Digital Preservation Open Forums
March-April, 2016

Cinda May, Indiana State University
Connie Rendfeld, Indiana State Library
Digital Preservation
Open Forums
March-April, 2016

Cinda May, Indiana State University
Connie Rendfeld, Indiana State Library

Today’s Agenda

- Welcome
- IN DPLA Update
- Digital Preservation Readiness Survey Results
- Digital Preservation 101 – What You Need to Know
- Preserving Indiana’s Memories: Introducing IndiPres, Indiana Digital Preservation

Digital Preservation Readiness Survey
September 2015

Courtesy of the Vigo County Historical Society
Digital Preservation 101: What You Need to Know

What is Digital Preservation?

- **Short Definition:** Digital preservation combines policies, strategies and actions that ensure access to digital content over time.
- **Medium Definition:** Digital preservation combines policies, strategies and actions to ensure access to content that is born digital or converted to digital form regardless of the challenges of file corruption, media failure and technological change. The goal of digital preservation is the most accurate rendering possible of authenticated content over time.

(ALCTS Preservation and Reformatting Section, Working Group on Defining Digital Preservation)

Why Preserve Digital Content?

Digital media does not have the same lifespan as print media and is far more fragile.
3 Primary Risk Factors for Digital Content

- Change and loss – accidental and intentional
- Obsolescence – as technology evolves
- Disasters – emergencies of all kinds

Where Do You Start?

- Inventory Your Collection
  - What Categories of Digital Content Do You Have? (Photographs, Documents, A/V, Data Sets, etc.)
  - Are there Analog Copies; or, is it Born Digital Content?
  - How Much Digital Content Do You Have? (MB/GB/TB)
  - What Format Types Are Represented? (TIFF, JPEG, PDF, DOC, WAV, etc.)
  - Where Is Your Digital Content Stored?

Creating an Inventory

- Decide who will be responsible for the inventory and who will need access to it
- Choose a database or spreadsheet software that is easy to use
- Create a simple format to document the information you want to collect
- Make certain that you will be able to edit or add information beyond the initial inventory
- Keep the inventory up-to-date
What Information Should You Record?

- Content Categories: Institutional records, Genealogy, Local History/Special Collections; Research data; Scholarly content (licensed and open), Locally created databases, etc.
- Descriptive Title of Content
- Format of Content
- Extent of Content
- Date of Creation, Date of Receipt, Date Range Coverage, Date of Inventory
- Location of Content: Online, Offline, 3rd Party

Sample Inventory

<table>
<thead>
<tr>
<th>Inventory Date</th>
<th>Category</th>
<th>Title/Description</th>
<th>Creation Date(s)</th>
<th>Location</th>
<th>Extent</th>
<th>Format(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-10</td>
<td>Special Collections</td>
<td>Eugene V. Debs Correspondence</td>
<td>1895-1926, Aligned 2008</td>
<td>Library: dark online server, production server</td>
<td>52 GB</td>
<td>Image/TIFF, Image/EPWs</td>
</tr>
<tr>
<td>2006-12</td>
<td>Institutional Repository</td>
<td>ETDs</td>
<td>2006-2016</td>
<td>Library: DSpace server, BackArchives</td>
<td>30 GB</td>
<td>PDFs</td>
</tr>
<tr>
<td>2015-08</td>
<td>Institutional Repository</td>
<td>ETDS</td>
<td>2015-2018</td>
<td>Library: DSpace server, BackArchives</td>
<td>30 GB</td>
<td>PDFs</td>
</tr>
</tbody>
</table>

What Needs to be Preserved?

- Not Everything Needs to be Preserved!
- Consider
  - What Items are Most Significant
  - How Often are they Requested by Patrons
  - How Old or New are the Items
  - How Extensive is the Overall Collection
Define Your Selection Criteria

- Ask Yourself
  - What is Central to Our Mission?
  - What is Our Collection Development Policy?
  - What are We Required to Keep and for How Long?
  - Is the Resource Unique?
  - Is the Resource Preserved Elsewhere?
  - What is the Value of the Resource?
  - Who Owns the Rights?
  - Who Needs to be Involved in the Selection Process?

As You Review Each Resource

- Determine if it Fits Your Established Selection Criteria
- Decide if it is Feasible for You to Preserve the Content
  - Do You Possess the Technical Infrastructure and Skill Sets?
  - Should You Outsource the Preservation of the Resource?
- Determine Your Ability to Provide Long Term Access
- Be Consistent in Your Review
- Document Your Decision(s) in Your Inventory

Preparing Digital Content for Archival (i.e. Long Term) Storage

- File Naming Conventions
- Formats
- Metadata
- Number of Copies
Santaguida, Vincent. Folder and File Naming Convention
10 Rules for Best Practice (2010)

- Avoid extra long folder names and complex hierarchical structures but use information-rich filenames instead.
- Put sufficient elements in the structure for easy retrieval and identification but do not overdo it.
- Use the underscore (_) as element delimiter. Do not use spaces or other characters such as: ! # $ % & _ @ ^ ` - + , ; = ] { }
- Use the hyphen (-) to delimit words within an element or capitalize the first letter of each word within an element.
- Elements should be ordered from general to specific detail of importance as much as possible.

Santaguida, Vincent. Folder and File Naming Convention
10 Rules for Best Practice (2010)

- The order of importance rule holds true when elements include date and time stamps. Dates should be ordered: YEAR, MONTH, DAY. (e.g. YYYYMMDD, YYMMDD, YYYYMM). Time should be ordered: HOUR, MINUTES, SECONDS (HHMMSS).
- Personal names within an element should have family name first followed by first names or initials.
- Abbreviate the content of elements whenever possible.
- An element for version control should start with V followed by at least 2 digits and should be placed as the last most element. To distinguish between working drafts (i.e. minor revisions) use Vx-
(0)-Vx-
(99) range and for final draft (i.e. major version release) use Vz-
(0)-Vz-
(99), (where x = 0-9).
- Prefix the names of the pertinent sub-folders to the file name of files that are being shared via email or portable storage devices.

File Name Conventions
North Carolina Dept of Cultural Resources.

