Digital Repository Best Practices for Cultural Heritage Organizations

Prepared by Katherine Kott
for the
Computer History Museum
February 3, 2012

Introduction: Assumptions and Definitions

This white paper is based on knowledge and experience I have gained in digital repository work and digital repository community involvement over the past decade. The contents of this report could be backed up by citations to journal articles and research papers. However, the intent is to offer practical advice for the Computer History Museum (CHM) to use in digital repository planning. Rather than interrupt the flow of the narrative with references, I have appended an annotated bibliography. Thanks to Heather Yager for surveying the literature, collecting copies of the resources, and extracting salient points from them.

The term *digital repository* is used to mean a variety of things. In academic environments, digital repositories are often systems where researchers deposit articles and papers. In archives, digital repositories are most often systems to manage digital assets--some born digital such as email correspondence, and some digital surrogates of other archival material such as photographs or documents.

In museums, digital surrogates from the museums' collections are commonly held in digital repositories. Some museums also collect born digital artifacts such as time-based art or software. Sometimes these repositories include preservation components, and sometimes they are focused on processing workflows. In certain cases, the repositories include "end user" access to the digital content, while in other situations access might be provided by another separate system.

For the purposes of this paper, and to meet CHM needs, we will define the digital repository as the systems and workflows that support *digital asset management* and *digital preservation.* Although access to CHM digital collections is important, end user access is not the focus of the digital repository planning project. Digital content production such as creating digital surrogates of objects in CHM collections, or production of digital video content is also excluded from this report.

I will begin the paper with background information about the history of digital repositories in cultural heritage organizations. This section will include lessons learned over the past two decades, and current thinking about digital repositories. Then I will suggest practices that should serve CHM well as the Museum plans for the next level of digital asset management and preservation for digital collections. I will present these recommendations within the context of emerging best practices in the cultural heritage community.

Background

Over the past two decades, leaders in museums, libraries, and archives have emphasized the role cultural heritage organizations must play in the management and preservation of digital collections, continuing the role they have long played in

the analog world. In the mid-1990s, various organizations and coalitions began to produce standards to inform digital preservation. These included standards for descriptive, technical, preservation, and structural metadata. Perhaps most importantly, this standards work also produced the Open Archival Information System (OAIS) reference model to define preservation functions.

In the early days, digital preservation was envisioned as a separate and specialized set of activities, focused on keeping digital objects safe and usable. More recently, preservation has come to be seen as a subset of "digital curation" activities by which digital objects are managed throughout the lifecycle, from creation and acquisition, to preservation, access, and re-use. This shift has been informed by a decade of digital preservation experience. Terminology is far from settled as this area continues to develop and grow. Ongoing lively discussions about what we mean by these various terms take place regularly on digital curation and preservation blogs and lists.

Digital preservation pioneers have reported on their experiences, producing reports of successes and failures in digital collection management and preservation. Repositories that were built to support the somewhat heavyweight and rigid standards developed in late 1990s and early 2000s with a focus on digital library collections have proved difficult for large institutions to scale and sustain, and impossible for smaller organizations to implement.

As a consequence, the field has moved away from the notion of a one-size-fits-all method for digital asset management and preservation for cultural heritage organizations. Instead, guidelines to help craft tailored solutions have begun to emerge. The Planning Tool for Trusted Digital Repositories (PLATTER) and other similar frameworks enable organizations to design solutions that are suitable for their specific needs. While these newer approaches are more flexible, the array of choices can be bewildering. The following best practices review provides specific recommendations for CHM and guidance for evaluating options.

Digital Repository Best Practices for the Computer History Museum

Each best practices section below begins with a specific recommendation for the Computer History Museum. The Computer History Museum certainly fits within a museum model in terms of its mission. However, the unique blend of collections the CHM holds, including archival materials, computer software, digital surrogates of objects in the Museum collections, and digital video produced in-house means CHM best practices should draw upon information from the library, archives, and museum communities. The following recommendations are derived from resources and experience across the cultural heritage spectrum.

- Begin with a planning tool
- Use the OAIS reference model

- Consider modular solutions
- Establish risk assessment policies and procedures
- Prefer throughput over exhaustive up-front preparation (just in time, not just in case)
    - Metadata
    - Validation and verification
    - Preservation practices
- Align digital collection policies and procedures with policies and procedures for non-digital collections
- Make storage basics a high priority
- Establish a clear rights policy

Begin with a planning tool:

Use the PLATTER framework as a guiding document for digital repository planning and operation.

