PLANNING AND MANAGING DIGITAL PROJECTS

John Nemmers and Mark Sullivan
University of Florida Smathers Libraries
Section 1: Overview & Project Planning
A Digital Project...

- Has a time frame (start/end dates)
- Has a scope/purpose
- Has a project manager
- Requires resources/inputs
- Has a work plan
- Produces digital outputs
Digital Project Breakdown

- 1/3 effort = project planning, pre-imaging preparation, preservation, oversight, evaluation
- 1/3 effort = description, metadata creation, database creation/maintenance
- 1/3 effort = digitization
Questions to Answer:

- Who will be served by project?
- Have we identified the Primary/Secondary/Other audiences?
- What will be digitized?
- What are our priorities?
- What will be the impact on the institution?
- How much should we digitize?
- What is our timeline? Start/end dates?
Questions to Answer:

- What resources do we need to support project?
- What are funding sources?
- Will project activities be undertaken in-house or outsourced?
- What standards and best practices will be adopted?
- What technologies will we use to create/maintain/deliver/preserve digital objects?
- What is our sustainability plan beyond the project end?
Initial Steps in the Process

- Identify holdings to be digitized
- Identify stakeholders and participants (internal/external)
- Create a planning team
- Determine project goals/objectives
- Environmental scan of other projects
- Identify needed resources
- Develop work plan with timeline of activities
Identify Holdings for Digital Projects

- Archival collections (papers, photos, AV, etc.)
- Special collections
- Reformatting microfilmed holdings
- Born digital holdings
- Digitization projects with partner holdings
Selective Digitization

- Focus on holdings with high research/educational/commercial value
- Often not feasible to digitize everything -> representative sample
- Not everything merits digitization
- Are the originals fragile/in jeopardy?
- Rights? No known copyright holder -> Take down notices are option.
- Permissions? Deed of gift?
What are Priorities

- Researcher demand?
- Preservation?
- Administrative mandate?
- Donor driven?
- Matching priorities of funding agency/partners?
- Exposing hidden collections to increase use?
Prioritizing Example: Everglades Collections at UF

- Collections in demand
- Hot subject area
- Preservation issues
- Easy formats
- Viable funding source
Potential Stakeholders/Participants

- Archivists/Curators
- Imaging personnel
- Metadata personnel
- Cataloging personnel
- IT/Web personnel
- Preservation or Conservation personnel
- Advisors (e.g. panel of subject specialists)
- Development/Fundraising personnel
- Grant writer/manager
- Administration
- Donors
- PR/Marketing personnel
- External partners
- Vendors
Creating a Team

- Need for clear leadership (project manager or co-managers)
- Consensus-building
- Group decision-making
- Expect resistance (both justified and not)
- Admin support usually = successful elimination of resistance
Collaboration & Partnerships

- Gain holdings/knowledge/support/IT from partners
- Shared knowledge/experiences can lead to better workflow/products
- Some granting agencies favor collaborative digital projects
- Requires better communication/negotiation of expectations/activities/outcomes
In-House vs. Outsourcing (or Hybrid)

- Outsourcing costs can be lower if infrastructure/expertise not already in-house
- Vendors set up to handle high volume; have equipment, training, software
- Vendors often offer multiple services in addition to imaging
- Vendor security risks – materials going off site/out of country (loss of control)
- Have to evaluate, select, contract and communicate with vendors
In-House vs. Outsourcing (or Hybrid)

- In-house projects lead to in-house expertise/infrastructure, but can require significant start-up costs.
- Complete control over in-house projects (fewer security risks).
- Other pros/cons?
- Hybrid approach: use vendors to handle one portion of project (e.g., only imaging).
Common Outsourced Activities

- Scanning
- Transcription (e.g., oral histories)
- Translation
- Metadata creation
- OCR
- Database management
- Online hosting
- Preservation
Planning Discussions

- Brainstorming!
- Consider all ideas/options
- Identify impacts
  - Staffing (extant or hire new or outsource)
  - Space, equipment, software
  - Impact on regular workflow
  - Holdings (e.g., preservation/access issues)
- Evaluate all info and make decisions
Base Decisions on Data

