

INTRODUCTION

This annotated bibliography has been compiled to acquaint you with the challenges associated with developing a digital preservation plan and repository, and successful strategies for overcoming those challenges.

There is some ambiguity in the terminology used to discuss digital preservation, curation, and repository implementation. For example, the terms “strategies” and “services” are used both at a macro level (to refer to broad-based institutional strategic planning and to the delineation of sets of services to be provided to user communities) and a micro level (to refer to specific and often technological methods of preservation). “Plans” and “planning” also take place at both levels. Additionally, the concept of a repository is subject to multiple and overlapping interpretations. There are institutional repositories, subject-based repositories, research repositories, national repository systems, and of course, archives. Digital preservation is also referred to as digital archiving. This list tries to clarify usage within the context of each subtopic covered.

The reading list moves from broader, more strategic considerations through models, tools, and concepts used in the preservation planning endeavor, and then combines these into specifics about preservation planning and case studies. Within some sections, a suggested reading order is provided.

OVERVIEW OF DIGITAL PRESERVATION NEEDS AND CHALLENGES

Material in this section provides an overview of the urgent need for digital preservation, the many types of materials encompassed, the threats to those materials, some of the challenges faced by institutions, a survey of high-level requirements for a successful preservation solution, and rationales for developing a sustainable, strategic approach.

Suggested Order:

Waters, Sinclair, *Sustainable Economics...*

Sinclair, Pauline. *The Digital Divide, Assessing Organisations' Preparations for Digital Preservation*. Rep. A Planets White Paper, Mar. 2010. Web. 15 Jan. 2013. <http://www.planets-project.eu/docs/reports/planets-market-survey-white-paper.pdf>.

Contained here are the results of a 2009 survey investigating the digital preservation activities and needs of 200 organizations. Although the respondents were primarily European archives and libraries, the survey has applicability

across borders and provides a recent and “real” entry point to the situation of contemporary organizations.

These organizations detail a pragmatic list of reasons to preserve digital content, challenges with the scale of digital content (volume and type), their current funding structures, and stages and characteristics of their implementations. Of particular interest are the findings concerning common expectations about what a digital preservation should do (#1 – maintain authenticity, reliability, and integrity of records) and the constraints felt by these organizations.

Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information. Rep. Blue Ribbon Task Force on Sustainable Digital Preservation and Access, Feb. 2010. Web. 20 Jan. 2013. http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf.

NOTE: Executive Summary, p. 1 – 5

This report focuses on the sustainability of digital preservation programs from an economic perspective. It distinguishes between four content domains that populate the field and the literature: scholarly discourse (published output), research data (primary inputs and first-order outputs), commercially owned content, collectively produced Web content. Each domain is evaluated and risks and recommendations are noted. After listing the “structural challenges” affecting all preservation strategies, the task force indicates three imperatives for a sustainable program.

Waters, Donald, and John Garrett, eds. *Preserving Digital Information: Report of the Task Force on Archiving Digital Information*. Rep. The Commission on Preservation and Access and The Research Libraries Group, 1 May 1996. Web. 15 Jan. 2013. <https://www.oclc.org/content/dam/research/activities/digpresstudy/final-report.pdf>.

Although this document is almost 20 years old, it continues to be a frequently referenced resource in digital preservation literature, perhaps because it makes a clear case for the importance of careful attention to developing preservation repositories and supporting standards, processes, and infrastructure to protect digital assets. It also successfully integrates the perspectives of archivists and technologists.

A case is made for preserving digital assets against a backdrop of the inherent issues associated with the creation of those assets due to the evolution of technology. A

spectrum of challenges associated with archiving digital information is succinctly addressed. The features that determine the integrity of information objects are enumerated and the concept of archival certification is introduced. Archival roles and responsibilities are discussed, some functionally tied to traditional archival repositories and some unique to the digital world. Migration strategies and cost discussions round out the document and perhaps show a little age. Recommendations for pilot projects, support structures, and best practices are worthwhile as many are ongoing efforts today.

CONSIDERATIONS: STRATEGIC APPROACH, SCOPE OF CONTENT, AND SERVICES

An institution's response in developing a strategic approach to digital preservation will depend on many unique, local environmental factors. The readings in this section address how to set the framework for the scope of a particular academic institution's digital preservation program(s) and repository implementation. They suggest that a repository should be tailored to the specific institution's profile and that consensus behind decisions must be carefully built. What is the mission of the repository? How will it fit within the functional areas of the university? What content should be included? What services should be offered? How might it collaborate within and beyond the boundaries of the university?

A new digital preservation plan and implementation of a repository afford the opportunity to carefully consider what to deploy and how to deploy it. Libraries are repurposing their position and function within academic institutions, and it is important to consider the options; there are many that may suit a library's purposes and its stakeholders' objectives. In the readings, the concept of services appears as a theme, tying back to achieving sustainability through successful demand for access/use. Life cycle management (curation) is introduced. Collaboration as a strategy is often suggested.

The Research Library's Role in Digital Repository Services: Final Report of the ARL Digital Repository Issues Task Force. Rep. Association of Research Libraries, Jan. 2009. Web. 4 Mar. 2013. <http://www.arl.org/bm~doc/repository-services-report.pdf>.

ARL members are urged to view digital repository services as a strategic priority for the future. The Report articulates why digital repository services are important, identifies key issues and choices to be addressed in formulating a plan and determining what objects and services to include, explores three strategies that are commonly used in the implementation of a repository, and analyzes environmental trends that will impact repository development by the year 2015.