The best place for an organization that is considering a digital repository service to begin is with a planning tool such as PLATTER. Essentially a questionnaire, the tool prompts for contextual information that can inform decisions. Without such a frame of reference, organizations can waste time planning systems that are not a good fit. Pitfalls include choosing metadata standards that would be difficult to implement or sustain, or getting distracted by advocates of specific technical solutions rather than focusing on the best option for the organization's specific needs.

Use the OAIS reference model:

Keep the OAIS reference model in mind to insure all critical repository functions are included in the repository solution CHM selects.

The Open Archival Information Standard (OAIS) continues to serve as a useful high-level reference model. The purpose of the OAIS model is to establish a metadata and packaging framework to insure that what is deposited can be retrieved and used in a given context.

Figure 1 provides a schematic representation of the OAIS functions. Content is packaged into a Submission Information Package (SIP). It is stored as an Archival Information Package (AIP) and can be served up in various ways through a Dissemination Information Package (DIP).

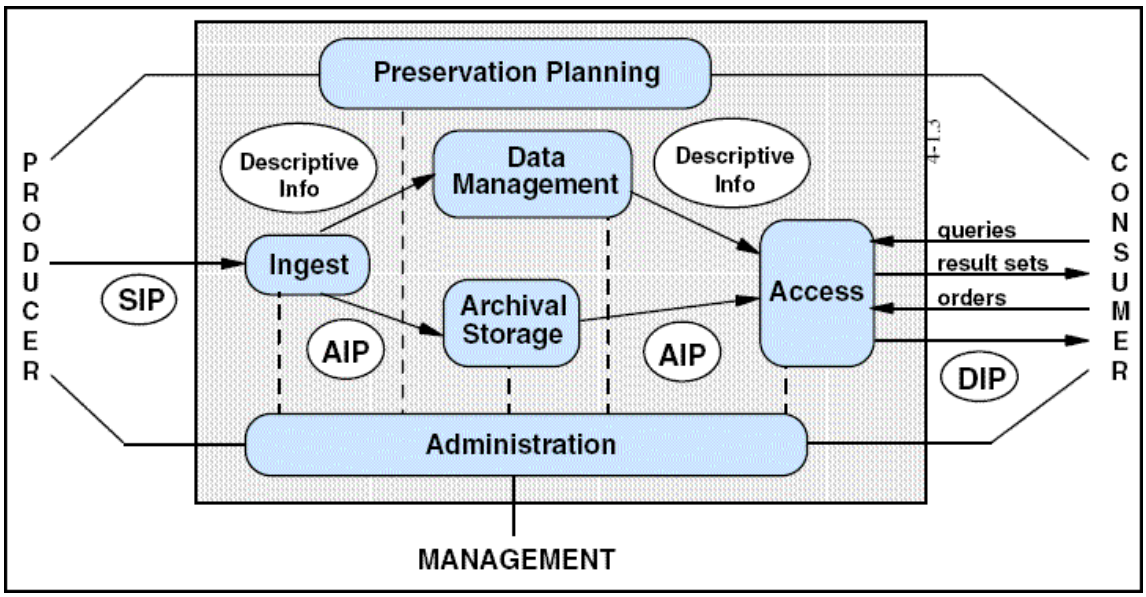

Figure 1: OAIS Reference Model.
PRESERV Project. Retrieved from http://preserv.eprints.org/guide/the_big_picture/?slide=2

Although OAIS is often associated with the use of specific metadata standards, Dublin Core for descriptive metadata, PREMIS for preservation metadata, and METS for structural metadata, the reference model is not prescriptive. It would be possible to implement an OAIS compliant system that uses other standards, EAD for descriptive metadata for example, or BagIt for packaging. Using OAIS as a reference model will insure that the digital repository can meet essential requirements; to be able to ingest digital objects, maintain them, and disseminate the objects to the depositor or other services in a predictable way.

Consider modular solutions:

When selecting software, CHM should consider modular solutions that can be integrated to support digital repository functions.

It is unlikely that CHM will find a single solution that will fit all digital repository requirements for digital asset management and preservation. In addition, monolithic solutions are expensive to implement and replace, extend and scale.

First generation digital repository systems were often designed monolithically, to serve a single purpose; a place to deposit papers, or a “dark archive” for preservation. In some cases, these systems have not been flexible enough in design to either scale up, or to interoperate with other systems that could provide complementary digital object lifecycle services.

The new generation of digital repository systems may still serve specialized functions, such as asset management or preservation, but these systems are

designed to be more extensible, with application programming interfaces (APIs) that enable separate but related systems to share data. In addition, many open source tools are available to fill specific requirements such as format verification or packaging for ingest.

Current best practice when building or selecting a digital repository system is to emphasize extensibility and flexibility. That way, as digital repository systems continue to mature, it is possible to swap out parts, without having to replace the whole. More and more, digital repository developers are being asked to integrate existing tools and services into workflows, and become active in the open source communities that enhance and maintain them.