- Review/gather statistics
  - Extent (e.g., # of page images)
  - Reference/Usage statistics
  - Scholarship produced based on holdings
- Can create surveys/suggestion forms
- Focus/advisory groups
Work Plan Development

- Work plan focuses on quantifiable goals (e.g., producing a certain # of images)
- What steps/activities will you implement to achieve goals?
- Shows how inputs/resources are used to produce outcomes/products
- Detailed budget (calculate all costs)
TIP! Review grant guidelines, which usually call for:

- Specific goals and objectives
- Scope of project
- Description of the materials and their importance/significance
- Budget templates
- Scanning/Metadata standards
- Specific work plan
- Evaluation criteria
Work Plan

- Projects usually have phases and benchmarks as part of a project timeline
- Project activities are tied to project goals
- Activities also tied to project resources
- Identify staff – who (extant or new), how much time needed, etc.
- Identify space, equipment, supplies needed (extant or new)
Work Plan

- Identify other costs (e.g., consultants, vendors, training, travel, shipping)
- Hidden costs (administrative, PR) – can be cost share in grants
- Technology issues: software/hardware/metadata/standards
Work Plan

- Work plan provides details about workflow / methodology
- Simple workflow:
## Project Timeline Example

<table>
<thead>
<tr>
<th>Months</th>
<th>Actions</th>
<th>Steps</th>
<th>Responsibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>Hire/train staff; buy equipment/supplies</td>
<td>Interview candidates; training</td>
<td>Project manager, digital supervisor</td>
</tr>
<tr>
<td>3-12</td>
<td>Scanning and Metadata creation</td>
<td>Scanning, QC, loading into system</td>
<td>Digital technicians; IT experts</td>
</tr>
<tr>
<td>13-16</td>
<td>Evaluation, promotion</td>
<td>Focus groups, listserv emails, presentations, journal article</td>
<td>Project manager; publicity dept, project staff</td>
</tr>
</tbody>
</table>
Gantt Charts

Source: http://en.wikipedia.org/wiki/File:GanttChartAnatomy.png
Possible Project Budget Items

- Salaries & Benefits
- Hourly wages for temp staff
- Vendors
- Consultants
- Equipment / supplies
- Training/travel
- Indirect Costs
Hardware Options

- Scanners (flatbed, ADF, slide, etc.)
- Digital cameras
- AV conversion
- Servers (storage, delivery, etc.)
Software Options

- **In-house development:**
  - Requires programmers/database admins/metadata specialists
  - Requires constant maintenance/testing

- **Off-the-shelf software:**
  - E.g., Photoshop, DigiTool
  - Requires knowledge/evaluation/comparison of vendors
  - How will software be upgraded/maintained?
Sustainability Considerations

- Digital technology is constantly/rapidly changing
- Upgrade/replacement of software/hardware
- Migration of digital objects/metadata
- Costs of long term digital preservation
- Costs of staffing over time (managing content, hardware/software)
Exercise 1: Entire Group

- Handout includes a description of holdings to be digitized, target audience and other pertinent info
- Create goals for project
- List activities to achieve those goals
Exercise 1: Goals & Activities

- Goal 1:
Section 2: Implementing Digital Projects
Implementing Digital Projects

- Imaging Theory & Best Practices
- Digitization Equipment
- Metadata for Digital Resources
- Digital Asset Management Software
- Archiving and Preserving Digital Projects
Section 2.1: Imaging Theory & Best Practices
Imaging Theory & Best Practices

- Bit Depth & Color Space
- Resolution
- File Types
- Image Compression
- OCR
- Questions
Bit Depth & Color Space

- Bi-tonal, “black and white”, 1 bit
- Greyscales
  - 8-bit (256 shades of gray)
  - 16-bit (65536 shades of gray)
- RGB (usually 24-bit)
- CMYK (usually 32-bit)
Bit Depth & Color Space

RGB “built” from 3 color channels

Image: © Nevit Dilmen found at Wikimedia commons
Bit Depth & Color Space

- Color Fidelity
- “Full Informational Capture”