Digital repository services provide scholars and researchers with a significant advantage in conducting research, delivering high quality teaching, and contributing scholarship

and services to society. This report's focus is on repository services rather than technologies. Key issues are building services around new content and old content in new forms, engaging with a larger networked environment, attending to the "demand side," and sustainability. Strategies explored are in-house development and deployment, collaborative or partnering approaches within and across institutions, and balancing local and vended services.

Walters, Tyler, and Katherine Skinner. *New Roles for New Times: Digital Curation for Preservation.* Rep. Association of Research Libraries, Mar. 2011. Web. 10 Mar. 2013. http://www.arl.org/bm~doc/nrnt_digital_curation17mar11.pdf.

The ARL herein elucidates the many ways research libraries can and should be involved in the lifecycle of digital materials to add value to the production of new research knowledge. It suggests that libraries focus on building infrastructures, content, and services and delineates specific roles for both libraries and librarians. It advocates for collaborative activities that are library-driven over which libraries retain authority and ownership. The report notes the usefulness of a digital curation perspective and discusses staff deployment.

The core of the report provides 13 case studies of collaboration strategies employed by digital preservation organizations that demonstrate a central repository model, a community-driven repository or network, and a third-party approach. Two appendices offer disciplinary considerations for digital curation within the sciences and within digital humanities.

Digital Repositories: Research Data Services

Tenopir, Carol, Ben Birch, and Suzie Allard. *Academic Libraries and Research Data Services: Current Practices and Plans for the Future.* Rep. Association of College and Research Libraries, June 2012. Web. 17 Mar. 2013. http://www.ala.org/acrl/sites/ala.org/acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf.

NOTE: Read Executive summary, Introduction, Related research (p. 3 – 13). Scan survey results. Read Actions for library directors and Conclusions (p. 41 – 42).

As a point of reference in planning for the scope of repository services, this ACRL report provides a baseline assessment of the current state of and future plans for research data services in academic libraries in the United States and Canada. This survey focuses on the sciences, but notes that other disciplines are also becoming data-intensive and computational. Services are delineated as both informational and technical.

As with other reports from professional organizations, this report urges academic libraries to play a more active role in digital/data-related processes. The authors suggest libraries

provide consulting services related to data management and curation, infrastructure for storage and curation, and support librarians in becoming members of research teams as data curation consultants. This report is clear in pointing to data curation as a top direction for academic libraries, especially those at doctoral degree-granting/research institutions.

Wong, Gabrielle K. "Exploring Research Data Hosting at the HKUST Institutional Repository." *Serials Review* 35.3 (2009): 125-32. Elsevier, Sept. 2009. Web. 28 Mar. 2013.
<http://repository.ust.hk/dspace/handle/1783.1/6096>.

Wong details research data services at the Hong Kong University of Science and Technology (HKUST) Institutional Repository, as well as a project she undertook, a survey of research data sets at major institutional repositories. The author provides useful guidelines for developing a repository's policy framework, structural measures, and content recruitment for those in charge of data archiving and management.

The article first outlines issues related to data preservation, then briefly describes the approaches various institutions have taken to data archiving: Purdue University's Distributed Data Curation Center, Johns Hopkins' compound object publication model, Australia's data curation continuum. Next, the article reviews the survey which examined types of data being archived at IRs, how data is accommodated in DSpace and other software platforms, and what steps LIS professionals can take to ensure data is usable, accessible, and discoverable. The author proposes a number of measures for refining data archiving services and takes an optimistic stance on the future of Institutional Repository management.

Institutional Repositories: Scholarly Discourse, Archives, (and Beyond)

Chavez, R., G. Crane, A. Sauer, A. Babeu, A. Packel, and G. Weaver. "Services Make the Repository." *Journal of Digital Information* 8.2 (2007): n. pag. Web.
<http://journals.tdl.org/jodi/index.php/jodi/article/view/195/179>.

Chavez et al. use the Perseus Project to make the case that an institutional repository (IR) is defined not by its content but by its services. The Perseus Digital Library is a large digital collection of Greco-Roman cultural heritage that provides users with a rich array of object behaviors. Perseus sought to augment services as it was moved into a Fedora-based IR. Much of the discussion of IRs centers on preservation strategies, establishing trust, and encouraging faculty participation. While these factors are vital to an IR's success, user services are also key in establishing a well used IR. In practical terms, use justifies the cost of establishing and maintaining IRs.

This emphasis on services connects with the OAIS emphasis on a defined user community. Chavez et al. point out that in order to attract submissions IRs must be relevant, and in order to be relevant IRs need to be service-oriented, which means being flexible enough to reach users where they are and shift services as user needs evolve.

Salo, Dorothea. "Innkeeper at the Roach Motel." *Library Trends* 57.2 (2008): 98-123. Print.

Entertaining and informative, Salo provides insights into the pitfalls of how institutional repositories have been implemented in the past, and suggestions to IR managers on how to avoid those pitfalls. The issues that have arisen and prevented a clearer success story for IRs are documented in detail. All constituents have played a part, including faculty, librarians, repository managers, repository software, professional organizations, publishers, and top-level administrators.

The author offers twelve strategies for "roads forward" for IRs. These include (but are not limited to) the suggestions that IRs must integrate with other library programs and priorities, that IR managers should involve themselves in campus cyberinfrastructure discussions and should help with research data management, and that support for IRs "begins at home."

Wise, Marie, Lisa Spiro, Geneva Henry, and Sidney Byrd. "Expanding Roles for the Institutional Repository." *OCLC Systems & Services* 23.2 (2007): 216-23. Print.