Establish risk assessment policies and procedures:

Risk assessment policies and procedures must strike a balance to preserve what is inherently valuable while recognizing resource constraints.

Risk assessment includes physical evaluation of digital material when it first arrives at the CHM, and analysis of file formats. Physical evaluation is essential to determine if media is stable. It is also necessary for determining whether the data can be read. For example, it may be difficult to migrate data from an obsolete hand-held device to a CHM storage system.

Policies and procedures could set time limits for attempting to preserve data from donated devices, or could suggest seeking support for recovering the data through outsourced services when such a donation is accepted. In the mean time, best practice would be to store the obsolete device. Although the data it contains may be valuable, migrating the data to the digital repository would become a lower priority until resources could be obtained.

File format risk assessment can inform the level of service the CHM defines for certain kinds of digital content. The CHM may wish to establish a list of well-understood file formats such as those that are generated in-house during media production or digitization. These file types would be considered low risk for successful preservation. While CHM will almost certainly ingest other less well-known file formats (interactive games, obsolete software) the donor agreements and associated service level agreements for these file types could state the risk associated with preserving them. A basic preservation system is likely to be able to store and retrieve files in a wide range of formats. Enhanced systems and services would be needed to create emulation environments, or to forward-migrate files.

Digital preservation tools and frameworks offer criteria for digital collection risk assessment, pre-and post-ingest. Using these frameworks as a guide, it is important to create local policies and procedure to assess risk levels for incoming objects and

collections. Risk assessment is important in prioritizing preservation treatment for digital material.

Prefer throughput over exhaustive up-front preparation (just in time, not just in case):

Use collection policies to inform decisions about levels of treatment for digital collections.

Early digital preservation system implementers learned that requiring extensive preparation for digital objects before they can be deposited often translates into backlogs of digital objects that are not under any kind of control or protection. Rather than expecting all objects to be reformatted, or described in detail, particularly when future use for the object may not be well understood, it is better to ingest the objects using streamlined low-barrier-to-entry processes. Often, this involves little more than providing minimal level metadata to insure that objects can be identified and retrieved, generating a unique identifier for each object, and validating, and virus-checking files.

Metadata Standards

Consider Dublin Core for descriptive metadata and PREMIS for preservation metadata. Technical and structural metadata format choices can be based on what is supported by the digital repository systems that are selected.

Metadata for digital repositories may vie with storage as the areas with the least best practices consensus. For descriptive and technical metadata, choose formats that meet management and access needs and that workflow tools support.

It is important that metadata content standards be flexible enough to allow digital collections to be ingested and processed quickly, especially if the material is vulnerable (e.g. on unstable media). In these cases, minimal descriptive metadata such as title, creator, donor, and date, and a system generated unique identifier is enough to insure the material can be retrieved for further processing in the future.

Thorough descriptive metadata for certain categories of digital assets, such as objects that will be used in educational programs, or included in exhibits may be required. However, expecting that all collections receive the same treatment could be an obstacle to timely ingest of at risk material.

Descriptive metadata standards include Dublin Core, Metadata Object Descriptive Schema (MODS), and Encoded Archival Description (EAD) for archival materials. Although a range of descriptive metadata formats have been developed for specialized material, because of the diversity of CHM collections, the CHM would be best served by choosing a flexible descriptive metadata format. Dublin Core would

be a good choice. In addition to being flexible, it is already in use at the CHM. If EAD records are needed for shared access to archival collections, these records may need to be created in a system that is separate from the core digital repository system as not all digital repository software packages support the EAD format.

For technical metadata such as capture information, the Museum should implement similarly flexible policies. When digital surrogates are generated in-house or through contracted agreements, capture information should be retained as technical metadata. In the case of donated objects or collections, this kind of information should be recorded if available. However, it may be more important to ingest without full technical metadata than to wait to ingest until this information could be obtained.

The array of choices for storing technical metadata is even more confusing than those for descriptive metadata. Technical metadata may be stored in an xml document, or wrapped in a METS wrapper. In Fedora-based systems, technical metadata can be stored in a "FOXML" datastream. Either option is perfectly reasonable. The systems and software that are best suited to CHM requirements will likely determine the technical metadata format.

PREMIS is the de facto standard for preservation metadata and the Museum should plan to implement some version of PREMIS for the preservation system. I recommend a lightweight PREMIS implementation such as the Stanford model. PREMIS supports a level of complexity in preservation metadata that would be difficult to sustain. Best practice is to create local preservation policies and procedures and document them within the PREMIS framework.