- Bi-tonal
- 8-bit Greyscale
- 24-bit Color

Meaningful color should be retained
Bit Depth... : Recommended

- (Almost) never scan 1-bit
- Completely grey items should (usually) be scanned 8-bit greyscale.
- Items with meaningful color should be scanned 24-bit RGB

- Trade-offs between quality and file size
Bit Depth... : Rationale

- **Text – Optical Character Recognition**

The Importance of Bit-Depth on Text Recognition: the Latin word *Feitis = Goodness*

1 Bit Image

This letter may be any of the following: *c-e-o-0*

8 Bit Image

This letter may be any of the following: *c-e-o-0*

24 Bit Image

The letter *e* appears now to be more probable.
Resolution

- Resolution of an image expressed in pixels
- PPI – pixels per inch
- DPI – dots per inch
## Resolution: Recommended

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Use For</th>
</tr>
</thead>
<tbody>
<tr>
<td>300 pixels per inch (ppi)</td>
<td>Printed text with normal sized fonts</td>
</tr>
<tr>
<td></td>
<td>Oversized documents and maps</td>
</tr>
<tr>
<td></td>
<td>Manuscripts with legible script</td>
</tr>
<tr>
<td>600 pixels per inch (ppi)</td>
<td>Photographs and select graphic arts</td>
</tr>
<tr>
<td></td>
<td>Printed text with very small fonts</td>
</tr>
<tr>
<td></td>
<td>Manuscripts with difficult scripts</td>
</tr>
</tbody>
</table>
Resolution: Rationale 1

- Newspaper graphics printed at 80 dpi
- Magazine graphics printed at 120 dpi
- High-end graphics printed at 300 dpi

- Scanning at 300 dpi is sufficient
Resolution: Rationale 2

- Text – Optical Character Recognition

Resolution and OCR Accuracy in high contrast images

- 75 ppi Image
  
  Label C of d Laid Pa

- 150 ppi Image
  
  Label C of d Laid Pa

- 300 ppi Image
  
  Label C of d Laid Pa

- 600 ppi Image
  
  Label C of d Laid Pa

OCR results

L ~ic ud L~ddPa
Resolution : Rationale 3

- Photographs
  - Use 600 dpi
  - Continuous-tone images
  - Unexpected use – capture all details
File Types

- Save archival masters as TIFF
- Internet delivery as JPEGs or JPEG2000s
Image Compression

- Save archival TIFFs as non-compressed
- “Lossy” vs. Lossless compression
OCR

- Optical Character Recognition
- Creation of plain text from an image file
- Just as important is the positional information!
  - Text highlighting
  - Text analysis
OCR : ALTO XML

- LOC XML schema / standard
- “Analyzed Layout and Text Object”
- Contains position (and style) of each word, with possible variants
- Can be embedded within a METS file
- Used by NDNP
Review of Topics

- Bit Depth & Color Space
- Resolution
- File Types
- Image Compression
- OCR
Questions?

How about born digital?
Section 2.2: Scanning Equipment
Scanning Equipment

- Flatbed scanners
- Sheet-feed scanners
- Book scanners
- Map scanners
- Microfilm
Flatbed Scanners

Microtek ScanMaker 9800XL
Epson Expression 10000XL
Sheet-feed Scanners

Panasonic KV-S2046C
Book Scanners

i2S CopiBook (24-bit color)

Konica Minolta PS7000
with grayscale up-grade
Oversized Document Scanners

Camera back, vacuum table, etc..

Betterlight Super 8K-HS
Microfilm Scanners
Section 2.3: Metadata for Digital Resources
Metadata for Digital Resources

- What is metadata?
- What is your digital resource?
- Standards
- Standards Comparison
- Tools
- Questions
What is metadata?
What is metadata?
What is metadata?
Library and Museum Metadata
Greek Vase