- Rule #1: Avoid using special characters in a file name. \ / * ? " < > | [ ] $ .
- Rule #2: Use underscores instead of periods or spaces.
- Rule #3: Err on the side of brevity. 25 characters or less
- Rule #4: The file name should include all necessary descriptive information independent of where it is stored.
- Rule #5: Include dates and format them consistently.
- Rule #6: To more easily manage drafts and revisions, include a version number on these documents.
- Rule #7: Be consistent.
The Importance of Metadata

- Metadata Enables Long Term Digital Preservation
  - Metadata is Unique to the Resource It Describes
  - Metadata Makes the Resource Understandable
  - Metadata Allows A Resource to be Traced Over Time

Number of Copies

How many copies are enough for you?
Minimum: 3 copies in 3 locations
Optimum: 7 copies in 7 locations

Examples of storage factors:
- Video files are too large to store 7 copies
- Possible legal restrictions
- Types of media used for storing the content

Storage Factors to Consider

- Cost (available resources for preservation)
- Quantity (size and number of files)
- Expertise (skills required to manage)
- Partners (achieving geographic distribution)
- Services (outsourcing)
Available Preservation Services

One of Google's server farms in Council Bluffs, Iowa, which provides over 115,000 square feet of space for servers running services like Search and YouTube

Selecting a Digital Preservation Solution

If you decide to use (build, join, buy) a repository

- No system is fully compliant to standards
- Range of types to consider:
  - general (any content) to special (format-specific)
  - open source to proprietary
  - easy to advanced installation and management
- Select best option for your content — for now
- Each option has pros and cons
Getting to the Bottom Line: 20 Cost Questions for Digital Preservation


Preserving Indiana's Memories: Introducing InDiPres, Indiana Digital Preservation

The mission of Indiana Digital Preservation (InDiPres) is to collaboratively manage and sustain a low-cost, secure, and geographically distributed archive for the long term preservation of locally sponsored digital resources in Indiana.
Membership

Any Indiana institution creating digital assets whose activities and objectives are consistent with Indiana Digital Preservation's mission and principles may join InDiPres. This includes but is not limited to universities, libraries, museums, historical societies, research centers, agencies of state and local government, and individual projects. The InDiPres membership year begins on July 1. Membership fees are due during the month of July.

Proposed Membership Model

Indiana State Library serves as lead institution with the ISL Foundation serving as the fiscal agent.

Indiana State University Library manages the LOCKSS storage cache on behalf of the participants in the collaborative.

Institutions pay $100.00/year to participate.

Participants share the cost of the MetaArchive Collaborative Membership ($2,500/year/3 year minimum).

Participants share the cost of the server housed at ISU (approximately $6,000/year with a 3 year replacement cycle).

Participants pay for individual storage needs ($0.59/GB/year).

Goal: To keep the overall cost per participant to $325.00 annually (or less) plus individual storage fees.

Goal: To make the Indiana Collaborative self-sustaining after 3 years.

Proposed Membership Fee Schedule

Based on a Minimum of 20 Participants

<table>
<thead>
<tr>
<th>Fee Type</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Collaborative Membership Fee</td>
<td>$100.00/year</td>
</tr>
<tr>
<td>Share of Server Cost</td>
<td>$100.00/year</td>
</tr>
<tr>
<td>(3 year replacement cycle, $6,000/year/20 participants)</td>
<td></td>
</tr>
<tr>
<td>Share of MetaArchive Collaborative Membership Fee</td>
<td>$125.00/year*</td>
</tr>
<tr>
<td>($2,500/year/20 participants)</td>
<td></td>
</tr>
<tr>
<td>+ Individual storage fee = $0.59/GB/year</td>
<td></td>
</tr>
<tr>
<td>Total set costs</td>
<td>$325.00/year</td>
</tr>
</tbody>
</table>

*As the number of participating organizations increases, the share of the annual MetaArchive membership fee will decrease (e.g. $2,500/25 = $100/year)
Benefits for Indiana Cultural Heritage Organizations

- Participation in a community-based digital preservation solution
- Participation in the governance of a statewide collaborative
- Direct involvement in the stewardship of local digital assets
- Distributed archiving across multiple, geographically distributed preservation sites
- Retrieval of content in case of catastrophic loss
- Access to educational opportunities related to digital preservation
- Opportunities to serve on committees and working groups
- Cost-sharing promotes affordability and long-term digital preservation for all

Thank You for Your Interest!

Cinda May
Chair, Special Collections
Indiana State University Library
812-237-2534
Cinda.May@indstate.edu

Connie Rendfeld
ISL Digital Initiatives Librarian
317-232-3694
crendfeld@library.in.gov

Costs

Indiana Digital Preservation requires minimal expenditures by its Members. The annual membership fee will be based on the following factors:

Members pay $100/year to participate in InDiPres.
Members share the costs of the MetaArchive Collaborative Membership ($2,500/year/3-year minimum).
Members share the cost of the server housed at ISU (approximately $6,900 with a 3-year replacement cycle).
Members pay for individual storage needs ($0.59/GB/year).

Example: the membership fee would be $325/year plus the cost of individual storage if there are 20 members.

New members will pay a pro-rated amount. Contact InDiPres for actual costs.
Indiana Digital Preservation (InDiPres)
Membership Agreement

Purpose
This Membership Agreement outlines the initial agreements that form the basis for membership in InDiPres.

Mission
The mission of Indiana Digital Preservation (InDiPres) is to collaboratively manage and sustain a low-cost, secure, and geographically distributed archive for the long term preservation of locally sponsored digital resources in Indiana.

Host Institutions
The Indiana State Library will serve as the lead institution for InDiPres, which will join the MetaArchive Cooperative Preservation Network as a collaborative member. The Indiana State Library Foundation will serve as the fiscal agent. Indiana State University Library will provide the ingest point for digital content submitted through InDiPres, as well as manage the LOCKSS server that will be a node in the MetaArchive network.

Membership
Any Indiana institution creating digital assets whose activities and objectives are consistent with Indiana Digital Preservation’s mission and principles may join InDiPres. This includes but is not limited to universities, libraries, museums, historical societies, research centers, agencies of state and local government, and individual projects. The InDiPres membership year begins on July 1. Membership fees are due during the month of July.

Costs
Indiana Digital Preservation requires minimal expenditures by its Members. The annual membership fee will be based on the following factors:

Members pay $100/year to participate in InDiPres
Members share the cost of the MetaArchive Collaborative Membership ($2,500/year/3 year minimum)
Members share the cost of the server housed at ISU (approximately $6,000 with a 3 year replacement cycle)
Members pay for individual storage needs ($0.59/GB/year)
Example: the membership fee would be $325/year plus the cost of individual storage, if there are 20 members.

New members will pay a pro-rated amount. Contact InDiPres for actual costs.
Applications for Membership
InDiPres will consider applications for membership from qualified institutions. Applications for membership will be reviewed by the InDiPres Steering Group, and applicants may be asked to submit additional information about their digital collections and technical capabilities before the application is considered for approval. Election to membership requires a two-thirds favorable vote by the InDiPres Steering Group. To complete its membership, a newly elected Member must sign and submit an InDiPres Membership Agreement affirming its acceptance of the conditions listed.