Best practices for structural metadata may also vary, from none required for single images, to a range of solutions for representing complex relationships. Requirements are often based on what is needed for proper display and use of the objects. Complex relationships can be recorded in different ways, depending on the system or workflow used for processing. For digital artifacts such as interactive games, best practices have yet to be established.

Many systems support METS for structural metadata, although Fedora-based systems require METS packages to be broken apart into separate data streams for ingest. Here again, best practice is to choose a system that will support the range and level of structural metadata needed for the digital repository; the metadata format is less important than the information that can be stored within it. Figure 2: Metadata in the Workflow shows where different metadata formats and containers come into play in a digital repository workflow.

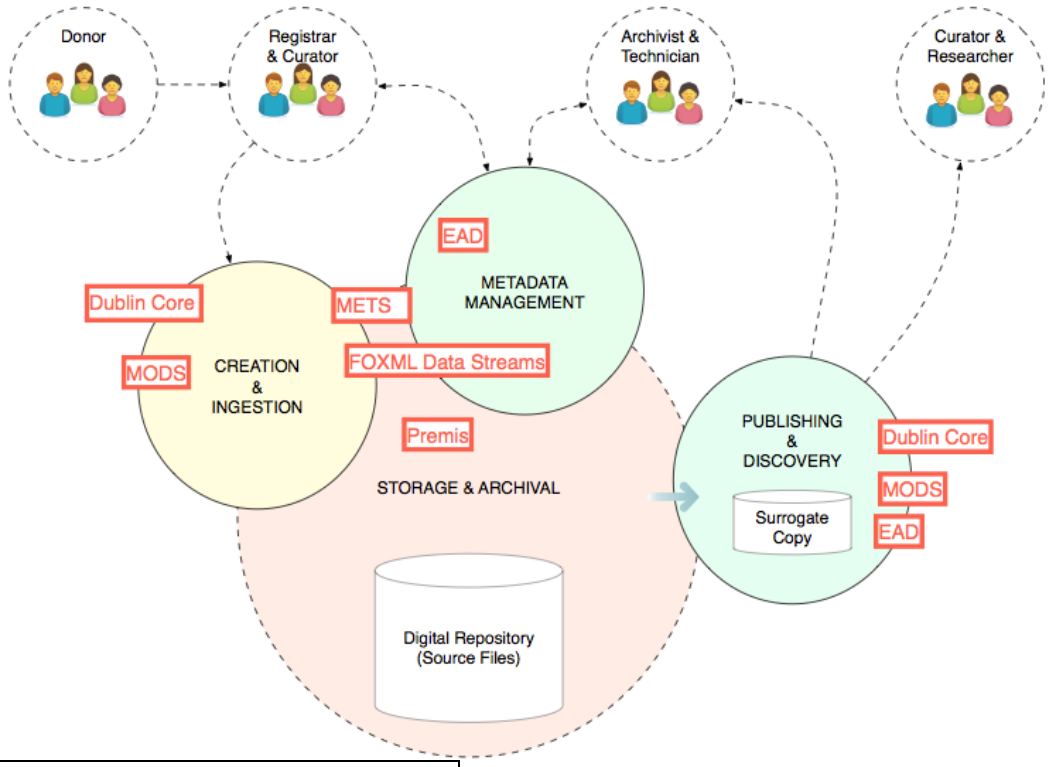

Figure 2: Metadata in the workflow

At this point, the CHM can afford to be flexible in thinking about what metadata formats to implement. EAD may be required to make archival collections discoverable, but METS might not be needed for structural metadata, for example. As digital repository software is evaluated, CHM staff should make certain the system provides support for the level of structural metadata that is required in a format that will be easy for other systems, such as the CHM website, to use.

Validation and Verification

Collection policies should determine the level of format validation the digital repository supports. When the CHM controls production of the artifacts, such as surrogates of physical objects, automatic format validation should be part of the workflow. Registries and validation tools for well-known formats are available and can be integrated into the digital repository workflow. For other collections such as historic software, format validation should not be required. Tools for checking do not necessarily exist, it would be expensive to produce them, and the software is unlikely to be rejected due to lack of validation. I therefore recommend that the Museum plan to implement different levels of format validation for some collection types than for others.

Following best practices for ingest, it is important to check for viruses, perform fixity checks on incoming files, and to generate a unique identifier for the object.

Preservation Practices

Unless future uses for digital objects are well understood, it is best to defer creation of emulation environments or other activities that involve research and development. While few organizations can afford to create emulation environments in operational workflows, as long as the objects are preserved, they can be retrieved for enhancement later, when specific use cases arise.