470 – 460 BC

35 centimeters

Odysseus and Eumaios the Swineherd

from Homer’s story of the Odyssey
Format: Greek Vase
Date: 470 –460 BC
Height: 35 centimeters
Title: [Greek Vase of Odysseus and Eumaiosthe Swineherd]
Notes: from Homer’s story of the Odyssey
Library and Museum Metadata
<table>
<thead>
<tr>
<th><strong>Title:</strong></th>
<th>L' Isle St. Domingue</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alternate Title:</strong></td>
<td>Atlas portatif Universel..</td>
</tr>
<tr>
<td><strong>Physical Desc:</strong></td>
<td>1 Map. : col; 17 x 22 cm on sheet 23 x 32 cm.</td>
</tr>
<tr>
<td><strong>Language:</strong></td>
<td>French</td>
</tr>
<tr>
<td><strong>Creator:</strong></td>
<td>Robert de Vaugondy, Gilles , 1688-1766</td>
</tr>
<tr>
<td><strong>Date:</strong></td>
<td>1749</td>
</tr>
<tr>
<td><strong>Place of Publ.:</strong></td>
<td>Paris 1749</td>
</tr>
<tr>
<td><strong>Subjects:</strong></td>
<td>Maps --Early works to 1800 --West Indies (lcsh)</td>
</tr>
<tr>
<td></td>
<td>Maps --Early works to 1800 --Hispaniola (lcsh)</td>
</tr>
<tr>
<td><strong>Genre:</strong></td>
<td>single map( marcgt)</td>
</tr>
<tr>
<td></td>
<td>Maps ( lcsh)</td>
</tr>
<tr>
<td></td>
<td>Early works to 1800 ( lcsh)</td>
</tr>
<tr>
<td><strong>Spatial Coverage:</strong></td>
<td>Haiti, Dominican Republic</td>
</tr>
<tr>
<td><strong>Note:</strong></td>
<td>Outline color.</td>
</tr>
</tbody>
</table>
Library and Museum Metadata

¿Avanzamos o retrocedemos?
Reflexiones sobre temas educativos

Radhamés Mejía
Este libro recoge una serie de artículos sobre temas educativos publicados en los años 2006 y 2007 en el periódico El Caribe.
Library Metadata

Economic theory, econometrics, and mathematical economics.

Ichihashi, Tatsuro.


x, 164 p. : ill. ; 24 cm. --
(Economic theory, econometrics, and mathematical economics)

Bibliography: p. 151-156.
Includes indexes.
ISBN 0-12-370180-5

1. Game theory. 2. Equilibrium (Economics) I. Title II. Series
What is your digital resource?
What is your digital resource?

- Physical archival containers
  - Archival box?
  - Archival folder?
  - Item-level?

- Scrapbooks
  - Pages
  - Photographs

- Newspapers
  - Issues
  - Article
Metadata Standards
Why metadata standards?
Standards

“The nice thing about standards is there are so many to choose from.”

- Andrew Tannenbaum (1988)
Types of Metadata Standards

- Bibliographic Description
- Metadata “Wrappers” and Transport
- Collection Descriptions
- Other Standards
  - Authority Standards
  - Proprietary Standards
  - etc…
Types of Metadata Standards

- Bibliographic Description
- Metadata “Wrappers” and Transport
- Collection Descriptions
- Other Standards
  - Authority Standards
  - Proprietary Standards
  - etc…
Bibliographic Description Standards

- MARC
- Dublin Core
- MODS
- VRA Core
Bibliographic Description: MARC

- MARC
  - Originally developed in 1960s by Library of Congress
  - Most embraced metadata standard in libraries

- MARC 21
  - Combination of USMARC and CAN/MARC
  - Metadata standard for 21st century

- MarcXML
  - Same MARC format, except encoded in XML
Bibliographic Description : MARC

- **Positives:**
  - Can encode a very granular amount of data
  - Very well adopted and works well with machine readers

- **Negatives**
  - Not very human-readable
  - High learning curve
  - Mixes data with display (as commonly implemented)
Bibliographic Description : DC

- **Dublin Core**
- Originally developed between 1994 and 1995 in Dublin Ohio by OCLC
- Used widely on web pages to assist search engines
- **Simplified Dublin Core**
  