Copyright and Intellectual Property
All Members bear the responsibility for determining ownership and their right to preserve content prior to submitting it for archival storage in InDiPres (i.e., intellectual property and copyright issues). Unauthorized duplication or distribution of cached content is a violation of the governance policy.

InDiPres strives diligently to comply with the spirit of the 1976 Copyright Law (title 17, U. S. Code), related guidelines issued by the U. S. Copyright Office, and subsequent U.S. laws affecting copyright. InDiPres is a dark archive. This means that it is not accessible outside of preservation routines and is available only to the InDiPres Members for purposes of preservation and to replace the originating institution’s local files when necessary. Each originating institution bears responsibility for handling copyright and intellectual property issues locally.

Indemnification
All Members agree to hold InDiPres and its Members harmless in the event of infringement, claims of infringement, loss of data, interoperability, and any other technical standards and governance claims by waiving any rights of recovery for any costs or damages associated with their relationships to and Agreement with InDiPres. Likewise, all Members agree to indemnify InDiPres and its Members to the extent permitted by law for any losses and costs incurred by InDiPres and its Members such as but not limited to legal fees, costs, and damage awards arising from infringement or other claims directly related to their activities in working with InDiPres and Members.

Withdrawing from InDiPres
All Members share some rights and responsibilities in common, including agreeing to retain membership through the end of the full three-year cycle with a one-year notice to cancel membership thereafter. This time frame provides some predictability for InDiPres as it continues to grow and will help improve services to all Members. Following the completion of its initial term (three years or more), any Member may withdraw from InDiPres by notifying the InDiPres Steering Group in writing of its intention to withdraw. It is understood that all digital content contributed to the MetaArchive Preservation Network by InDiPres Members for preservation purposes may remain in the Network indefinitely. The MetaArchive Steering Committee reserves the right to remove content from the MetaArchive Preservation Network, in, but not limited to, such cases as: a) content is deemed unsuitable; b) content was submitted without proper copyright clearance; or c) if the submitting Member fails to fulfill its responsibilities to the Cooperative as stated in this Charter and in the signed Membership Agreement.
Disaster Recovery

In the event of data loss by an InDiPres Member, that Member may recover the content it has submitted to the MetaArchive Preservation Network through InDiPres. InDiPres and the MetaArchive Preservation Network are not intended as substitutes for a robust local backup and recovery regimen. However, in the event of data loss InDiPres will work with MetaArchive Central Staff to recover a member’s data in full from the MetaArchive Preservation Network and archive(s) to which it has contributed content. Data available for recovery includes the metadata and the digital objects that the InDiPres Member has submitted to the MetaArchive Preservation Network through InDiPres and that the MetaArchive Members have successfully ingested. Recovery time is dependent on the size of the collection to be recovered and restored.

To ensure that all data submitted to the MetaArchive network is successfully ingested, and thus is available for extraction and replication by a Member, the MetaArchive Cooperative makes reports available to Members, annually and upon request, through an auto-generated, web-based tool. These reports relay information concerning what content has been ingested and by which caches. It is each Member’s responsibility to track and audit the integrity of its own collections using these report tools (e.g., if the report shows that a submitted collection has not yet been ingested by at least six caches, the Member should bring this to the attention of the MetaArchive Cooperative’s Central Staff).
Signatures Page

This page should be completed by an individual with organizational signature authority for the organization joining InDiPres. Signature indicates agreement to the terms and conditions of the document entitled "Membership Agreement" of which this is the last page.

Upon completion of this form, an invoice will be sent separately to your organization.

Please print clearly or type information in the spaces provided below, and return this signature page by mail to InDiPres, Indiana State Library, 140 N. Senate Ave., Indianapolis IN 46204.

Member Information:

<table>
<thead>
<tr>
<th>Organization</th>
<th>Billing Address</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Authorized Representative’s Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Representative’s Name (print)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contact Person</th>
<th>E-mail/Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Membership Fee</th>
<th>Effective Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Signed by Authorized Representative of InDiPres: ____________________________

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The NDSA Levels of Digital Preservation: An Explanation and Uses


Abstract

The "Levels of Digital Preservation" being refined now by the National Digital Stewardship Alliance (NDSA), is a tiered set of recommendations on how organizations should begin to build or enhance their digital preservation activities. A work in progress, it is intended to be a relatively easy-to-use set of guidelines useful not only for those just beginning to think about preserving their digital assets, but also for institutions planning the next steps in enhancing their existing digital preservation systems and workflows. It allows institutions to assess the level of preservation achieved for specific materials in their custody. It is not designed to assess the robustness of digital preservation programs as a whole since it does not cover such things as policies, staffing, or organizational support. The guidelines are organized into five functional areas that are at the heart of digital preservation systems: storage and geographic location, file fixity and data integrity, information security, metadata, and file formats.

This paper presents the Levels, explains the context of the project's development within the NDSA, describes the rationale behind each of the guidelines and why they were prioritized the way they were, suggests how the guidelines may be used, and compares and contrasts the Levels to other ways of assessing stages of digital preservation. Other assessment models include Nancy McGovern and Anne Kenney's "The Five Organizational Stages of Digital Preservation," Charles Dollar and Lori Ashley's "Digital Preservation Capability Maturity Model," and OCLC Research's 2012 report, "You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media." Finally, the paper requests feedback on the work in progress and outlines planned future work.

Introduction and Background

The National Digital Stewardship Alliance (NDSA), a diverse group of over 140 organizations whose mission is to "establish, maintain, and advance the capacity to preserve our nation's digital resources for the benefit of present and future generations" [1] has recently developed the NDSA Levels of Digital Preservation. The Levels of Digital Preservation are a tiered set of guidelines and practices intended to offer clear, baseline instructions on preserving digital content at four progressive levels of sophistication across five different functional areas. The recommended activities within the Levels are agnostic towards both content type and technology, focused on specific preservation actions (as opposed to organizational requirements), and are designed to offer a practical blueprint that can be utilized by institutions of all sizes and resource levels to perform digital preservation. The primary goal of the Levels of Digital Preservation chart is to meet the need for straightforward, accessible practices that are more substantial than the conventional digital archiving advice geared towards individuals, but less daunting and demanding than those required for certification as a trustworthy digital repository. This paper describes the Levels of Digital Preservation's origins and development within the NDSA, explains its purpose and goals, reviews related digital preservation models, presents the levels, and explicates them. The paper includes suggestions for using the Levels and implementing its activities. The paper closes with future plans for encouraging further community feedback and supporting the continued evolution and refinement of the Levels.