Align digital collection policies and procedures with policies and procedures for non-digital collections:

Reference existing Museum policies and procedures when creating policies and procedures for digital collection management and preservation to insure that collections in all areas are handled with the same level of attention.

Particularly in smaller organizations, aligning digital collection policies and procedures with existing policies and procedures can leverage scarce resources. Specialized systems will be needed to store digital assets, and asset management systems that are used for analog resources may not be extensible to digital assets. As a result, processing workflows for digital collections may differ from those for analog resources.

It may make sense to create and maintain separate documentation for digital repository systems, processes and procedures. Even so, the more similar workflows and processes can be, the easier it is to cross-train staff and provide flexibility for handling both analog and digital collections, including providing metadata to discovery and delivery systems.

Make storage basics a high priority:

It is imperative that all digital collections be migrated to a robust storage system that is backed up.

Some CHM collections are currently at risk because they are not being stored on professional IT grade media. Before CHM will be ready to implement a digital repository, it is imperative that all digital collections be migrated to a robust storage system that is backed up, preferably to tape that is then stored in an off-site location. Once this fundamental best practice is in place, the system can be enhanced with preservation management capabilities.

When selecting preservation software for the CHM, it may be necessary to consider proprietary solutions, even though open source is preferred. Specific recommendations from the Storage Consultant will be useful in this area.

Storage may be the area where best practices remain the least clear. That said, the first priority for a digital repository system is to insure that files are organized in a way that makes retrieval possible at the object as well as the file level. In addition, a

regular and robust backup procedure using proven mainstream IT technologies and media such as data tape is critical to digital repository operation.

At the next level, a storage system that includes preservation functions should be implemented. Despite the growing popularity of cloud storage for personal collections of music, photographs, and other data, preservation professionals have been slow to adopt cloud solutions, even for redundancy. Thus, current best practice is to replicate data, preferably using different technology stacks to create each copy. There is no hard and fast rule about how many copies are enough and there is little in the literature to suggest that three copies are better than two, for example.

"Lots of copies keeps stuff safe", the slogan behind the LOCKKSS acronym may describe a suitable practice for data that is already natively redundant, and relatively small. Licensed text-based content that the LOCKSS project has focused on is ubiquitous. A number of organizations can capture and replicate this data in a peer-to-peer network without much effort. The same strategy will not work for unique collections, especially collections involving large files.

Consequently, best practice in this area must be informed by what is practical and affordable for the organization. A preservation master on spinning disk or tape, and at least one other copy that is stored in a separate location, but is accessible by a system that allows audit would meet minimum level requirements for a preservation system. Beyond that, questions about how many additional copies are needed, and whether access copies such as lower resolution versions of image files need to be preserved must be considered in the context of organizational mission and resources.

Establish a clear rights policy:

The Computer History Museum must create rights policies consistent with risk assessment levels and resource allocation.

Rights issues are complex. Although best practices can provide some guidance, the CHM must decide the level of risk it is willing to assume, what level of resources the organization can dedicate to resolving rights questions, and how vulnerable the resources are.

For example, will digital collections or commercial software that appear to have been copied be accepted? If the objects have no provenance, or the provenance is uncertain, will CHM ingest them for preservation but not provide access? Will rights be sought to copy software for preservation purposes, or will the CHM take the position that copying for preservation is allowable?

Digital preservation professionals argue that best practices from a preservation standpoint make it important for organizations to preserve digital material, even if

the rights to do so are not entirely clear. However, most people in the field recognize that the organization's legal advisors must make the final determination what the organization's policies should be.

The Museum has already established rights policies and procedures within the *Guidelines for Documenting Digital Exhibit Artifacts*. These policies and procedures are aligned with current best practices and could serve as a model for generalized rights policies and procedures for the digital repository.

Conclusion

There are not yet, and perhaps never will be a set of unified digital repository best practices. Organizational needs are too diverse to support a single group of recommendations that would function well in all environments. However, broad frameworks and reference models do provide overall guidance. Creating local policies and procedures that fit within general frameworks such as PLATTER and OAIS will insure that a digital repository meets functional requirements that are accepted by the community. In addition, a modular design will create a digital repository that is sustainable and extensible. With preservation as a key service, it is also imperative as a best practice, to implement a robust storage and backup system following professional IT practices.

Beyond these fundamentals, best practices for digital repositories involve selecting systems to support workflows for efficient and effective asset management and preservation. This is where organizational needs must be translated into policies and procedures for managing and preserving digital collections within the context of the Museum's mission.