<table>
<thead>
<tr>
<th>Title</th>
<th>Creator</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Publisher</td>
<td>Contributor</td>
</tr>
<tr>
<td>Date</td>
<td>Type</td>
<td>Format</td>
</tr>
<tr>
<td>Identifier</td>
<td>Source</td>
<td>Language</td>
</tr>
<tr>
<td>Relation</td>
<td>Coverage</td>
<td>Rights</td>
</tr>
</tbody>
</table>
Bibliographic Description : DC

- **Qualified Dublin Core**
  - Added three elements (Audience, Provenance, RightsHolder)
  - More importantly, added some basic refinements

- **Positives**
  - Widely accepted and easy to read and encode

- **Negatives**
  - Even with qualified Dublin core, difficult to encode complex data
  - Lack of data refinement leads to loss of UI options
Bibliographic Description : MODS

- **Metadata Object Description Scheme**
- Developed in 2002 by Library of Congress
- Beginning to be the de facto standard for digital libraries (MODS/METS)

**Positives**
- Can handle a (large) subset of MARC tags
- Handles complex objects and can easily be extended

**Negatives**
- Hard to do round trip portability from MARC → MODS → MARC
Bibliographic Description : VRA Core

- **Visual Resource Association**
- Developed in 1996 by Visual Resource Association
- Used for describing visual/cultural materials
- Includes all the standard tags
  - Date, Title, Description, Rights, Subject, etc..
- Also more unique tags
  - Cultural Context
  - Style
  - Technique
Bibliographic Description Standards

- MARC
- Dublin Core
- MODS
- VRA Core
Types of Metadata Standards

- Bibliographic Description
- Metadata “Wrappers” and Transport
- Collection Descriptions
- Other Standards
  - Authority Standards
  - Proprietary Standards
  - etc…
Metadata “Wrappers” and Transport

Contents
- Files and Structural data
- Administrative data

Choices
- METS
- OAI-PMH
- No wrapper necessary
Metadata “Wrappers” : METS

- Metadata Encoding & Transmission Standard
- Contains
  - Descriptions section(s)
  - Administrative section(s)
  - File section(s)
  - Structure map(s)
- Accepts any XML schema
Metadata “Wrappers” : OAI-PMH

- Open Archives Initiative - Protocol for Metadata Harvesting (2000-2001)
- Defines a protocol (over HTTP)
- Main Verbs
  - Identify
  - ListMetadataFormats – usually dublin core (99%)
  - ListSets
  - ListRecords
Types of Metadata Standards

- Bibliographic Description
- Metadata “Wrappers” and Transport
- Collection Descriptions
- Other Standards
  - Authority Standards
  - Proprietary Standards
  - etc…
Collection Metadata Standards

- (METS)

- EAD
  - Encoded Archival Description
  - Electronic Finding Guide/Aid
  - Main Sections
    - Descriptions
    - Container List – can link to digital objects
Types of Metadata Standards

- Bibliographic Description
- Metadata “Wrappers” and Transport
- Collection Descriptions
- Other Standards
  - Authority Standards
  - Proprietary Standards
  - etc…
Still to come...

- Authority Metadata Standards
  - NACO
  - MADS
  - EAC

- Other Standards
  - KML
  - DarwinCore
  - Z39.50 / ZING

- Standards are like weeds....
Proprietary and Custom Formats

- Proprietary File Formats
  - Greenstone Document File (doc.xml)
  - Fedora Object File (FOXml)

- Custom Schemas
  - SobekCM Schema (UFDC and dLOC)
Back to the beginning?

MODS?

Dublin Core?

KML?

VRA Core?

MADS?

EAC?

MARC?

EAD?

METS?