At the core of the Levels of Digital Preservation's creation and development is the collaborative spirit that underpins the NDSA. As an alliance composed of a variety of institutions, from large research universities to small cultural heritage institutions, from non-profit organizations to commercial partners, the NDSA offers an ideal environment to develop a resource beneficial to the varied types of individuals and institutions operating or hoping to initiate digital preservation programs. The NDSA contains a multiplicity of skills, expertise, and experience in a diverse membership dedicated to the many tasks and responsibilities involved in digital stewardship. This diversity was invaluable in conceptualizing and articulating the goals and final form of the Levels of Digital Preservation. This project was also notable as being the first NDSA-wide collaboration, as it featured members drawn from all five of the NDSA's Working Groups: Content, Standards, Infrastructure, Innovation, and Outreach. The team working on this project thus reflected the diversity of the NDSA membership and it ability to work in cooperative, cross-discipline groups.

The project to define levels of digital preservation originated when a number of different NDSA members recognized the need for practical, actionable, and scalable digital preservation guidance that was accessible both to those just getting started and those with fully implemented preservation programs. In informally surveying NDSA members, as well as researching existing digital preservation models, the project team identified a number of desired goals for the Levels of Digital Preservation. The team wanted the Levels to feature actions independent of specific formats, content types, and storage systems, thus enhancing their usability across domains. The team wanted the levels of digital preservation to be comprehensive in scope, but also simple and practical as far as specific actions. The Levels should be able to
<table>
<thead>
<tr>
<th><strong>Table 1: Version 1 of the Levels of Digital Preservation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1 (Protect your data)</strong></td>
</tr>
</tbody>
</table>
| **Storage and Geographic Location** | - Two complete copies that are not collocated  
   - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system | - At least three complete copies  
   - At least one copy in a different geographic location  
   - Document your storage system(s) and storage media and what you need to use them | - At least three copies in geographic locations with different disaster threats  
   - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems |
| **File Fidelity and Data Integrity** | - Check file fixity on ingest if it has been provided with the content  
   - Create fixity info if it wasn't provided with the content | - Check fixity on all ingests  
   - Use write-blockers when working with original media  
   - Virus-check high risk content | - Check fixity of content at fixed intervals  
   - Maintain logs of fixity info; supply audit on demand  
   - Ability to detect corrupted data  
   - Virus-check all content |
| **Information Security** | - Identify who has read, write, move and delete authorization to individual files  
   - Restrict who has those authorizations to individual files | - Document access restrictions for content | - Perform audit of logs |  
   | **Metadata** | - Inventory of content and its storage location  
   - Ensure backup and non-collocation of inventory | - Store administrative metadata  
   - Store transformative metadata and log events | - Store standard technical and descriptive metadata |  
   | **File Formats** | - When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs | - Inventory of file formats in use | - Perform format migrations, emulation and similar activities as needed |  
   |

**General Structure of the Levels: Categories and Tiers**

The overall structure of the chart is progressive -- the actions in the first level are either necessary prerequisites for those in the second to fourth levels or are themselves the most pressing activities to accomplish first. The five general categories (Storage and Geographic Location, File Fidelity and Data Integrity, Information Security, Metadata and File Formats) were agreed upon early in the project. These areas were identified by the Levels team as the broad conceptual areas of focus for thinking through technical and immediate threats to digital preservation. In this respect, these categories are the categories that the subject matter experts on the Levels team use to categorize their own work. This is how the team members describe the risks and threats that they work to mitigate.

In relation to other work, some readers might ask why issues with rights and/or policies have been excluded. From the start, the team was primarily concerned with technical issues; the goal was to identify the technical functions and features one would want to see occurring somewhere to ensure long term access to digital content, not the social or legal structure that would be in place to
sustain those activities. Again, the goal of this project is not to provide a plan for digital preservation but to provide a chart to help anyone interested in long term access to digital information evaluate how they are doing in terms of mitigating risk of loss and identify concrete technical next steps they can take to move all or part of their operation to the next level.

Broadly speaking, as one moves up each of the tiers from Level 1 to Level 4, one is moving from the basic need to ensure bit preservation towards broader requirements for keeping track of digital content and being able to ensure that it can be made available over longer periods of time. While the names for the five general categories in the grid were agreed upon early in this work, there was difference of opinion on the extent to which the labels for each of the levels (Protect Your Data, Know Your Data, Monitor Your Data, and Repair Your Data) should be included. Some in the team wanted to leave the labels out and strictly work to organize the document according to the perceived biggest risks to mitigate loss. Others in the group felt that the labels helped conceptually organize the grid and helped to explain the general overarching goals of each level. The conceptual value of the categories won out, and remains part of the chart. However, it is important to note that the labels applied to each level are rough characterizations and not edicts about exactly what should go in a given level. In any case when the conceptual purity of ordering particular activities was in conflict with the pragmatic realities of what the team thought needed to be addressed first, the team sided with the pragmatic action over conceptual purity.

**Detailed Explanation of Levels and Tiers**

In what follows the reasoning behind each of the particular features of each individual level is briefly articulated.

**Storage and Geographic Location**

The first factor on the grid focuses attention on the storage of digital information. As one moves up the levels one is keeping additional copies, which helps to hedge against threats of loss due to bit rot and failures in storage media and systems. Similarly, as one moves up the levels one incorporates additional geographic locations to hedge against regional threats (like natural and manmade disasters) to storage systems. At the very base level, the first step one should take to ensure access to materials in the future would take is to create a second copy. Thus that requirement is the first item on the chart.

Aside from these two general trends across this category, the first level asks that one get data that comes in on heterogeneous media (optical disks, external hard drives, etc.) off of the removable media and into a storage system. The team identified this as an essential first step, as this kind of heterogeneous storage media is at risk of failure and requires significant manual effort to ensure data integrity. Further, this first step is necessary as a means to enable the kinds of preservation action required in many other parts of the levels document. The term "storage system" is intentionally vague as the team did not want to focus too much on any particular storage technology. Given the nature of the full set of requirements in the levels document, storage system should generally be understood as either a nearline or online system using either all spinning disk or some combination of spinning disk and magnetic tape.

Levels 2, 3, and 4 have additional requirements that focus on ensuring the longevity of storage systems: first requiring documentation of the system, then requiring an obsolescence monitoring process for storage systems and media and finally a comprehensive plan for keeping content on currently accessible media or systems. The intention in making these steps incremental is largely to spread out a set of activities that would all be nice to have but which become increasingly complex and require the work of the previous level to be possible.