Appendix: Annotated Bibliography

Digital Repository best practices are still very much a work in progress. Although it may appear that there is little consensus, best practices are emerging through the efforts of digital preservation pioneers. Based on this early experience, the community has shifted from a focus on digital preservation to digital curation. This shift puts preservation into context with the full digital object lifecycle. Along with this shift, best practices have become less absolute, and more situational. In the following bibliography, the resources issued after about 2005 are more likely to reflect current thinking in the community than those produced before that date.

Organizational web sites

There is no one organization in charge of promulgating digital repository best practices for the cultural heritage community. However, a number of organizations play key roles. These organizations frequently post white papers on their web sites and sometimes offer blogs and webinars on digital repository topics.

The **Library of Congress Digital Preservation** web site links to white papers on a range of digital repository topics such as digital format sustainability and rights. The Library of Congress also lists relevant conference and educational offerings on this site, links to **The Signal** blog—a great resource for keeping up to date on digital repository issues. Library of Congress @ndiipp recent tweets appear on the website as well, highlighting timely topics of interest. Although focused on digital preservation programs in libraries and archives, the web site contains much practical advice of use to the broader cultural heritage community, and even to individuals interested in preserving personal digital collections such as photographs.

Based in the UK, the **Digital Curation Centre** focuses on managing digital research data. However, their white papers and reference models can be helpful in understanding management and preservation choices for born-digital content. DCC also lists training opportunities, although their listings tend to be for events held in Europe.

Traditionally associated with libraries, OCLC Research has expanded its mission to include archives and museums. Their focus is much broader than digital repository management, so mining the web site for relevant information can be somewhat time consuming. OCLC is a good resource for metadata standards information, and can also be a useful launch point for information on standards for unique identifiers. Brian Lavoie has contributed to numerous articles and papers on the economics and sustainability of digital preservation. Reading a few of these recent reports can

enhance understanding of the ongoing nature of digital repository management and operation.

Before the Research Libraries Group (RLG) merged with OCLC, the two organizations jointly produced Trusted Digital Repositories: Attributes and Responsibilities. Although published in 2002, this white paper has informed many current frameworks for establishing trust in repository operations and is still considered a classic.

The Coalition for Networked Information (CNI) is a membership organization aimed at academic libraries and information technology organizations. The Executive Director, Cliff Lynch, monitors issues of interest to these types of organization, including digital curation. Of late, CNI has focused significant attention on preserving scientific data. However, earlier task force reports and papers on building sustainable digital preservation solutions may be of interest. The web site also links to videos and podcasts of talks from CNI project briefings from CNI meetings held twice a year.

Other notable resources

Every year, Charles W. Bailey, Jr. publishes a Digital Curation and Preservation Bibliography. The 2010 edition cites over 500 selected resources, published from 2000-2010.

Members of the Digital curation Google group discuss everything from specific technical implementations such as debugging the BagIt tool, to how the term "curation" has come into general parlance. Over 500 members strong, this list is the place to be to learn about trending topics, upcoming training opportunities and to get advice from other members of the digital curation community.

Selected material on specific topics

The articles listed below informed the best practices recommendations in the *Digital Repository Best Practices for Cultural Heritage Organizations* report.

Repository Basics

Ball, A. (2010). *Preservation and curation in institutional repositories* (version 1.3). Edinburgh, UK: Digital Curation Centre.

While this white paper focuses on preservation of scientific data, it provides a useful review of the state of digital repository development as of 2010. The paper is descriptive rather than prescriptive, although it suggests preferences in some areas such as metadata for complex objects by omitting some commonly used standards.

The report includes information about three specific repository solutions, EPrints, DSpace, and Fedora. It goes on to inventory digital preservation models such as OAIS, architectures, and planning tools, including PLATTER.

The metadata section of the report includes information about format registries that might be of use when deciding which preservation formats to support for objects the CHM controls (see format verification section of Best Practices document). The article covers Dublin Core as a descriptive metadata option in the most detail. Predictably, PREMIS is emphasized for preservation metadata. Somewhat surprisingly, there is not mention of METS for complex metadata representation, with OAI-ORE the only standard that is described.

The remainder of the paper reviews tools for extracting and capturing a range of metadata, from technical to descriptive. It goes on to discuss what might be called levels of preservation, from bit preservation to emulation to migration and wraps up with an inventory of audit tools and resources for understanding costs and benefits of digital repositories.

Consultative Committee for Space Data Systems. (2002). *Reference model for an Open Archival Information System* (OAIS). Washington, DC: CCSDS.

This reference model continues to define the essential functions a digital repository should support.

Cothey, V. (2010). Digital curation at Gloucestershire Archives: From ingest to production by way of trusted storage. *Journal of the Society of Archivists.* 31(2), p. 207-228.
This article is an excellent how-to use case for using OAIS and PLATTER to implement a digital repository.