To wrap?
Evidence-based

- Universities within Florida (10 schools)
- Digital Objects
  - MARC (linked to catalog)
  - Dublin Core (no wrapper, simple objects)
  - METS wrapper (for complex items)
  - MODS
- EAD for Finding Aids / Guides
- OAI-PMH Harvesting
“One standard to rule them all”
“One standard to rule them all”

METS Wrapper
“One standard to rule them all”

METS Wrapper

DC $\rightarrow$ MODS

Additional custom metadata
“One standard to rule them all”

METS Wrapper

Dublin Core

Additional custom metadata
“One standard to rule them all”

METS Wrapper

MODS

Additional custom metadata
“One standard to rule them all”

METS Wrapper

- MODS
- VRA Core
- Darwin Core
- Additional custom metadata
“One standard to rule them all”

Archival

METS

MODS

SobekCM

Service

Greenstone
“One standard to rule them all”
Standards Comparison

Link to external comparison

http://digital.uflib.ufl.edu/development/standards_comparison.pdf
Lessons / Standards Conclusions

- Try not to tie your metadata to your system
  - Archival vs. Service

- Don’t be afraid of extending for your own needs, but continue to follow the standards

- Don’t cripple your metadata / Prepare for the most information you would want
  - METS / MODS v. simple Dublin Core
Tools
Metadata Tools

- Library of Congress site of METS utilities

- SobekCM METS Editor (UF)
  - http://ufdc.ufl.edu/metseditor

- MarcEdit Tool
Review of Topics

- What is metadata?
- What is your digital resource?
- Standards
- Standards Comparison
- Tools
Questions?
Exercise 2: Individual

- Handout includes descriptive info for an object that will be digitized
- Map from non-standard source metadata to Dublin Core
Section 2.4: Digital Asset Management System
DAMS

- ContentDM
- DigiTool
- Omeka
- Greenstone
- Fedora
- Dspace
- Custom (SobekCM)
Archival Information System

- Archon
- Archivist Toolkit
- ArchiveSpace
Section 2.5: Archiving and Preservation of Digital Projects
Digitization: One Puzzle Piece

- Digitization is one piece of a digital preservation plan, but only one piece.
- Without the rest of the puzzle, digitization is ultimately useless for preserving and archiving our historical and cultural resources.
- Include Digital Preservation and full lifecycle of the image and data in all digitization plans.
Preservation and Long-Term Access Through Networked Services Project Survey

- Surveyed over 200 organizations, mostly European archives and libraries
  - Digital content volume to increase 25-fold over next decade
  - 70% of respondents expect to hold over 100TB of content
  - Only 47% of organizations had an allocated budget for digital preservation
  - [http://www.planets-project.eu/publications](http://www.planets-project.eu/publications)

- Large-Scale Digitization
Digital Preservation Is... 

Digital preservation is:
- Active
- Constant process
- Exponentially multi-tiered
- About ensuring continuing accessibility
- Relatively new and somewhat in its infancy
Digital Preservation is NOT...

- Compact Disks
  - 2002 – 6,000 CD’s (2.6 TB)
  - 2008 – 23,000 CD/DVD (48 TB)
  - Unmanageable
  - Slowly deteriorating
- Two (or 3 or 4) copies
Table of Contents

- What to preserve? How do I know it is preserved?
- Object-level Preservation
- File-level Preservation
- Organizational: Trustworthy Digital Repositories
- Recommendations
- Questions
What to Preserve?

How do I know it is preserved?
What to preserve?
How do I know it is preserved?

- Text-level preservation?

Aujourd'hui premier janvier dix huit cent quatre, le Général en chef de l'Armée indigène, accompagné des généraux, chefs de l'armée, convoqués à l'effet de prendre les mesures qui doivent tendre au bonheur du pays:

Après avoir fait connaître aux généraux assemblés ses véritables intentions d'assurer à jamais aux indigènes d'Haïti un gouvernement stable, objet de sa plus vive sollicitude : ce qu'il a fait à un discours qui tend à faire connaître aux puissances étrangères la résolution de rendre le pays indépendant, et de jouir d'une liberté consacrée par le sang du peuple de cette île ; et, après avoir recueilli les avis, a demandé que chacun des généraux assemblés prononçât le serment de renoncer à jamais à la France, de mourir plutôt que de vivre sous sa domination, et de combattre jusqu'au dernier soupir pour l'indépendance.
What to preserve?  
How do I know it is preserved?

- Text-level preservation?