Establishing an institutional repository (IR) requires the successful integration of various institutions, software platforms, third-party services, and user communities. Wise et al. offer the Travelers in the Middle East Archive (TIMEA) out of Rice University as a case study in IR integration.

TIMEA is a digitized collection of eighteenth-twentieth century accounts of western travel in the Middle East that includes texts, images, and contextual information. As one of many digital collections at Rice, TIMEA must be considered in a broad context. TIMEA integrates with two other systems, Connexions (a separate repository of education modules) and ArchIMS (for dynamic GIS data) to present a unified, full-service interface to the user. To help connect objects across these systems, TIMEA uses a custom web interface, XML support, and a combination of Dublin Core and METS metadata. TIMEA demonstrates the practical and technological challenges in establishing an IR that maintains its individuality and serves its unique designated community while also embedding itself in broader digital collections.

FUNCTIONAL FRAMEWORK: MODELS FOR DIGITAL PRESERVATION AND CURATION

There are two models for digital archiving/preservation/curation that have been particularly well-received. These two models provide the functional framework from which more detailed implementation planning proceeds once overall scope, policies, and constraints have been determined. Ultimately, those most intimate with the preservation planning and repository design process should study the [Digital Curation Centre website](#) and the [Open Archival Information System documentation](#).

Higgins, Sarah. "The DCC Curation Lifecycle Model." *International Journal of Digital Curation* 3.1 (2008): n. pag. Web. 04 Mar. 2013. <http://www.ijdc.net/index.php/ijdc/article/view/69>.

The life cycle of a digital object is a vital concept in digital preservation and curation and varies greatly from traditional preservation. This article includes explanations of common terms and key ideas essential for an understanding of digital preservation and the curation lifecycle. It suggests that the Digital Curation Centre (DCC) model should be used to evaluate and inform digital preservation and curation decisions.

Lifecycle models focus on preservation beginning at the birth of born digital objects and continuing through their lifespan. The DCC model is adaptable, allowing for varying levels of granularity. It defines roles and responsibilities of staff members. Furthermore, the DCC model details which tasks need to be completed sequentially, which tasks can be done on an as-needed basis, and explains why. Note that the model complements models such as OAIS and ISO standards.

McMeekin, Sharon M. "With a Little Help from OAIS: Starting Down the Digital Curation Path." *Journal of the Society of Archivists* 32, No. 2 (2011): 241-253. Print.

The Reference Model for an Open Archival Information System (OAIS) is highlighted in this article. The author seeks to alleviate concerns about digital curation: a fear of costs, the perception of a frequent need for intervention, the need to keep up with changing formats, and a belief that the item must be identical to the original. The case study takes place in an archive, providing support for the applicability of this model across domains.

The Royal Commission on the Ancient and Historical Monuments of Scotland (RCAHMS) sought to develop a trusted digital repository (TDR) with limited resources. RCAHMS chose OAIS because it defines the functions required to implement a digital preservation plan without limiting organizations to specific technological solutions or procedures. Once the OAIS model was understood, RCAHMS sought stakeholder buy-in and created and implemented a Digital Archive Policy (DAP). After a repository is functional,

new concerns must be addressed, such as continued resources/funding and future technological capabilities.

AUTHENTIC DIGITAL OBJECTS AND TRUSTWORTHY REPOSITORIES

Authenticity of the digital object and trustworthiness of the digital repository are critical factors to the success of a preservation plan and the implementation of a preservation solution.

Center for Research Libraries. "TRAC Metrics." *Archiving and Preservation: Digital Archives*. N.p., n.d. Web. 30 Mar. 2013. <http://crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

The Center for Research Libraries (CRL) uses Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) to evaluate digital repositories. TRAC was developed by OCLC and based on OAIS and test audits completed by CRL. This resource provides a structured list of TRAC metrics, and also links to the full documentation for TRAC and OAIS. The metric is divided into three main sections: organizational infrastructure, digital object management, and technologies, infrastructure and security. In all, the metric includes 50 plus measures of trustworthiness, each of which consists of a general principle and an example of evidence a repository might provide to demonstrate adherence to that principle. Existing institutions expanding their services by creating a digital repository will already have many of the principles – such as written and reviewed policies, procedures and manuals – in place. Other principles address issues specific to digital repositories such as a timetable for reviewing new technological developments.

Dobratz, Susanne, Astrid Schoger, and Stefan Strathmann. "The Nestor Catalogue of Criteria for Trusted Digital Repository Evaluation and Certification." *Journal of Digital Information* 8.2 (2007): n. pag. Web. <http://journals.tdl.org/jodi/index.php/jodi/article/view/199/180>.

The Nestor Catalogue, developed in Germany, provides a set of criteria by which long-term storage repositories can complete self-evaluations at any stage of the planning or implementation process including conception, planning and specification, realization and implementation, and evaluation. These steps may overlap and be repeated because preservation is not a one-time accomplishment but an ongoing process.

The aim of the nestor project is to promote trustworthy repositories in the absence of a centrally recognized formal certification program. The catalogue is intended for commercial and non-profit entities at cultural heritage institutions. The concept of trustworthiness is based on four pillars: integrity, authenticity, confidentiality and availability. All of these need to be defined in relation to the OAIS

concept of a designated community. Nestor provides a 14-point framework for repository evaluation that covers the definition of goals and the designated community, adherence to laws and rules, quality management, object management, strategic planning, archival storage, data management, and IT infrastructure.

Dryden, Jean. "Measuring Trust: Standards For Trusted Digital Repositories." *Journal Of Archival Organization* 9.2 (2011): 127-130. Print.