The Shift from Preservation to Curation

Abrams, S., Cruse, P., & Kunze, J. (2009). Preservation is not a place. *International journal of digital curation,* 4(1). Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/98

The authors explain the design changes they are making in the California Digital Library digital repository prompted by community shifts that emphasize curation rather than preservation.

Lee, C.A. & Tibbo, H. (2007). Digital curation & trusted repositories. *Journal of digital information,* 8(2). Retrieved from http://journals.tdl.org/jodi/article/view/229/183

This article is an introduction to a special issue of JoDI. It puts digital curation in context and explains the community shift from preservation to curation.

Rubin, N. (2009). Preserving digital public television: Not just an archive. *Library trends*, 57(3). Retrieved from https://www.ideals.illinois.edu/handle/2142/13605

Specific to the television production industry, this article includes general information about the way moving to digital production causes a shift in thinking about preserving digital video. The shift involves providing consistent access to what might previously have been thought of as archival material. This reinforces the lifecycle management, or curation approach.

The project was a pioneering effort to create file wrappers (structural metadata) for digital video and recommend preservation metadata standards for this type of material. The experimental repository handled both high-resolution masters, and low-resolution distribution copies of each digital title. To meet the local needs of the project, PBCore was used as the descriptive metadata format, PREMIS was used for preservation metadata, METSRights was used for rights, and METS was used as the wrapper. These choices were presented in the context of this particular project, not necessarily as a best practice for digital video curation. The article touches on rights and sustainability issues but does not recommend best practices in either of these areas.

Best Practices for Establishing Trust

Jantz, R. & Giarlo, M. (2007). Digital archiving and preservation: Technologies and processes for a trusted repository. *Journal of archival organization*. 4(1-2) p. 193-21. doi:10.1300/J201v04n01_1

Jantz and Giarlo put what it means to develop a trusted repository in context through the example of the Fedora-based preservation repository developed at Rutgers. While they restrict their examples and discussion to surrogate digital objects, not born-digital content, many of the solutions they suggest would be broadly applicable. They make the case that good storage policies need input from archivists and information technology professionals, and emphasize the importance of persistent identifiers in establishing authenticity, and reliability.

Research Libraries Group. (2002). *Trusted digital repositories: Attributes and responsibilities.* RLG: Mountain View, CA. Retrieved from http://www.oclc.org/research/activities/past/rlg/trustedrepositories.pdf

Although this paper was written before the preservation to curation shift, and RLG is now part of OCLC, this report stands as a useful framework for thinking about how an organization can fulfill its responsibilities to the community it serves in terms of preservation services.

Rosenthal, C., Hutař, J. & Blekinge-Rasmussen. (2008). Planning a Trusted Digital Repository with PLATTER. Digital Preservation Europe. Retrieved from http://www.slideshare.net/DigitalPreservationEurope/platter-colin-rosenthal-presentation

This slide presentation explains how the PLATTER framework can be used to establish organizational policies and procedures to insure repository services are trustworthy.

Trend toward Modular Solutions

Cramer, T., & Kott, K. (2010). Designing and implementing second generation digital preservation services: A scalable model for the Stanford Digital Repository. *D-Lib Magazine*, 16(9/10). Retrieved from http://www.dlib.org/dlib/september10/cramer/09cramer.html

In addition to presenting arguments for building modular digital asset management systems, this article makes the case for minimal treatment of some collections to insure they can be ingested for preservation in a timely manner.

Mikhail, Y., Adly, N., & Nagi, M. (2011). DAR: Institutional repository in action. In S. Gradmann, F. Borri, C. Meghini, & H. Schuldt (Eds.), *Research and advanced technology for digital libraries.* (pp. 348-359). Berlin: Springer-Verlag. Retrieved from http://www.bibalex.org/isis/UploadedFiles/Publications/DLF36_DAR%20Institutional%20Repository%20Integration%20in%20Action.pdf

Bibliotheca Alexandrina (BA) also developed a modular solution for their "Digital Asset Repository" (DAR). They translate the OAIS model into four somewhat differently divided components:

- *Digital Assets Factory (DAF) for digitization and ingestion,*
- *Digital Assets Metadata (DAM) subsystem for metadata management,*
- *Digital Assets Publishing (DAP) components to enable discovery and delivery systems, and*
- *Digital Assets Keeper (DAK) for object storage and versioning.*

The article explains how they implemented a repository, including architecture (Fedora-based) and metadata choices.