Today, January 1, 1804, the General in Chief of the Indigenous Army, accompanied by generals and army chiefs convoked in order to take measures tending to the happiness of the country:

After having made known to the assembled generals his true intention of forever ensuring to the natives of Haiti a stable government — the object of his greatest solicitude, which he did in a speech that made known to foreign powers the resolution to render the country independent, and to enjoy the liberty consecrated by the blood of the people of this island; and, after having gathered their opinions, asked each of the assembled generals to pronounce a vow to forever renounce France; to die rather than to live under its domination; and to fight for independence with their last breath.
What to preserve?
How do I know it is preserved?
What to preserve?
How do I know it is preserved?

- “It is the actual information content of a document, data-set, or sound or video recording that should be preserved, not the Microsoft Word file, or the Excel spreadsheet, or the QuickTime movie.”


- Also the context and layout though.
Object-Level Preservation
Object-Level Preservation

- More than just the images and text (but also!)

- Descriptive Information
  - Describes the intellectual entity (title, author)
  - Same information used for discovery and display of object

- Structural Information
  - Table of contents, page order

- File Information
  - File size
  - Checksums and digital signatures for fixity and authenticity checks

- Administrative Information
  - Provenance and access conditions for the object
Object-Level Metadata

- **METS**
  - Metadata Encoding and Transmission Standard
  - Descriptive Metadata Container
    - MODS (Metadata Object Description Standard)
    - Dublin Core
    - VRACore (Visual Resource Association)
    - MarcXML
  - Structural Metadata
  - Technical File Metadata
  - Administrative Metadata
PREMIS

- PREservation Metadata : Implementation Strategies
- Metadata needed to support preservation
- Contains what “most working repositories are likely to need to know” (Dappert and Enders)
- HUGE data dictionary – must pick and choose
- Describes historical preservation actions/agents
- Provides information needed for any future preservation actions
- Can be encoded in METS file
PREMIS
File-Level Preservation
Necessary Preservation File Information

- Example PREMIS Image Record
  - Checksums and Size
  - Preservation Level (full, bitwise, etc.)
  - Behaviors
  - Format (MIME, Format Registry, Validation Info)
  - Creating Applications
  - Capture Information (MIX)
  - Environments (software, operating system, processor, hardware)
  - All migrations events
Any sound preservation strategy must deal with the ephemeral nature of file formats

Migration v. Emulation

Must either constantly migrate forward or at least have a migration strategy

Migration strategy is also ephemeral and represents best current options
Recommended File Formats

- Four basic criteria
  - Specification must be freely available
  - Must be no patents or licenses on the format
  - Other digital repositories using or endorsed format
  - Variety of tools for writing and rendering format
Recommended File Formats

- Migration strategy is also ephemeral and represents best current options

- Archivematica (UNESCO preservation system)
  - Presented with a .WMV (Windows Media Video) proprietary file
  - Current strategy: Convert to MPEG-2
  - Future strategy: Convert to Motion JPEG2000
Organizational: Trustworthy Digital Repositories
Alexandria Library
Alexandria Library
Digital Preservation Community Process Models

- **Open Archival Information System (OAIS)**
  - Draft OAIS reference model published in 1999
  - Primary benchmark and chief process model for the preservation of digital assets

- **InterPARES**
  - Focuses on records maintenance

- **Digital Curation Centre: Curation Lifecycle Model**
  - Defines three levels of preservation and best practices for each
OAIS Reference Model

- Provides framework for understanding of archival concepts needed for long term digital information preservation and access
- Provides concepts needed by non-archival organizations to be effective participants in the preservation process
- Provides framework for describing and comparing different strategies and techniques
- Provides basis for comparing data models of digital information preserved and how they change over time
- Guides the identification and production of OAIS-related standards
An OAIS must:

- Negotiate for and accept appropriate information from producers
- Obtain enough rights to ensure long term preservation
- Determine the designated community
- Ensure preserved objects are independently understandable to designated community
- Follow documented policies and procedures which ensure the information is preserved against all reasonable contingencies
- Ensure object can be disseminated as authenticated copies of the original or the original
- Make preservation information available to designated community
Toward Certified Digital Repositories

Where are we?