Dryden provides an illuminating, brief history of the challenges posed by the proliferation of digital technology since the 1980s and responses to those challenges. As early as the 1980s, archivists recognized the challenges posed by the variety and fragility of digital media. In 1996, the Task Force on Archiving Digital Information reported that digital preservation would require significant changes to existing infrastructure, as well as standardization. Dryden covers the OAIS model, the Trusted Repositories Audit & Certification: Criteria and Checklist (TRAC), the nestor project, the DRAMBORA standard (Digital Repository Audit Method Based on Risk Assessment), and finally efforts of the Consultive Committee for Space Data Systems (CCDS) to unite these tools under one standard, ISO 16363. ISO 16363 approaches an international standard for trustworthy digital repositories, but challenges with best practices, file formats and intellectual property rights remain.

Jantz, Ronald. "An Institutional Framework for Creating Authentic Digital Objects." *International Journal of Digital Curation* 4.1 (2009). Web.
<http://www.ijdc.net/index.php/ijdc/article/view/103>.

Jantz discusses methods for establishing the authenticity of digital objects and the steps an institution can take to ensure access to such objects. He explores the processes, the redefinition of roles within the institution, and the technologies required to preserve authentic works of digital scholarship.

The digital archiving and preservation process has three phases: capture and description, archiving and ingest of the object, and preservation over the life of the object. Authenticity must be addressed in each of these phases. With digital objects, this includes the addition of markers to the record that help verify authenticity, such as public key encryption and time stamping. The creator of the metadata and the digital object should digitally sign the object. During the lifecycle of a digital object, it may need to be migrated to a new format. Authenticity must be managed during this lifecycle.

Moore, R. W., and M. Smith. "Automated Validation of Trusted Digital Repository Assessment Criteria." *Journal of Digital Information* 8.2 (2007): n. pag. Web.
<http://dspace.mit.edu/bitstream/handle/1721.1/39091/Moore-Smith.htm?sequence=1>.

Beginning with an explanation of the RLG/NARA trusted digital repository (TDR) certification checklist and how it might be applied to digital repository policy, the authors note that the TDR does not provide guidance for implementing management policies. They nonetheless delineate how to validate a repository's trustworthiness by providing guidelines to implementing rule-based data consistency management.

One of the main concepts addressed is the "preservation environment," that is, a suite of software that protects digital objects from damage from outside forces while also preserving integrity, authenticity, and ability to be displayed. The authors discuss the Dspace digital asset management software, which when used in conjunction with the Storage Resource Broker (SRB) distributed data management software, illustrates the concept of the preservation environment. The article also details the architecture of the Rule-Oriented Data System (iRODS), currently under development at the San Diego Supercomputer Center. iRODS incorporates management policies as rules that guide its data management services.

FILE FORMATS, CHARACTERIZATION, AND REGISTRIES

Digital information presents itself in a wide-ranging variety of formats. The tasks of choosing formats when digital data is created, identifying formats as data comes into a repository, maintaining formats, changing formats, and so on throughout the life cycle of a digital "document" present a myriad of challenges.

McLellan, Evelyn Peters. "Selecting Formats for Digital Preservation: Lessons Learned during the Archivemata Project." *Information Standards Quarterly* 22, no. 2 (2010): 30-33.

McLellan enumerates factors to be considered when choosing file formats for preservation, and discusses the pros and cons of specific file types based on the work of developing Archivemata, a Canadian digital preservation software project. Archivemata is open-source software that supports emulation and migration, as well as the PREMIS, METS, Dublin Core, EAD, and ISAD(G) metadata standards. McLellan focuses on the normalization process whereby Archivemata uses a variety of open-source tools to migrate files into preservation-friendly formats upon ingest. The Archivemata project selects file formats with freely available specifications, no patent or licenses, widespread acceptance by the preservation community, and that have a variety of writing and rendering tools. The last criterion is the most challenging because of the widespread use of proprietary file formats. McLellan discusses common challenges with normalizing proprietary file formats, particularly focusing on moving picture files, Microsoft Office files, and PDF/A files.

Rosenthal, David S. H. "Format Obsolescence: Assessing the Threat and the Defenses." *Library Hi Tech* 28, no. 2 (2010): 195-210.
<http://lockss.stanford.edu/locksswiki/files/LibraryHighTech2010.pdf>

Providing a good overview of the issues associated with format obsolescence and offering an alternative to the current preservation paradigm for handling file format obsolescence, Rosenthal argues format obsolescence is in actuality a rare occurrence.

The current preservation paradigm informed by the fear of format obsolescence is to validate formats, collect metadata, create obsolescence notifications, use registries of format specifications to select format conversions, and migrate files to a "less doomed format" as necessary. Rosenthal argues that file formats at risk for obsolescence are likely young and/or obscure, registries will not be notified of their impending obsolescence, no renderer likely exists, and files will be difficult to verify.

Rosenthal concludes we cannot expect perfect renderability or flawless migration. Instead, we should aim to balance preservation costs with benefits. To reduce costs, Rosenthal suggests trimming preservation down to the essentials and delaying preservation actions as long as possible.