**File Fixity and Data Integrity**

One of the most essential components of digital preservation is being able to attest to the fixity and integrity of the materials being preserved. This is a foundational component of digital preservation, but for many organizations checking the fixity of content remains a challenge. The goal of this category is to provide a series of steps that will take an organization to a stage where it is acting robustly to ensure the fixity of their content.

At the first level, the recommendation is simply to check the fixity on ingest if fixity information is provided for content (likely MD5 or SHA-1 cryptographic hashes) or to generate fixity information if none were provided. This is a necessary first step for an organization to validate that the content they preserved is what they intended to preserve. Many organizations are accomplishing this by making use of tools like Bagger [6] or usage of the BagIt specification [7] to package digital content.

From there, the next levels bring in additional activities to help further ensure the integrity of content. Most notably, Level 2 requires fixity checks on all ingests, and Levels 3 and 4 move into increasingly strong requirements for ongoing checking of digital content. The requirements in Levels 3 and 4 shift from placing trust in the quality and performance of particular storage media and shift to thinking of preservation as being ensured through repeated ongoing checking of content. This provides the added level of assurance and the ability to confidently assert the fixity of content one is stewarding.

**Information Security**

The information security section focuses primarily on understanding who has access to content, who can perform what actions on that content and enforcing these access restrictions. It starts with basic and simple steps to identify who can do what to the content. This is essential as without having procedures in place to restrict what can be done with content, one invites the risk of someone incorrectly deleting content. From there, Level 2 progresses to access restrictions. Level 3 suggests keeping logs of actions, which helps to bring an organization’s approach in line with archival best practices. Level 4 brings in the added requirement of auditing logs of those actions which helps to double check that intended activities and actions are actually happening.

Like many of the other sections, these levels were arrived at largely by establishing what one would need to have in place as a prerequisite for more advanced requirements and calibrated to minimize risk in relation to the other risks the team perceived across the other categories.
Metadata

While issues related to metadata appear in many of the other levels, it was decided that it was critical to give the issue its own row in the chart. The team defines metadata broadly, including everything from inventory information about the location of files, broader sets of administrative metadata (for example when and how it was created, and who can access it), transformative metadata documenting and logging events that have resulted in changes in objects, to technical and descriptive metadata, and ultimately preservation metadata.

In organizing the levels in the order presented the team suggests the most essential metadata at the lower levels and at higher levels, the additional layers of metadata that will make content both better protected and more identifiable and accessible. It is worth noting, that in most systems nearly all of this metadata (with the exception of descriptive metadata) can and should be generated and processed computationally and not manually.

File Formats

Digital objects are intimately dependent on the structure and nature of their file formats. The file formats section of the chart is the section that underwent the most change and revision in the process of public review of the Levels. The recommendation team settled on allows for the possibility that not all formats will need migration or emulation, but that at Level 4, preservation interventions will be actively pursued for any formats that require them.

The first level simply suggests that organizations, when possible, encourage the use of limited sets of known and open file formats. This is particularly the case in situations where an organization is digitizing material and has considerable say in what formats to use. For that kind of work, authoritative sources such as the Federal Agencies Digitization Guidelines [8] should be consulted for additional file format advice. With that noted, in many other contexts, including collecting born digital archival material from heterogeneous tangible media or web archiving it would not be feasible to force any changes in file formats as a basic requirement.

Successive levels begin to document the formats in use, monitor them for obsolescence issues, and ultimately to engage in migrations, support emulation, or look into other modes of ensuring that preserved content is usable and accessible in the future. The team placed actions based on format obsolescence in the fourth level for several reasons. First, while obsolescence is a formidable problem, it requires one to have made it through the hurdles of basic bit preservation and data management. In short, if a file cannot be opened, it is still a file in one’s possession. Aside from this, as issues around migration and emulation are areas of extensive ongoing research and development it is likely that while an organization is addressing issues related to the first three levels there will be substantive advances in working with some of the file formats they are stewarding. It’s better to get one’s house in order first and then join in ongoing discussions of when particular approaches to migration and emulation are applicable to particular goals.

Using the Guidelines

The initial use case envisioned by the Levels team was as a reference for organizations to consult when prioritizing enhancements to digital preservation systems. Although the guidelines are still relatively new, they are proving not only to be useful for that purpose, but also in unanticipated ways. Other possible uses for the preservation levels are described in this section.

Use: Identify where there is and isn’t general consensus in the preservation community

To get to this first version of the Levels, there was a great deal of discussion and debate among the NDSA members who produced it. Numerous versions of the Levels were produced and modified based on these discussions. During the course of this work, the Levels evolved into a product that still had a few items without unanimous support, but in general the team was behind the guidelines. After the Levels were published to the Web, another round of discussion and debate ensued, this time including practitioners from around the world, over important topics such as the usefulness (or not) of validating file formats, whether or not format normalization or ingest is an activity we should all be striving to implement, and how many different copies need to be in locations with different disaster threats. Where there seemed to be consensus, the suggestions were incorporated into what became version 1. It is expected that as the Levels are reviewed and debated in additional forums over time, and the consensus is incorporated into revisions, they will continue to reflect the community’s digital preservation best practices.

Use: Educate and develop guidelines for content creators and contributors

A thorough reading of the Levels chart reveals that there is a direct relationship between the activities and efforts of content creators (e.g. use of open formats and codecs, degree of content description) and the level of preservation service that can be provided for the content. For example, if content creators supply some descriptive metadata about the content, there is the possibility that the content can receive Level 3 service. An organization could use this chart directly as an educational tool or transform its information into guidelines to show content creators how they can contribute to better preservation for their content.

Use: Validate preservation guidance given locally

One person giving feedback on version 1 of the Levels said, “It’s the kind of thing we need at the NPS to be able to demonstrate to practitioners and managers that we aren’t just ‘making things up’ when it comes to preservation recommendations.” [9] Because the guidelines presented in the Levels chart were produced by digital preservation practitioners, it is likely that they will intersect in whole or part with recommendations and advice given locally within our institutions. Where this occurs practitioners can point to the Levels chart as evidence that their local advice is in sync with the larger preservation community’s thinking and practices.
Use: Develop requirements for third-party preservation service providers

The Levels chart defines some of the core minimum requirements for preservation. These guidelines could be re-expressed from a content holder's perspective as requirements when soliciting or negotiating preservation services from external companies or organizations. For example an RFP could specify that all of the activities described in the File Fixity and Data Integrity row be implemented.