- Granular standards (PREMIS)
- Process Models (OAIS, InterPARES)

In 2000, RLG and OCLC published *Trusted Digital Repositories: Attributes and Responsibilities*

This inches us closer, by defining a Trusted Digital Repository and expands the discussion to include the policies, standards, as well as technology infrastructure as a trusted system.
Developing Metrics for Certification

- Nestor Catalogue of Criteria for Trusted Digital Repositories
  - Network of Expertise in Long-term Storage of Digital Resources
  - Provides:
    - Information and criteria for self-assessment for digital repositories
    - Training Tools
    - Iterative process – Documentation, Transparency, Adequacy, and Measurability
Developing Metrics for Certification

- **DRAMBORA**
  - Digital Repository Audit Model Based on Risk Assessment
  - Digital Curation Center (DCC) and Digital Preservation Europe (DPE)
  - Focuses on risk identification, risk prioritization, and risk mitigation of risks through six step process of self-assessment.
  - Developed from a TRAC-based self-assessment originally
Developing Metrics for Certification

- **TRAC**
  - Trusted Repositories Audit & Certification Criteria & Checklist
    - Research Libraries Group (RLG) and U.S. National Archives and Record Administration (NARA)
    - 2005 - Grant from Andrew W. Mellon Foundation to develop procedures and activities to audit and certify digital archives.
    - 2007 - Final publication includes 84 criteria
      - Organizational infrastructure
      - Digital object management
      - Technologies (IT and security)
    - Two repositories (Portico and HathiTrust) initially audited
    - Portico certified as a trustworthy digital repository for CRL community
Toward Standards

- Conformity Assessment Requirements for Bodies Providing Audit and Certification of Management Systems (ISO/IEC 17021) is current international standard applied.

- Audit and Certification of Trustworthy Digital Repositories (CCSDS 652.0-R1, ISO/DIS 16363) still under development as an ISO standard.

- Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories also under development as an ISO standard.
Archiving
Recommendation
Recommendations

- Be aware of the ongoing cost of preservation and plan accordingly

- Be a part of something larger
  - Subject-based trusted repositories
    - LLMC (Law Library Microform Consortium)
  - Regional and International trusted repositories
    - HathiTrust
    - Florida Dark Archive / State Universities
Review of Topics

- What to preserve? How do I know it is preserved?
- Object-level Preservation
- File-level Preservation
- Organizational: Trustworthy Digital Repositories
- Recommendations
Questions?
Section 3: Evaluation & Project Wrap-up
Project Wrap-Up

- Evaluation of success/failure
- Promoting both products and the project itself
- Final report for funding agencies
- Future testing to improve products/services (e.g., usability, focus groups, surveys)
Project Wrap-Up: Questions

- What was final per-item cost?
- What would you change if you had it do over again?
- Were the products worth the expense/effort?
- What did we learn from failures/mistakes?
Assessing/Measuring Success

- Determine criteria for measuring success at start of project
- Choose method most appropriate for project (or required by funding agency):
  - Summative Evaluation
  - Outcomes-Based Evaluation
- Best evaluation is both during and after
- Difficult to assess if user needs met
Summative Evaluation

- Did we meet/not meet goals of the digital project?
- Did we meet goals on time?
- Were resources (staffing, budget) enough to get job done?
- Includes evaluation of workflow and activities
- Usually includes both quantitative and qualitative measures
Outcomes Based Evaluation

- Shows if goals of the digital project were met/not met
- User-centric (how have patrons benefitted)
- Focus is on impacts/benefits of project rather than on activities
- Includes evaluation of staffing, workflow, budget (were they enough to get job done)
- Includes quantitative and qualitative measures
- Encouraged by many grants
Components of Outcomes-Based Evaluation

- Inputs/Resources
- Activities
- Outputs/Products
- Quantifiable outcomes (e.g., # of users able to access holdings will increase)
- Evaluation indicators
- Methods/Sources for Evaluation
Promotion/Publicity

- Fairly standard for most projects
- Traditional methods: newsletters, releases, listservs, presentations, journal articles, etc.
- Newer methods: facebook, blogs, youtube, flickr, other social networking
Questions?
Exercise 3: Entire Group

- Use list of project goals/activities created in Exercise 1
- Create evaluation plan (summative or outcome-based)