Stanescu, A. "Assessing the Durability of Formats in a Digital Preservation Environment: The INFORM Methodology." *D-Lib Magazine* 10.11 (2004). Web. 7 Apr. 2013.
<http://www.dlib.org/dlib/november04/stanescu/11stanescu.html>

This reading provides insight into the assessment of the durability of digital preservation formats. OCLC developed INFORM (INvestigation of FOrmatS based on Risk Management), a methodology for investigating and measuring the risk factors of digital formats. The assessment yields a risk exposure, useful in informing preservation decisions. According to this methodology, there are six classes of risk: digital object format (includes risks introduced by the file specification, compression algorithms, proprietary vs. open formats, DRM, encryption, and digital signatures), software (encompasses risks introduced by software components), hardware components (type of media, CPU, I/O cards and peripherals), risks inherent in working with associated organizations, the digital archive itself and the migration process (when not covered by any other category). Each of these risk factors is assessed in terms of the probability of its occurring and, if it were to occur, the impact. The resulting computation is the risk exposure.

Thompson, Dave. "A Pragmatic Approach to Preferred File Formats for Acquisition." *Ariadne*, no. 63 (2010). Web.
<http://www.ariadne.ac.uk/issue63/thompson>

Thompson demonstrates a practical strategy for dealing with the issue of diverse file formats on acquisition of digital materials. He provides an overview of the Wellcome Library's approach, a pragmatic approach accepting the reality that the library has little to no influence over the file formats of submissions, just as it has little control over the format of traditional items submitted for archiving. Indeed, creators likely have little control over file format because of cost constraints, business practices, and functional requirements. This creates three types of challenges for preservation: variety of file formats, proprietary formats, and format obsolescence.

Instead of selecting specific file formats acceptable for submission, Wellcome sets out nine principles for evaluating the suitability of a born-digital object for preservation, six of which cover technical considerations while the remaining three evaluate content. This approach has the advantage of being flexible and aligning with archival appraisal practices for analogue materials.

METADATA

Librarians and archivists are well-acquainted with metadata in their respective settings. These readings are intended to provide an overview of concepts relative to preservation metadata in a repository setting. The material covers PREMIS, PREMIS with METS, and the role of archivists in achieving preservation goals, particularly with research data.

Suggested Order: Caplan, Guenther, Wilson

Caplan, Priscilla. *Understanding PREMIS*. Tech. Library of Congress Network Development and MARC Standards Office, 1 Feb. 2009. Web. 10 Mar. 2013.
<http://www.loc.gov/standards/premis/understanding-premis.pdf>

It's critical for those designing and developing a preservation repository to understand what preservation metadata for digital content should be collected, stored, maintained, and exchanged over the long-term. The PREMIS guidelines are relevant whether the specific institutional implementation software is selected from a vendor, is an open source solution, or is developed in-house. The guidelines were developed to be technically neutral, that is, applicable across a wide range of repository contexts. *Understanding PREMIS* provides an overview of the Data Dictionary conventions and entries, the PREMIS data model, and how PREMIS is used. There is a discussion of the interaction of PREMIS, PREMIS-provided XML, and METS that should be reviewed. Finally, Caplan explains what it means for a repository to be "PREMIS-conformant," a goal of many repository development efforts. The PREMIS model and data dictionary provide a framework for thinking about preservation metadata deployment.

Guenther, Rebecca. "Battle of the Buzzwords: Flexibility vs. Interoperability When Implementing PREMIS in METS." *D-Lib Magazine* 14.7/8 (2008): n. pag. July-Aug. 2008. Web. 10 Mar. 2013.
<http://www.dlib.org/dlib/july08/guenther/07guenther.html>

Guenther discusses the ambiguities of using PREMIS with METS, both highly regarded standards for metadata deployment. She underscores that the technical neutrality of PREMIS provides flexibility that must be adapted to the local repository policies and systems decisions. In addition to the data dictionary and model, PREMIS provides XML schemas that may be used alone or with other XML schemas. METS is a respected XML schema, a customizable metadata encoding standard used by a variety of institutions for digital information submission, storage, and especially exchange. This interoperability is a hallmark of METS. But to take advantage of the flexibility of PREMIS and the interoperability of METS presents some challenges. They don't always perfectly mesh. Guenther provides insight into these challenges and addresses the development of a set of guidelines to aid repository developers in meshing the two.

Wilson, Andrew. "How Much Is Enough: Metadata for Preserving Digital Data." *Journal of Library Metadata* 10.2-3 (2010): 205-17. *Routledge (UIUC)*. Web. 15 Mar. 2013.

Wilson suggests that repositories will not do enough to preserve authentic digital data if they do not enlist the expertise of the archival community. He highlights the important role archivists can play in ensuring appropriate metadata is developed to achieve the long-term data preservation goals of institutions with repositories serving the research community.

Different from archival description, recordkeeping metadata has as its primary purpose ensuring a record remains usable over time by focusing on properties of reliability, usability, integrity and authenticity. Recent digital preservation initiatives arising from information technology rather than information management backgrounds seem to have "discovered" the centrality of metadata with no reference to models and standards for preserving digital resources that have been in place in the archival community for a decade. The author states that PREMIS-conformance is insufficient to secure digital objects and that the METS administrative data section, while suitable for transmission, is inadequate for ensuring authenticity.

DIGITAL PRESERVATION AND CURATION: STRATEGIES, METHODS, PLANS, FACTORS

With an understanding of the preservation needs and challenges facing institutions with digital collections, after considering the strategic responses of not only professional organizations, but institutions in active pursuit of goals, after an introductory review of relevant planning models, and with a general

knowledge of some of the tools used to enable digital preservation, it's time to take a closer look at specific methods and plans, as reported by those involved in preservation and curation solutions.

Services, Strategies, Infrastructure

This section offers coverage of specific solutions offered by practitioners in the field. Articles detail how a repository will function, methods of achieving goals, how to evaluate solutions and processes, and more.