Use: Assess compliance with preservation best practices and identify key areas to improve

The Levels chart looks deceptively simple but in actuality it can support multiple types of assessments. The unit of assessment is flexible. It can be used to assess the preservation capabilities of an entire preservation repository or one component of the repository (e.g. storage). Or it can be used to assess the degree of preservation received by particular collections or streams of content. The portion of the chart referenced can vary also. The Level 1 column can be used alone to reference recommended first steps. Or a single row could be referenced to drill down in a particular area (e.g. just “Storage and Geographic Location”). Alternatively the full chart could be used to do an overall assessment.

Unlike some of the other assessment models, the result of an assessment using the Levels isn’t likely to be a single score, e.g Level 2. The chart is composed of five different functional areas (Metadata, etc.) that are not necessarily correlated within a given implementation. For example an institution may find that their repository is at Level 2 for Information Security but Level 3 for metadata. In addition, within a functional area, the levels do not in all cases build on prior levels. An institution could find that it complies with Level 3 for metadata but not Level 2. Finally, many of the cells contain multiple guidelines. An institution may find that it only partly complies with Level 2 for File Fixity and Data Integrity. For these reasons the levels are better viewed as progressive stages or levels of service instead of “scores”. They can be used to identify broad areas to improve, identify areas of service excellence and pinpoint specific enhancements to make in order to comply with best practices. In addition they can be used to demonstrate the effect of large enhancement projects and to track progress over time.

The Levels team is hoping to learn how others are using the Levels to perform assessments of their preservation repositories. One method that has proven to be useful in one of the team members' institutions is described here. Within this institution, a large repository enhancement project (called project x here) is in progress. Each of the functional areas was reviewed sequentially from Level 1 through Level 4. For each cell, one of 5 values was written to a summary table:

- PASS - means that we already are doing these activities
- PASS (improved after project x) - means that we are already doing these activities but we will have an even better implementation after project x is complete
- PASS (after project x) - means that we will be doing these activities after project x is complete
- INCOMPLETE - means that we are doing the activity somewhat but not in an entirely satisfactory way
- FAIL - means that we are not doing this activity

As a result of coding each cell with one of these values and coloring the cells to make the patterns obvious visually, the summary table provides a powerful visualization of not only how the repository compares to this set of best practices, but also the effect of the enhancement project in progressing to where we want to be. Often enhancements to preservation systems can be largely “behind the scenes” so visualizations like this summary table provides can help justify the costs and effort by communicating the value to the organization. Information about additional purposes the Levels of Digital Preservation could serve or examples of actual use of the levels would be useful feedback.

Feedback and Future Work

As noted throughout this paper, the Levels were developed in a collaborative environment and they are still a work in progress. The Levels team actively invites comments and suggestions for how to improve the document. This paper and the associated Archiving 2013 presentation are mechanisms for the Levels team to spread the word further and invite more experts in the field to review and help refine the document. Levels team members are presenting the levels to other professional organizations and requesting their input during the course of 2013.

Readers can e-mail comments and suggestions to the paper’s authors at the addresses provided below. Revisions will continue until the Levels stabilize at a broad consensus view of the progression of technical steps recommended for decreasing the risk to digital materials. Comments received by August 31, 2013 will be considered in drafting the next version of the document.

- Jefferson Bailey jbailey@metro.org
- Andrea Goethals andrea_goethals@harvard.edu
- Trevor Owens trow@loc.gov
- Meg Phillips meg.phillips@nara.gov

In addition to revising the Levels chart, the project team plans several other types of future work. Based on earlier feedback, the team plans to incorporate definitions of terms used in the chart, and resources available to plan or execute each step. The team is considering including information about how the Levels might be used (as described above) in the document itself. The team would also like to provide a version of the chart online that allows the user to drill down on each cell of each level to access definitions and resources relevant to that topic. Feedback on these plans would be welcome as well.

Acknowledgments

The authors would like to thank the other members of the NDSA “supergroup” who helped to develop the Levels: Abbie Grotke (Library of Congress), Amy Kirchhoff (ITHAKA), Kris Klein (California State Library), Jane Mandelbaum (Library of Congress), Shawn Rounds (Minnesota Historical Society), Robin Ruggaber (University of Virginia) and Linda Tadic (Audiovisual Archive Network); as well as the members of the NDSA Content Working Group who laid the foundation for this project. In addition, thanks are in order to the individuals who provided valuable feedback on earlier versions of the Levels via the Library of Congress’ Signal blog. [10]
References


Author Biography

Best Practices for File-Naming

Just like paper files, electronic files need to be well-organized and labeled correctly so that they are identifiable and accessible by all employees. This is especially important for government offices in order to comply with legal requirements to ensure the availability, integrity, accessibility, and, if appropriate, the confidentiality of public records. G.S §132-8.1 provides that each agency shall establish and maintain an active, continuing program for the economical and efficient management of records of that agency.

Efficient management of electronic records begins with accurate file-naming. The following file-naming convention incorporates best practices in electronic records management and information technology. This document explores the general characteristics of records, how records are used and referenced, and the file-naming rules that should be applied to all electronic records.

An electronic record is machine-readable, meaning that it requires hardware and software to be accessed and read. Organization is especially important so that these records can be found and retrieved. Electronic records include documents, spreadsheets, databases, images, video, and audio. If not managed, a computer assigns a unique name for these files when saved, but these names do not provide a context for the file, nor are they logical. For example, the default file name for a Microsoft Word document consists of the first few words written on the first page. Images are frequently labeled with sequential numbers. These types of file names do not promote accessibility and ease of identification. For ease of explanation, the examples in these guidelines will focus primarily on documents; however, the standards outlined below apply to all file formats. In addition, this document will frequently use “record” and “file” interchangeably, as the records discussed here are electronic records and, therefore, generally files on a computer.

Keep in mind that this document is intended to be used primarily by individuals creating records on a daily basis. The rules discussed below are best practices and therefore highly recommended; however, every rule may not be relevant to every office. Regardless, this document should provide a foundation for developing a consistent and easy to use file-naming standard to be implemented in any office.

Records will be accessed by others.
As mentioned earlier, good file names are essential to accessibility. Many offices utilize a network server to store files so that they are accessible from multiple locations by various people. This requires that file names (as well as folder structures) make sense to more than just the creator. A file name should be clear to everyone in the department or agency in which the file was created. A record should be distinguishable from files with similar subjects as well as different versions of the same file.