Metadata Resources

Donaldson, D., Conway, P. (2010). Implementing PREMIS: A case study of the Florida Digital Archive. *Library Hi Tech* 28(2) p. 273-28. doi 10.1108/0737883101104767

While this article focuses on how developers at the Florida Center for Library Automation implemented PREMIS when developing a preservation repository, it

provides context for understanding PREMIS as an information model. PREMIS itself does not dictate preservation policy, nor does is offer implementation solutions. It simply provides a framework for expressing preservation and other metadata (including rights metadata) in xml.

PREMIS Editorial Committee. (2011). *Introduction and supporting material from PREMIS Data Dictionary for Preservation Metadata.* Version 2.1. Library of Congress: Washington, DC. Retrieved from http://www.loc.gov/standards/premis/v2/premis-report-2-1.pdf

By removing nearly 200 pages from the full data dictionary, the Editorial Committee has created a manageable length document that explains how to use PREMIS for preservation metadata. The report goes far beyond the practical application of PREMIS. It puts PREMIS in the context of digital preservation, defines preservation concepts, and includes a preservation glossary.

Riley, J. (2009). Seeing standards: A visualization of the metadata universe. Bloomington, IN: Indiana University Libraries.
Retrieved from http://www.dlib.indiana.edu/~jenlrile/metadatamap/

This poster maps all the various metadata standards in use in various cultural heritage sectors. It gives a flavor for the range of choices available. The accompanying glossary defines the standards and provides information about what standards bodies are responsible for maintaining them. Helpful for identifying metadata options.

Smith-Yoshimura, K. (2007). *RLG Program Descriptive Metadata Practices Report.* Dublin, OH: OCLC Programs and Research. Retrieved from http://www.oclc.org/research/publications/library/2007/2007-03.pdf

An inventory of systems, descriptive metadata structures and content standards used by RLG members describes a field without much cohesion, although Dublin Core does seem to be emerging as the most prevalent non-MARC metadata structure.

Wilson, A. (2010). How much is enough: Metadata for preserving digital data. *Journal of library metadata.* 10, p. 205-217. doi 10.1080/19386389.2010.506395

Wilson writes with a view towards recommending preservation metadata standards for preserving digital data based on sound archival record-keeping practices. He critiques PREMIS for the shortfalls he sees in the standard's capacity to handle provenance information, which he sees as critical to establishing authenticity for digital objects. He is also critical of PREMIS and a particular implementation of PREMIS in METS for falling short of what is needed to prove reliability, usability and integrity. In his view, the metadata model should make provision for access. Not everyone would agree with this approach, seeing other systems outside the preservation system as the place where discovery and delivery metadata is maintained and backed up, not necessarily preserved.

Assessment, Appraisal, Upstream Requirements

Ide, M. & Weisse, L. (2006). *Recommended appraisal guidelines for selecting born-digital master programs for preservation and deposit with the Library of Congress.* Boston: WGBH. Retrieved from http://cn2.wnet.org/thirteen/ptvdigitalarchive/files/2009/10/appraisal-guidelines-final.pdf

While developed for a specific appraisal process, this white paper contains background material on the archival appraisal process in general. It also puts appraisal of public television programming in context with digital moving image and audio material. It provides a useful model for creating appraisal policies and procedures for complex, born-digital content.

Marmor, J., Van Malssen, K. & Goldman, D. (2011). *A preservation-compliant media asset management system for television production.* New York: WNET. Retrieved from http://www.thirteen.org/ptvdigitalarchive/uncategorized/designing-and-deploying-a-preservation-compliant-media-asset-management-system-for-television-production/127/

This white paper makes recommendations for combining preparation of media files for preservation with the television production process. It is based on the experience of ingesting digital output from television production into the New York University preservation repository. The NYU repository contains diverse content, but requires that metadata for ingest be presented in a standard format. The incoming metadata from different television production workflows required laborious re-work for ingest.

The resulting recommendation was to push requirements for standardization upstream, suggesting that production workflows should include creation and capture of standardized technical and descriptive metadata. Creating incentives for production to take on the additional work of creating improved metadata could prove problematic unless resources for production and preservation are viewed holistically, from a resource standpoint.

Ross, H. & Thompson, D. (2010). Automating the appraisal of digital materials. *Library Hi Tech.* 28(2) p. 313-32. doi 10.1108/0737883101104770

On the way to exploring whether or not automated appraisal of digital materials is feasible, this article makes a number of sensible preservation policy recommendations. The authors note that following sound archival principles is essential to good preservation policy and that OAIS should be used as a digital preservation framework. They suggest that both intellectual appraisal, evaluating the intrinsic value of the material, and technical appraisal, the question of whether the organization can afford to preserve it, should be considered. While tools such as those developed through the PLANETS project, and JHOVE can be used for format validation, integrating such tools into an appraisal workflow involves a complex set of tasks by experienced developers.