Abrams, Stephen, John Kunze, and David Loy. "An Emergent Micro-Services Approach to Digital Curation Infrastructure." *International Journal of Digital Curation* 5.1 (2010): 172-86. Web. 17 Mar. 2013.
<http://www.ijdc.net/index.php/ijdc/article/view/154>

Abrams et al provide a view into the reinvented University of California's Digital Library's digital preservation program that is instructive on several levels. The California Curation Center (UC3), serving 10 UC campuses, replaces a preservation perspective with a curation perspective. The strategic lessons of this article are that a curation solution should be comprehensive in scope, but adaptable to "local" needs and practices and flexible in the face of inevitable changes in both technology and user expectations. The practical lessons of this article are that this can be accomplished through devolving functions into "micro-services" that can be developed and deployed (and replaced) more easily and assembled in configurations suitable to differing "local" campus and departmental environments. The article mentions some of the design principles. For a more comprehensive, entertaining, and less technical report of this endeavor, read the White Paper found [here](#).

Barateiro, Jose, Goncalo Antunes, Filipe Freitas, and Jose Borbinha. "Designing Digital Preservation Solutions: A Risk Management-Based Approach." *International Journal of Digital Curation* 5.1 (2010): n. pag. Web. 10 Mar. 2013.
<http://www.ijdc.net/index.php/ijdc/article/view/143>.

Acknowledging that the digital preservation community uses risk management concepts to assess existing repositories, the authors suggests that this be taken a step further, and that risk management concepts be used to design and assess preservation solutions, too. This article is instructive in three ways. First, as prelude to their proposals, the authors summarize digital preservation system requirements (e.g., authenticity assurance) and the main threats (e.g., disasters) and vulnerabilities (e.g., obsolescence). Second, they present a succinct discussion of digital preservation techniques and strategies to address the threats and vulnerabilities, rather broadly conceived (e.g., migration and emulation, as well as metadata and auditing). Finally, they mesh all this information into a table that shows how the

techniques can be used to address the risks associated with the threats and vulnerabilities.

Becker, Christoph, Hannes Kulovits, Mark Guttenbrunner, Stephan Strodl, Andreas Rauber, and Hans Hofman. "Systematic Planning for Digital Preservation: Evaluating Potential Strategies and Building Preservation Plans." *International Journal of Digital Libraries* 10 (2009): 133-57. Springer. Web. 5 Mar. 2013.

This article explicates a concrete, systematic method of defining and developing a preservation plan that relies on empirical evidence and constructs a repeatable and transparent workflow, ideally supporting automated documentation of decisions and actions. It details why such a plan is needed, the dynamic and precise elements it should include, how these elements will define and continue to refine the repository, the executable triggers behind the repository's functions, and the importance of measurable preservation goals. This methodology specifically addresses the preservation planning functional entity of the OAIS Model and supports criteria for trustworthy repositories as defined by TRAC and nestor. Requirements are quantifiable. Alternatives are evaluated using controlled experimentation. Results are analyzed, aggregated, and consolidated. The preservation plan includes a complete evidence base of decision making. The article concludes with an examination of three case studies, their results, and lessons learned.

Cramer, Tom, and Katherine Kott. "Designing and Implementing Second Generation Digital Preservation Services: A Scalable Model for the Stanford Digital Repository." *D-Lib Magazine*, Sept.-Oct. 2010. Web. 3 Feb. 2013.
<http://www.dlib.org/dlib/september10/cramer/09cramer.html>

The SDR was the first noteworthy digital preservation project undertaken by Stanford, and has succeeded in its original purpose of holding massive quantities of digital objects (over 80 TB). Content ranges across many digital types (e.g., library collections, theses, research data, and university archival materials). After taking into account three years of SDR's operational experience, as well as changes in Stanford's technological and organizational environment, the authors propose a structural, architectural, and strategic revision. In their analysis of SDR's live operational history, the authors first summarize the challenges the system has overcome (such as storage and data modeling problems). Thereafter, they describe how SDR was redesigned in 2009 with the aid of the TRAC Checklist as a preservation model and the adoption of Fedora as a systems design platform. As of the writing of the article, Stanford was actively aiming to simplify their data models and architecture.

Institutional Repositories: Scholarly Discourse

This section examines the traditional core function of institutional repositories – deposit of scholarly discourse

materials. The two articles attempt solutions to the problem of ensuring faculty deposits: one is automated and one is mediated.

Anuradha, K. "Design and Development of Institutional Repositories: A Case Study." *The International Information & Library Review* 37.3 (2005): 169-78. Print.

Anuradha walks through the process of developing and maintaining an Institutional Repository (IR) named PRABHAVI at the Indian Institute of Science (IISc) in Bangalore, India. Anuradha focuses on the technical requirements for an IR that automatically captures works by IISc faculty, as opposed to relying on uploads by faculty or IR staff. Although PRABHAVI is largely automated, human intervention is also required. Because of the difficulty in accurately identifying duplicates, potential duplicates identified by the algorithm are checked by IR staff for confirmation. IISc faculty also have the option to manually upload digital objects that were missed by PRABHAVI's automated process. PRABHAVI's automated process represents one answer to the concern that authors will be reluctant to self-archive. It also could be a practical approach to collecting previously published works when an IR is first established. Anuradha briefly comments on the role of IRs in helping developing countries contribute to mainstream knowledge generation.

