# InterPARES 3 Project

International Research on Permanent Authentic Records in Electronic Systems

TEAM Canada

**Title:** **General Study 01 – Past and Present Digital Preservation Projects Devoted to the Long-term Preservation of Digital Records and Digital Information**

**Status:** Final (public)

**Version:** 1.4

**Date Created:** November 2007

**Last Revised:** January 2009

**Author:** The InterPARES 3 Project, TEAM Canada

**Writer(s):** Donald C. Force
School of Library, Archival and Information Studies,
The University of British Columbia

**Project Component:** Research

**URL:** http://www.interpares.org/rws/display_file.cfm?doc=ip3_canada_gs01_digital_preservation_projects.doc

# Document Control

| Version history | | | |
|---|---|---|---|
| Version | Date | By | Version notes |
| 1.0 | 2007-11-09 | D. Force | Submitted for review |
| 1.1 | 2007-01-12 | D. Force | Added appendix |
| 1.3 | 2008-04-22 | D. Force | Incorporated revisions suggested by L. Duranti |
| 1.4 | 2009-01-26 | R. Preston | Minor content and copy editing |

# Table of Contents

## Introduction

Since 1999, the InterPARES Project has been examining the complex issue of preserving digital records. From 1999-2001, the first phase of the Project examined issues of authenticity and reliability from the preserver's perspective. InterPARES 2, 2002-2007, shifted views and looked at how record creators created and interacted with records generated in interactive, experiential and dynamic systems. For the next five years, 2007-2012, InterPARES 3 will "translate the theory and method of digital preservation drawn from research to date into concrete action plans for existing bodies of records that are to be kept over the long term by archives—and archival/records units within organization—endowed with limited resources."[1]

InterPARES is not alone in its efforts to address the issues concerned with the long-term preservation of digital records. Since 1999, various research projects and related initiatives have been devoted to similar purposes. This bibliographic report discusses the current projects and some of their findings. While the following pages briefly cover these projects, readers should also consult the annotated bibliography in Appendix 1 for additional resources and information. Given the dynamic nature of the digital preservation environment, readers should consider this a constantly evolving document, to which periodic revisions will be made to incorporate newly discovered or updated information, including information about new projects.

## Citations of InterPARES Research

An overall assessment of the current projects dealing with the long-term preservation of digital information shows that the findings of the first two phases of InterPARES have not yet, to any appreciable degree, permeated the archival and information science communities. Furthermore, InterPARES documentation has sparsely been cited by these other projects and, in most cases, has only been cited by those individuals who have had some involvement with InterPARES at one stage or another.[2] A notable exception is K. Hänström's 2007 article "Authenticity in a Digital World: Long Term Preservation of Public Records,"[3] which discusses the results of a Swedish study that focused "on the problem of authenticity concerning long term preservation of electronic official documents [in an effort] to find out whether authenticity problems can occur when governmental agencies create and later transfer electronic records to the [Swedish] National Archives for long term preservation." In this study, Hänström analyzes Swedish legislation to identify the legal requirements on the preservation of authentic digital

[1] Luciana Duranti (2007), "Theoretical Elaborations into Archival Management in Canada (TEAM Canada): Implementing the theory of authentic records in digital systems in small and medium-sized archival organizations," SSHRC-CURA InterPARES 3 Project Proposal, 11. Available at http://www.interpares.org/rws/display_file.cfm?doc=CURA_IP3_Proposal_Parts_1-7_FINAL.pdf.

[2] See, for example, Malcolm Todd (2006), "Power, identity, integrity, authenticity, and the archives: a comparative study of the application of archival methodologies to contemporary privacy," *Archivaria* 61 (Spring): 181–214; Michèle V. Cloonan and Shelby Sanett (2005), "The preservation of digital content," *Libraries and the Academy 5*(2): 213–237; Francesca Marini (2007), "Archivists, librarians, and theatre research," *Archivaria* 63 (Spring): 7–33, esp. footnotes 77 and 87; and Eliot Wilczek and Kevin Glick (2006), "Tufts and Yale: Fedora and the Preservation of University Records." Available at http://dca.tufts.edu/features/nhprc/index.html. See especially Part 1.4: Glossary, available at http://repository01.lib.tufts.edu:8080/fedora/get/tufts:UA069.004.001.00004/bdef:TuftsPDF/getPDF.

