
Digital Preservation ???

• Should I do this?
• How do I do this?
• What’s the right “____”? 
• Who should do this?
• How do I talk to “_____” about doing this?
• How do I know if I made the right choice?
Primary Areas of Concern

• Standards and best practices
• Workflows and staffing
• Storage for preservation
• Rights and usage
• Funding
How we answer...

GOOD → BETTER → BEST
Resources

• Start with “The Classics:
  – NISO’s A Framework of Guidance for Building Good Digital Collections
  – CRL’s Trustworthy Repositories Audit & Certification: Criteria and Checklist

• Consider some new additions to the pack...
**Digitization Guidelines**

- **Audio Digitization System Performance**
  Draft for Public Comment | Comment period closes April 15, 2012
  This document discusses the relevant metrics and measurement methods for analog-to-digital converters. Future performance-related documents will discuss the problem of interstitial errors, i.e., accidental loss or transformations of audio samples within the digitizing system before the data stream is written to file.

- **Minimal Descriptive Embedded Metadata in Digital Still Images**
  Reviewed and Recommended by Working Group | March 23, 2012
  Guidelines created by the EMDeWG (Embedded Metadata Working Group) of the Smithsonian Institution and recommended by the Still Image Working Group. This document defines the minimum proposed descriptive embedded metadata for digital still images.

- **Embedding Metadata in Broadcast WAVE Files**
  Draft for public comment | Comment period closes February 21, 2012
  Revision to the guideline for metadata to be embedded in Broadcast WAVE files that reproduce historical and cultural heritage digital sound recordings.

- **Technical Guidelines for the Still Image Digitization of Cultural Heritage Materials**
  Updated by Working Group | August 24, 2010
  Creation of Raster Image Master Files represents shared best practices.

- **New & Events**
  - [Information or Artifact? Four-part blog on digitizing books and photos](http://www.digitizationguidelines.gov/
    Information or Artifact? Four-part blog on digitizing books and photos)
    October 17-20, 2011
  - [Evaluating Still Image Digitization and Digitization Equipment](http://www.digitizationguidelines.gov/
    Evaluating Still Image Digitization and Digitization Equipment)
    (PDF, 2.5 MB) | August 16, 2010
  - [Information Standards Quarterly article](http://www.digitizationguidelines.gov/
    Information Standards Quarterly article)
    (PDF, 528 KB) | Spring 2010
  - [Federal Computer Week article](http://www.digitizationguidelines.gov/
    Federal Computer Week article)
    February 5, 2010

- **Resources**
  - Digital Imaging Standards
  - Evaluating Still Image Digitization and Digitization Equipment (PDF, 2.5 MB)
  - "Format Considerations in Audio Visual Preservation Reformating" (PDF, 528 KB)

**About This Initiative**

- **Still Image Working Group**
  This group is involved in a cooperative effort to develop common digitization guidelines for still image materials.

- **Audio-Visual Working Group**
  The goal for this working group is to identify, establish, and disseminate information about standards and practices for the digital reformatting of audio-visual materials.

URL: [http://www.digitizationguidelines.gov](http://www.digitizationguidelines.gov)
Sustainability of Digital Formats
Planning for Library of Congress Collections

Format Description Categories >> Browse Alphabetical List

Browse alphabetical list

A

- **AAF**
  - **AAF 1.1**, Advanced Authoring Format (AAF), Version 1.1

- **AAC**
  - **AAC MP2**, Advanced Audio Coding, MPEG-2
  - **AAC ADIF**, Advanced Audio Coding, MPEG-2, Audio Data Interchange Format
  - **AAC MP4**, Advanced Audio Coding, MPEG-4
  - **AAC MP4 LC**, AAC (MPEG-4) Low Complexity Object
  - **MP4 FF 2 AAC**, MPEG-4 File Format, V.2, with Advanced Audio Coding
  - **QTA AAC**, QuickTime Audio, AAC Codec

- **AC-3**
  - **AC-3 A**, AC-3 Compressed Audio (Dolby Digital), Revision A

- **AES3**
  - **AES3**, AES3, Digital Audio Interface Format
  - **AES3 SMPTE**, AES3 Digital Audio Interface, SMPTE Extensions

- **AIFF**
  - **AIFF**, Audio Interchange File Format
  - **AIFF LPCM**, AIFF File Format with LPCM Audio

- **A-Law**
  - **A-Law**, A-Law Compressed Sound Format

- **AMR**
  - **AMR**, Adaptive Multi-Rate Speech Codec
Main Page

From Archivematica

What is Archivematica?

Archivematica is a comprehensive digital preservation system. Archivematica uses a micro-services design pattern to provide an integrated suite of free and open-source tools that allows users to process digital objects from ingest to access in compliance with the ISO-OAI/S functional model.

Users monitor and control the micro-services via a web-based dashboard. Archivematica uses METS, PREMIS, Dublin Core and other best practice metadata standards. Archivematica implements media type preservation plans based on an analysis of the significant characteristics of file formats.