When other individuals access a record, they may be using different operating systems (Microsoft Windows, Mac OSX), different versions of the system (e.g. Windows Vista, Windows XP), or different software (e.g. Microsoft Word, OpenOffice.org Writer, Notepad); therefore, it is important to follow rules that will allow a file to be recognized in as many different environments as possible.
Rule #1: Avoid using special characters in a file name. / : * ? " < > | [ ] & $ , .

The characters listed above are frequently used for specific tasks in an electronic environment. For example, a forward slash is used to identify folder levels in Microsoft products, while Mac operating systems use the colon. Periods are used in front of file-name extensions to denote file formats such as .jpg and .doc; using them in a file name could result in lost files or errors.

Rule #2: Use underscores instead of periods or spaces.

As mentioned above, periods already have a specific function in a file name, which is to tell the computer program where the file-name extension begins. Spaces are frequently translated in a Web environment to be read as "%20". For example,

Naming tutorial.doc

would appear as

Naming%20tutorial.doc

if it were available online. This alteration can cause confusion in identifying the actual file name. Spaces in file names can also cause broken links, because word processing tools like Microsoft Word, and e-mail clients like Microsoft Outlook, recognize spaces as an opportunity to move to another line. Therefore, a link to

\Ah1\Intranet\ar\naming conventions

could become

\Ah1\Intranet\ar\naming

conventions

Another difference that is found among operating systems and software is the acceptable length of file names. Some systems allow up to 256 characters, while others allow far fewer.

Rule #3: Err on the side of brevity.

Generally about 25 characters is a sufficient length to capture enough descriptive information for naming a record.

Records will be moved from their original location.

Files are frequently copied to other folders, downloaded, and emailed. It is important to ensure that the file name, independent of the folder where the original file lives, is sufficiently descriptive. Electronic records are usually organized in a series of folders. For example:

World_War_I\Posters\Owens\0001.tif
While this is a very organized way of storing records, it is only efficient as long as the files stay in their original folders in their original context. When multiple staff works on a project (or staff is in a position to get distracted) it is very easy to misfile a document. As soon as “0001.tif” is copied to another folder, or emailed to an agency, the context provided by the folders in which the document is nested is lost. Context is particularly important in legal situations because it provides authenticity and trustworthiness of the record. Losing the context of a record can possibly compromise its trustworthiness and therefore its validity or admissibility for court proceedings.

For instance, if the following files were pulled out of their appropriate folders, they would appear to be the same file:

World_War_\Posters\Owens\0001.tif
World_War_\Posters\RedCross\0001.tif

Rule #4: The file name should include all necessary descriptive information independent of where it is stored:

Incorrect: 0001.tif
Correct: wwl_poster_owens_0001.tif

Government records will have an associated retention schedule.

The point of organizing an agency’s electronic records is to enable accessibility not only by current users, but by future users as well. Records retention schedules are applied to electronic records just as they are to paper records. Having the relevant date associated with the file is essential. Though many operating systems store this information with the file, as users move the file among folders and computers and as the file is re-saved as revisions are made, those dates change. A file could have dates that do not make much sense to its original creation.

For example, the image to the right displays the properties for a file named ‘default’. The date modified appears to be before the date created. This disparity appeared when the file was copied from a server to the desktop of a PC.

The best way to prevent confusion is to embed the relevant date (the date that the file was created or revised) in the file name itself.
Rule #5: Include dates and format them consistently.

Some workgroups might find it more useful to have the date at the start of the file name, while others might prefer it at the end. Either way, it is a useful sorting tool when the files are organized. Just be sure to keep it consistent.

The best way to list the date is based on an international standard – ISO 8601. ISO 8601 specifies numeric representations of date and time to be used in electronic format. The international standard date notation is:

YYYY_MM_DD       or       YYYYMMDD

YYYY is the year, MM is the month of the year between 01 (January) and 12 (December), and DD is the day of the month between 01 and 31. For example, January 5, 2008 is written as 2008_01_05 or 20080105.

This format allows ease of sorting and comparing files by date and prevents confusion with other date formats (especially in other formats that use just two digits for the year).

For example, this document could be named:

filenaming_20080507

to reflect that this draft was last edited on May 5, 2008.

Records will have multiple versions.
A file will frequently have multiple versions, especially when it is created by a workgroup.

Rule #5: To more easily manage drafts and revisions, include a version number on these documents.

The easiest way to do this is to use the letter “v” to represent “version number.” Then, “v01, v02, v03” can be added as needed to a file and the main file name can stay the same. This is much more effective than other common additions like “update,” “new,” “old,” etc. An exception to this rule is using “FINAL” to indicate the final version of the document. This can be helpful to quickly identify the most accurate version of the document. When using “FINAL”, be sure to use it instead of the version number, rather than in addition to it.

Rule #6: Be consistent.

The most important rule of file-naming is to be consistent. Some choices will need to be made about organization that affects the entire workgroup – where to include the date, what abbreviations to use, etc. Regardless of what the group decides, it is only effective if everyone follows the rules consistently.
There will be exceptions.
One notable consideration is the batch-scanning process. This process typically relies on a program that sets its own parameters on the file names allowed. If this is the case, take advantage of the folder hierarchy and, when possible, apply the rules outlined here to folder names. Avoid spaces and special characters. Be consistent throughout the project; consider developing a file-naming standard for all batch-scanning projects.

Remember: this document is not going to apply absolutely to every situation; it should be used as a guide to encourage discussion in offices and workgroups to develop file-naming practices that work for best in those specific environments.

Still have questions? Contact us.

Department of Cultural Resources | Office of Archives and History

Government Records Branch

Mailing Address: 4615 Mail Service Center | Raleigh, NC 27699-4615

Physical Location: 215 N. Blount Street | Raleigh, NC 27601

Phone: 919-807-7350 | Fax: 919-715-3627 | Courier: 51-81-20

E-mail Address: records@ncmail.net

Additional Resources


Kuhn, M. “A summary of the international standard date and time notation.” http://www.cl.cam.ac.uk/~mgk25/iso-time.html

Getting to the Bottom Line:  
20 Cost Questions for Digital Preservation  
Printer-Friendly Version

About This Resource  
The following questions are being made freely available by the MetaArchive Cooperative Outreach Committee to assist institutions with their comparative analyses of various digital preservation solutions. Ask these of any digital preservation solution provider. Features and functionality are important, but those are often the easy pieces of information to learn about. Identifying and comparing short- and long-term costs, including a variety of up-front, recurring, and sometimes hidden fees, can make Getting to the Bottom Line a difficult task. Users of this resource are encouraged to make use of all or as many of the questions as desired.