Bevan, Simon J. "Developing an Institutional Repository: Cranfield QUEprints – a Case Study." *OCLC Systems & Services* 23.2 (2007): 170-82. Web.
<http://search.ebscohost.com.proxy2.library.illinois.edu/login.aspx?direct=true&db=aph&AN=25780644&site=ehost-live>.

This case study of the QUEprints IR at Cranfield University in the UK demonstrates a successful utilization of library staff for archiving as opposed to a self-archiving system whereby faculty submit their own work to the IR. Cranfield chose to build the IR on DSpace software because of its broad-based support in the community. During the planning process, concerns arose over the likelihood of faculty adopting a self-archiving system. Consequently, Cranfield chose a mediated archiving system whereby all materials are added to the IR by library staff. The main concern of such a system is cost, but QUEprints was established by reallocating existing library resources. Cranfield determined continuing support was merited because of the increase in use of the system over its first few years which correlated with increases in deposits of materials. Once QUEprints is better established, it is hoped that faculty will see its value and become more open to changing to a self-archiving system.

Digital Repositories: Libraries, Archives (and Beyond)

In considering a strategic approach, examples of institutional repositories that are pushing the scope of content and services beyond traditional boundaries are provided. Inter- and intra-institutional collaboration have been discussed. The next

example provides insight into integrating a range of varied digital resources from previously distinct entities.

Keskitalo, Esa-Pekka. "Costs and Benefits of a Shared Digital Long-Term Preservation System." *Liber Quarterly* 21.1 (2007): n. pag. 2007. Web. 30 Mar. 2013.
<http://liber.library.uu.nl/index.php/lq/article/view/8008/8346>.

Keskitalo summarizes the results of a cost-benefit analysis of creating a National Digital Library (NDL) in Finland, integrating digital materials of libraries, archives and museums. The NDL proposes to combine five existing systems that in total serve over 200 libraries, archives, and museums. The NDL is intended to be flexible and capable of preserving different kinds of materials and working with a variety of end users. The analysis projects significant savings of connecting the systems. While the actual costs of a national repository may not be relevant to university library professionals, the factors considered in the analysis – the amount of materials, software replacement, migrations, server space, and human resources costs – are pertinent to projects of all sizes. The article also provides a basic introduction to the LIFE (lifecycle information for e-literature) model created by University College London, the British Library and the Joint Information Systems Committee (JISC) to define the lifecycle of digital materials.

Human Factors in Preservation

An important consideration in planning a repository is the human component, including, for example, building relationships, education and training, and staffing, both during implementation and ongoing to achieve sustainability.

Berman, Fran, Ardys Kozbial, Robert H. McDonald, and Brian E.C. Schottlaender. "The Need to Formalize Trust Relationships in Digital Repositories." *EDUCAUSE Review* 43.3 (2008): 10-11. Print.

While many articles mention collaborative relationships, this one focuses on how to form successful partnerships and includes the roles and responsibilities each partner should have, how these should be defined, and how trust can be built between the partners.

The Chronopolis pilot is a data grid that uses data replication as a method of data preservation, similar to LOCKSS. Each participant in Chronopolis can take multiple roles, including user, partner, and provider. These roles can even vary for a partner depending on the data collection. Defining these roles and having clear expectations for the outcomes of each has been a key to the success of the collaboration. Moreover, in order for participants to work together, there is a need to formalize trust relationships. The article articulates four components of trust: credibility, reliability, intimacy and self-interest, with the last being weighted most heavily. Each participating institution should have a formal trust relationship with the other partners.

Carlson, Jake. "Demystifying the Data Interview: Developing a Foundation for Reference Librarians to Talk with Researchers about Their Data." *Reference Services Review* 40.1 (2012): 7-23.
http://docs.lib.purdue.edu/lib_research/153/.

Carlson's goal is for librarians to have the skills and ability to conduct discussions with researchers to ensure that their research results are managed, shared and curated in the best possible way. While geared toward reference librarians and the Data Curation Profiles Toolkit used at Purdue University, the article provides succinct definitions to data curation terms and concepts useful to all employees in the digital preservation field, and foregrounds the importance of the relationship between researcher and librarian.

Because institutional repositories tend to be underused, librarians, particularly reference librarians, can serve as a liaison between researchers and institutional repositories by using skills already common within their field, including negotiation skills, advocacy, marketing, promotion, expectation management, data services, documentation, and best practices adherence. Librarians can then assume the role of trusted data consultant, being involved during the entire lifecycle of the data, ensuring that the data is curated in a manner that suits the community standards and provides long-term access.

Kennedy, Marie R. "Reformatting Preservation Departments: The Effect of Digitization on Workload and Staff." *College & Research Libraries* 66.6 (2005): 543-551.
<http://crl.acrl.org/content/66/6/543.full.pdf+html>.

This study elucidates how digitization is affecting the workload of preservation departments. Workloads and staffing are key to understanding the organizational changes needed to accommodate the digitization of materials, an essential part of the planning process in forming a digital repository. Eighteen member libraries of the Association of Research Libraries (ARL) accumulated data over five years to track the number of reformatting tasks completed and staffing trends in preservation departments, where the majority of digitization work is completed.

Cloud Storage and Digital Preservation

Cloud storage is being suggested by many to be a part of a digital preservation scheme.

Rosenthal, David S. H., and Daniel L. Vargas. "Distributed Digital Preservation in the Cloud." 8th Annual Digital Curation Conference, Jan. 2013. Web.
<http://www.lockss.org/locksswp/wp-content/uploads/2013/01/IDCC2013.pdf>.