The overview section provides a detailed description of Archivematica's functionality and technical architecture. This screen cast gives a demo of the Archivematica 0.7.1-alpha release.
## Media type preservation plans

<table>
<thead>
<tr>
<th>Media type</th>
<th>File formats</th>
<th>Preservation format(s)</th>
<th>Access format(s)</th>
<th>Normalization tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>AC3, AIFF, MP3, WAV, WMA</td>
<td>WAVE (LPCM)</td>
<td>MP3</td>
<td>FFmpeg</td>
</tr>
<tr>
<td>Email</td>
<td>PST</td>
<td>MBOX</td>
<td>MBOX</td>
<td>readpst</td>
</tr>
<tr>
<td>Office Open XML</td>
<td>DOCX, PPTX, XLSX</td>
<td>Original format</td>
<td>PDF for PPTX</td>
<td>OpenOffice</td>
</tr>
<tr>
<td>Plain text</td>
<td>TXT</td>
<td>Original format</td>
<td>Original format</td>
<td>None</td>
</tr>
<tr>
<td>Portable Document Format</td>
<td>PDF</td>
<td>PDF/A</td>
<td>Original format</td>
<td>Ghostscript</td>
</tr>
<tr>
<td>Presentation files</td>
<td>PPT</td>
<td>Original format</td>
<td>PDF</td>
<td>OpenOffice</td>
</tr>
<tr>
<td>Raster images</td>
<td>BMP, GIF, JPG, JP2*, PCT, PNG*, PSD, TIFF, TGA</td>
<td>Uncompressed TIFF</td>
<td>JPEG</td>
<td>ImageMagick</td>
</tr>
<tr>
<td>Raw camera files/Digital Negative format**</td>
<td>3FR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MRW, NEF, ORF, PEF, RAF, RAW, X3F</td>
<td>Original format</td>
<td>JPEG</td>
<td>ImageMagick/uFRaw</td>
</tr>
<tr>
<td>Spreadsheets</td>
<td>XLS</td>
<td>Original format</td>
<td>Original format</td>
<td>OpenOffice</td>
</tr>
<tr>
<td>Vector images</td>
<td>AI, EPS, SVG</td>
<td>SVG</td>
<td>PDF</td>
<td>Inkscape</td>
</tr>
<tr>
<td>Video</td>
<td>AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV</td>
<td>FFV1/LPCM in MKV</td>
<td>MPEG-1</td>
<td>FFmpeg</td>
</tr>
<tr>
<td>Word processing files</td>
<td>DOC, WPD, RTF</td>
<td>ODF (WPD and RTF)</td>
<td>PDF</td>
<td>OpenOffice</td>
</tr>
</tbody>
</table>

* (*PNG and JPEG2000 are not normalized to a preservation format)
* (**in development*
New Roles for New Times:

Digital Curation for Preservation

March 2011

Tyler Walters
Katherine Skinner
CODE OF BEST PRACTICES IN FAIR USE FOR ACADEMIC AND RESEARCH LIBRARIES

JANUARY 2012
## Storage Needs

<table>
<thead>
<tr>
<th>Original Color Film</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>35mm</td>
<td>50 MB</td>
</tr>
<tr>
<td>120 square</td>
<td>80 MB</td>
</tr>
<tr>
<td>120 6x4.5</td>
<td>60 MB</td>
</tr>
<tr>
<td>120 6x9</td>
<td>90 MB</td>
</tr>
<tr>
<td>4x5</td>
<td>135 MB</td>
</tr>
<tr>
<td>8x10</td>
<td>240 MB</td>
</tr>
</tbody>
</table>
Bit-level, Content, and Format Preservation

- Store
- Refresh
- Migrate
  - LC’s Sustainability Formats
- Emulate
  - mimeTypes
  - Pronom, Global Digital Format Registry
- Encapsulation
- Preserve the Technology
- Digital Forensics
- “3-3-3” Strategy
Preservation Infrastructure Strategies

- Redundancy
- Disaster recovery planning
- Diversify funding sources
- Document activities
- Maintain and grow skill sets and knowledge bases
- Policies, Policies, Policies
- Planning, Planning, Planning...
- Saving for a “Rainy Day”
Amazon Simple Storage Service (Amazon S3)

Amazon S3 is storage for the Internet. It is designed to make web-scale computing easier for developers.

Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, secure, fast, inexpensive infrastructure that Amazon uses to run its own global network of web sites. The service aims to maximize benefits of scale and to pass those benefits on to developers.

This page contains the following categories of information. Click to jump down:

- **Amazon S3 Functionality**
  - Protecting Your Data
  - Pricing
  - Getting Started with Amazon S3
  - Transferring Large Amounts of Data

- **Common Use Cases**
- **Resources**
- **Amazon S3 Design Requirements**
- **Intended Usage and Restrictions**

Amazon S3 Functionality

Amazon S3 is intentionally built with a minimal feature set.

- Write, read, and delete objects containing from 1 byte to 5 terabytes of data each. The number of objects you can store is unlimited.
- Each object is stored in a bucket and retrieved via a unique, developer-assigned key.
- A bucket can be stored in one of several Regions. You can choose a Region to optimize for latency, minimize costs, or address regulatory requirements. Amazon S3 is currently available in the US Standard, EU (Ireland), US West (Northern California), Asia Pacific (Singapore), Asia Pacific (Tokyo) and GovCloud (US) Regions. The US Standard Region automatically routes requests to facilities in Northern Virginia or the Pacific Northwest using network maps.
- Objects stored in a Region never leave the Region unless you transfer them out. For example, objects stored in the EU (Ireland) Region never leave the EU.
- Authentication mechanisms are provided to ensure that data is kept secure from unauthorized access. Objects can be made private or public, and rights can be granted to specific users.
- Options for secure data upload/download and encryption of data at rest are provided for additional data...
Preserved Collection

Georgia Tech Photograph Collection

The collection includes more than 5,400 images that document the institution's rich history. These scanned images are stored as JPG and TIFF files in the Archives' instance of DSpace (viewable at [Georgia Tech History Digital Portal](#)).
URL: Google Forms for Contact Info

http://is.gd/DVNssh