[3] See Kenneth Hänström (2007), "Autenticitet i en digital värld: långsiktsbevarande av allmänna handlingar [Authenticity in a Digital World: Long Term Preservation of Public Records]," *Human IT* 9(1): 67–109. An English abstract is available at http://www.hb.se/bhs/ith/1-9/kh.htm. The Swedish text article is available at http://www.hb.se/bhs/ith/1-9/kh.pdf.

government documents and tests these legal requirements against the findings of the InterPARES 1 Project's Authenticity Task Force, particularly its *Template for Analysis*[4] and its *Requirements for Assessing and Maintaining the Authenticity of Electronic Records* (i.e., the benchmark and baseline requirements).[5]

Several terms and definitions from the InterPARES 2 Project are cited in the newly released MoReq2 Specification.[6] As well, both the InterPARES 2 Project and the "Preservation of the Integrity of Electronic Records Project" (i.e., the UBC-MAS Project), the precursor project to phase one of the InterPARES Project, are listed in Appendix 1 of MoReq2 as two of the ten key references consulted during preparation of the MorReq2 Specification. A number of terms and definitions from the InterPARES 1 Project's Glossary are also cited in the ANSI/ARMA 16-2007 standard "The Digital Records Conversion Process: Program Planning, Requirements, Procedures."[7] This standard, which outlines requirements for "ensuring that electronic records remain authentic and trustworthy as they are converted from one digital recordkeeping system to another,"[8] was approved as an American National Standards Institute (ANSI) standard on 1 March 2007. Likewise, some definitions from the InterPARES 2 Project's Glossary are cited in a 2006 article by Swedish archivist, Ulf Andersson, Archives and Records Manager for the pharmaceutical company, AstraZeneca.[9] Finally, it is noted that all the most recently issued and the current draft ISO TC46-SC11 standards cite InterPARES among their references, especially those concerning metadata[10] and long-term preservation of digital records.[11]

Several individuals and organizations also cite the existence of the InterPARES Project and/or the availability of its products. See, for example, reference to InterPARES 1 and 2 resources by UNESCO's *Memory of the World* project.[12] Some other project-related examples include the Network of Expertise in Long-Term Storage and Long-Term availability of Digital Resources (NESTOR) project in Germany, which provides a passing reference to the InterPARES Project in its 2007 handbook "Eine Kleine Enzyklopädie der Digitalen Langzeitarchivierung [A Small Encyclopedia of Digital Long-term Archiving],"[13] and the Preservation and Reformatting Section (PARS) of the Association for Library Collections & Technical Services (ALCTS), which cites the InterPARES Project as one of several resources consulted during its efforts to draft a definition for the concept "digital preservation."[14] The InterPARES Project is identified as an external liaison organization for the International Organization for Standardization's (ISO) Technical Committee 46, Subcommittee 11 (Archives/Records Management).[15] Finally, Yale

[4] See http://www.interpares.org/book/interpares_book_j_app01.pdf.
[5] See http://www.interpares.org/book/interpares_book_k_app02.pdf.
[6] See Serco Consulting (2008), "Model Requirements for the Management of Electronic Records, Update and Extension, 2008: MoReq2 Specification," European Commission. Available at http://www.moreq2.eu/downloads.htm.
[7] See ARMA Standards Development Committee Task Force, *The Digital Records Conversion Process: Program Planning, Requirements, Procedures (ANSI/ARMA 16-2007)* (Lenexa, KS: ARMA International, 2007). See especially "Appendix C - Terms and Definitions Comparison Table," 28–32.
[8] See http://www.arma.org/bookstore/productdetail2.cfm?ProductID=2229.
[9] See Ulf Andersson (2006), A Concept for Governing the Complete Life Cycle of Information," AstraZeneca, 10 September 2006 (unpublished), 4.
[10] See, for example, ISO/TS 23081-2:2007, Information and documentation – Records management processes – Metadata for records – Part 2: Conceptual and implementation issues.
[11] See, for example, ISO/CD TR 26102, Information and documentation – Requirements for long-term preservation of electronic records.
[12] See http://portal.unesco.org/ci/en/ev.php-URL_ID=24616&URL_DO=DO_TOPIC&URL_SECTION=201.html.
[13] Version 0.1, p. 10. Available at http://nestor.sub.uni-goettingen.de/handbuch/nestor-Handbuch_01.pdf.
[14] See PARS' blog page, "Defining Digital Preservation: Background Documents on the Definition of Digital Preservation" at http://blogs.ala.org/digipres.php?cat=327.
[15] See http://www.iso.org/iso/about/organizations_in_liaison/organizations_in_liaison_details.htm?id=514509.