This article covers using cloud storage as a form of digital preservation within a "Lots of Copies Keep Stuff Safe" distributed digital preservation scheme. Cloud storage has

become increasingly common, thus, any new digital repository should explore it as an option and understand the potential pros and cons, with regard to cost, implementation, and security of objects held within the repository. Special focus is paid to the economic and technical feasibility of cloud storage vs. the use of local disks. This article makes cloud storage understandable and highlights both the positives and negatives of its use.

COPYRIGHT ISSUES

This section addresses intellectual property issues that are relevant to those undertaking digital preservation activities.

Besek, June M. "Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment." Washington, DC: Council on Library and Information Resources, 2003. Web. <http://www.clir.org/pubs/reports/pub112/pub112.pdf>.

This article, commissioned by the National Digital Information Infrastructure and Preservation program, provides a comprehensive overview of copyright issues relevant to the creation of a digital archive. It elucidates how to navigate the legal complexities of digital librarianship projects, points out which intellectual-property areas are ambiguous, and addresses how librarians might go about studying and unpacking such issues. The article clearly lists ways in which librarians might obtain a work and also offers concise and helpful summaries of the contractual and copyright constraints that are applicable in a variety of common acquisition scenarios. The article's intended readers are those in the process of developing digital repositories who need both an overview of the fundamental information about copyright statutes and practical guidance on how to apply this information to their preservation project plans.

TUTORIALS, VIDEOS, WEBSITES

The following list of online resources should be explored.

Association for Library Collections and Technical Services. American Library Association, n.d. Web. 5 Mar. 2013. <http://www.ala.org/alcts/>.

This American Library Association (ALA) divisional website contains information about digital preservation and is a good source for both those new to the field and seasoned professionals. Notable are free webinars on topics of interest, such as "The Role of Long-Term Storage in Digital Curation," "The Art of Selecting Digital Content to Preserve," and "Steps in a Digital Preservation Workflow."

Additionally, the website is an excellent source of information about PARS: the Preservation and Reformatting Section of the ALA. PARS focuses on ensuring access to library materials through applying new technologies. Preservation of digital materials is a large part of this

mission. There are links to interest groups, where one can seek the wisdom of others in the field. Topics include acquisitions, collection development, digital preservation, electronic resources and metadata.

Center for Research Libraries. *Archiving and Preservation: Digital Archives*. N.p., n.d. Web. 24 Mar. 2013. <http://crl.edu/archiving-preservation/digital-archives>.

The Center for Research Libraries (CRL) provides a variety of digital archiving resources in support of its mission to support research in the humanities, sciences and social sciences. Resources include profiles of established repositories, reports, guiding principles, metrics and assessment tools. An advisory panel of professionals from research and academic libraries across the U.S. vets tools and materials. CRL provides profiles and audit reports on repositories such as HathiTrust, LOCKSS, Portico, the Associated Press, LexisNexis, the Internet Archive and the Government Printing Office. Tools include the Trusted Repository Audit Checklist (TRAC) and Trusted Digital Repository Checklist (TDR) also known as ISO 16363. CRL focuses primarily on establishing trusted repositories through auditing and metrics. For professionals who already possess a working knowledge of digital repositories, CRL provides a wealth of tools and profiles.

Digital Preservation. Library of Congress, n.d. Web. <http://www.digitalpreservation.gov/>.

This Digital Preservation page is a portal to a wealth of resources on the subject provided by the Library of Congress. The resources are part of the National Digital Information Infrastructure and Preservation Program, which aims to preserve digital resources at the national level via partnerships with other organizations in the public and private sectors. The site is better suited to in-depth learning than to casual browsing; however, the Digital Preservation Outreach and Education section (<http://www.digitalpreservation.gov/education/curriculum.html>) provides six short videos illustrating how the LOC identifies, stores, protects, manages and provides access to its own digital collections. Individuals can also keep up-to-date by viewing digital preservation podcasts (<http://www.digitalpreservation.gov/multimedia/>) and subscribing to The Signal blog (<http://blogs.loc.gov/digitalpreservation/>), which not only covers digital preservation news at the LOC, but frequently links to outside news, resources and events.

"Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems Tutorial." *Digital Preservation Management Workshops and Tutorial*. Cornell University, n.d. Web. 13 Mar. 2013. http://www.dpworkshop.org/dpm-eng/eng_index.html.

This informative tutorial from Cornell University contains definitions, key concepts, practical advice and exercises. It is an excellent introduction for those new to the field, as it will give readers an understanding of the terms, issues, and priorities, preparing them to explore additional readings with a solid foundation.

The site provides the background, concepts, and definitions necessary to understand the Reference Model for an Open Archival Information System (OAIS) and Attributes of a Trusted Digital Repository (TDR). Challenges to successful digital preservation are discussed, including organizational, technical, and resource challenges. The technical issues of obsolescence and physical threats are explored, including format, software, hardware and media obsolescence.

The Digital Preservation Network. Digital Preservation Network, n.d. Web. 11 Mar. 2013.
<http://www.dpn.org/>.

This website features a collaborative effort among the nation's leading universities to develop a federated approach to preservation to ensure the scholarly record is available to future generations. Local repositories are contributing nodes; DPN creates federated replicating nodes that contain redundant dark copies of all deposits. While this effort is still under development, membership in this organization might prove useful through exposure to and partnering with institutions at the forefront of digital preservation. Additionally, some services could be provided through the DPN rather than developed locally.

ACKNOWLEDGEMENTS

Compiled and annotated by Sandi Caldron, Liz Corley, Jan Emery, and Jennifer Jacobsen-Wood.

COPYRIGHT STATEMENT

Rights belong to the authors.