1. What are the solution provider's licensing, subscription or membership fees?  
   a. Have these fees increased or decreased over the past three years, and why?  
   b. How often is the fee structure reviewed? And how are fees set?  
   c. How are customers/subscribers/members consulted during any such reviews?

2. Are additional memberships required to participate in the solution?  
   a. If so, what are the fees associated with those memberships?

3. Is there a minimum licensing/subscription/membership term?

4. What are the solution provider's storage fees?  
   a. Have these fees increased or decreased over the past three years, and why?  
   b. How often is the fee structure reviewed? And how are fees set?  
   c. How are customers/subscribers/members consulted during any such reviews?

5. Are there limits on the number or size of collections that can be deposited?  
   a. Is there a maximum limit on the amount of storage that can be utilized?

6. How does the solution provide for increases in storage capacity?  
   a. And how do these increases affect fees, if at all?

7. Do any fees paid (licensing/subscription/membership/storage) include geographically distributed copies of content?  
   a. If not, what's the additional cost for this option?  
   b. If so, how many copies, and in what geographic locations?

8. Am I responsible for obtaining any additional hardware or software at my own expense in order to work with the solution?  
   a. If Yes, what are the average hardware/software costs?  
   b. Does the solution provider publish specifications or requirements?  
   c. What is the recommended replacement cycle?  
   d. Are there any additional costs for security (e.g., encrypted storage, etc.)?

9. What are a few specific examples of tasks, services, or resources that my fees for this solution are supporting (e.g., staff salaries, infrastructure upgrades, research and development, etc.)?

10. On what schedule are customers/subscribers/members billed?
11. Does the solution provider charge a one-time setup, implementation, or initial ingest fee, above and beyond any ongoing fees?

12. On average, how long does it take to begin using the solution once a contract or service license agreement (SLA) has been signed?
   a. What steps are involved?

13. Does the solution provide basic documentation or instructions on getting started?

14. Does the solution provider prepare content for ingest?
   a. If so, do any fees being paid cover tasks like verifying inventories, performing fixity checks, and/or repairing any files that may get damaged in the process of sending content to the solution provider?
   b. If not, does the solution provider cover/defer any costs to train or outsource?

15. To what degree are the steps related to ingest, description, preservation, etc. automated?

16. If the solution is a non-profit (or run by one), are copies of the annual operating budget and financial statement shared with the members?
   a. If not, are there plans to?

17. In terms of sustainability, does the solution provider have a strategic plan, succession plan, or disaster recovery plan?
   a. If so, how up-to-date are such plans?
   b. Has the solution provider engaged in any audits or risk assessments?
   c. Are any of the plans or audit/assessment results publicly available?

18. Is there a charge for retrieving content from the solution?
   a. If yes, how much is it?
   b. How is this charge calculated and what does it cover?
   c. Under what terms? Are there any restrictions or limitations?

19. Is there a charge for deleting content from the solution?
   a. If yes, how much is it?
   b. How is this charge calculated and what does it cover?
   c. Under what terms? Are there any restrictions or limitations?

20. SELF-QUESTION: Am I, as a customer/subscriber/member, also paying for local backups above and beyond the preservation service costs?
   a. Recognizing that both backup and preservation are important and have their own unique and important places in an overall strategy, how can you best balance both costs?
Digital Preservation Resources List

General Resources


POWRR Preserving Objects With Restricted Resources IMLS Grant Project. http://digitalpowrr.niu.edu/
Appraisal, Assessment & Selection Issues


Data Asset Framework http://www.data-audit.eu/


Oxford University. Assessment Criteria for Digitization, 1999:
http://www.bodley.ox.ac.uk/scoping/assessment.html

Audits & Trusted Repositories

http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
[NOTE: ISO 16363 version of TRAC approved fall 2011]

Data Seal of Approval: http://www.datasealofapproval.org/en/

MetaArchive Trusted Repository Audit (2010):
http://metaarchive.org/sites/default/files/MetaArchive_TRAC_Checklist.pdf


Digital Content Inventories: Examples (note these are used for submission of digital content rather than for inventory purposes):

California Digital Library (CDL) / UC Libraries. “Digital Assets Submission Inventory,”
http://www.cdlib.org/services/dsc/contribute/docs/submission.inventory.rtf

File Formats, Sustainability

*Sustainability of Digital Formats, Planning for Library of Congress Collections:*
http://www.digitalpreservation.gov/formats/sustain/sustain.shtm

Brown, Adrian. *Selecting file formats for long term preservation, August, 2008:*

File Identification


PRONOM – “The Technical Registry” :


File Naming Conventions

Oregon State University. CWS Training. “What are Good File Naming Conventions?”
http://oregonstate.edu/cws/training/faq/what-are-good-file-naming-conventions.


University of Colorado Boulder. “Guidelines on File Naming Conventions for Digital Collections.”
Implemented March 4, 2008.
http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf.
University of Illinois at Urbana-Champaign. “2.0 Best Practices for File Naming.” 
http://www.library.illinois.edu/dcc/pdfs/best_practicespdfs/02_best_practices_for_file_naming_opt.pdf.

**Intellectual Property Issues**


Copyright Crash Course: http://copyright.lib.utexas.edu/


**Long-term Access Issues**


**Metadata**

http://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isq22no2.pdf

"Understanding Metadata," National Information Standards Organization, 2004

**Policies and Planning**

http://aabc.ca/resources/archivists-toolkit/preservation/

http://library.columbia.edu/services/preservation/dlpolicy.html

http://dspace.library.cornell.edu/bitstream/1813/11230/1/cul-dp-framework.pdf


http://www.icohere-meeting.net/WebConference/RecordingDefault.aspx?c_psrid=E959DF86834A

Northeast Document Conservation Center (NEDCC), Preservation Leaflet 1.3: The Needs Assessment Survey
http://www.nedcc.org/free-resources/preservation-leaflets/1.-planning-and-prioritizing/1.3-the-needs-assessment-survey

http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/preservation/policies/

Yale University Library. “Policy for the Digital Preservation.”
http://www.library.yale.edu/iac/DPC/revpolicy2-19-07.pdf

Storage Issues


“General Considerations for Choosing File Formats,” Andrea Goethals, Harvard University Library, July 20
http://library.harvard.edu/sites/default/files/general_format_considerations.pdf

http://library.harvard.edu/sites/default/files/OIS_recs_for_file_names.pdf
"OIS Recommendations for PDF Files Created for Long-term Preservation and Access," Andrea Goethals, Harvard University Library, February 2010
http://library.harvard.edu/sites/default/files/OIS_recs_for_pdf.pdf

Tools