University's Beinecke Rare Book and Manuscript Library provides a notable example of another institution incorporating an InterPARES resource into a tangible product. In fact, the Beinecke used the InterPARES 2 Project's *Creator Guidelines* as the basis for its recently drafted "Authors' Guidelines for Digital Preservation."[16]

Examples of individuals unaffiliated with InterPARES who cite the Project include Stephen Mason, barrister and former Director of the Digital Evidence Research Programme at the British Institute of International and Comparative Law,[17] who, in his 2006 article, "Proof of the Authenticity of a Document in Electronic Format Introduced as Evidence," cites the InterPARES Project in his bibliography and the Project's Web site (as well as the US InterPARES Web site) in the article's "Some web sites of interest" section.[18] Another example of an individual citing the Project is provided by British Columbia archivist-historian, David Mattison, who cites several InterPARES 2 Project resources (*Policy Framework*, *Creator Guidelines*, *Preserver Guidelines*) in his *The Ten Thousand Year Blog*.[19]

Finally, there are a couple of other situations where InterPARES has been cited for reasons other than its findings or products. For example, in 2004, the Electronic Resource and Preservation Access Network (ERPANET), which sought to gather information on the awareness of digital preservation, digital preservation strategies and the future requirements for digital preservation, considered, in the process of deciding how best to collect these data, using the methodology implemented by InterPARES but decided it was "too detailed and lengthy" and too narrow in scope for ERPANET's own interests.[20] Another article referenced InterPARES as an example of a successful "collaborative research programme."[21]

The relative dearth of literature citing InterPARES may be attributed to several factors. First, given the recent conclusion of InterPARES 2, it may be too early to gage the reaction and usage of its findings within the different professional spheres. A good example of this may be S. Katuu's 2006 "Interdisciplinary Investigation of the Authenticity and Long-term Preservation of Electronic Records," which summarizes the work of InterPARES 1 and discusses what InterPARES 2 sought to accomplish.[22] Second, as will be discussed below, the recent trend in long-term preservation projects seems less concerned with digital or electronic records and more focused on the long-term preservation of "digital information" and "data." Third, it is likely that some other projects use the findings of InterPARES without saying so explicitly.

## Existing Projects Addressing the Long-term Preservation of Digital Records

Where once projects such as the Victorian Electronic Records Strategy (VERS), CURL Exemplars in Digital Archives (Cedars), Creative Archiving at Michigan & Leeds: Emulating the Old on the New (CAMiLEON), and projects at the University of British Columbia, Pittsburgh,

[16] See http://www.library.yale.edu/~nkuhl/AuthorsGuidelines.pdf.
[17] See http://www.biicl.org/about/.
[18] Stephen Mason (2006), "Proof of the Authenticity of a Document in Electronic Format Introduced as Evidence," ARMA International Education Foundation, October 2006. Available at http://www.armaedfoundation.org/images/Proof_of_authenticity_of_a_document.pdf.
[19] See http://www.davidmattison.ca/wordpress/?p=2049.
[20] ERPANET (2003), "Broadcasting," erpastudies, 8. Available at http://www.erpanet.org/studies/docs/erpaStudy_broadcasting.pdf.
[21] Karen Anderson (2007), "Global Archive and Record-keeping Research Agendas: Encouraging Participation and Getting over the Hurdles," *The Journal of the Society of Archivists* 28(1): 35–46.
[22] Shadrack Katuu (2006), "Interdisciplinary Investigation of the Authenticity and Long-term Preservation of Electronic Records." *South African Journal of Information Management* 8(2).

Indiana University and Yale dominated the landscape of electronic records in late 1990s and the first few years of the twenty-first century, today the scene is much different. Currently, there are only a few projects specifically devoted to the long-term preservation of digital records. One of the most recent projects has been "Preserving the Records of the E-Legislature," carried out by the Minnesota Historical Society. Begun in 2005 and concluded in 2007, this project sought to develop and "implement a plan to preserve … digital legislature records."[23] At the time of this report, there was little other available information regarding the procedures and findings of the project and its final report will not be available until sometime in early 2008. In addition to this project, its primary funder, the National Historical Publications and Records Commission (NHPRC), has also supported various other digital records preservation projects in the United States, although few have created Web sites that document their progress and findings.[24] Despite the dearth of current, on-going projects solely devoted to digital records, a more recent trend has been to take a more granular approach to the issue.

One of the more theoretical projects that analyzed a couple of components of digital records was the Significant Properties (InSPECT) project. Supported by the Arts and Humanities Data Service (AHDS), the project is scheduled to operate from April 2007 through September 2008 and aims to "examine the whole concept of significant properties, determine which properties are significant for a range of object types and assess the importance of each of these for future representation of the object…."[25] By looking at raster images, e-mails, structured texts and digital audio, the project hopes to identify the essential characteristics of a digital object so that better preservation techniques may be applied to ensure the long-term survival of such objects. At the time of this report, the project had not yet published any of its findings. Other projects have taken a less theoretical approach to specific forms of digital records.

The increased usage and dependence on e-mail has provided researchers with a specific situation in which they may test and apply established theories and conceptions of digital records.[26] In 2003, Loughborough University orchestrated the "Institutional Records Management and E-mail" project. Within a short six-month span, the project performed six different case studies at the University that examined "current and possible practices for dealing with e-mail at a corporate, department, or group level rather than a personal level, and evaluated various technical options for archiving e-mail in line with the policies and legal requirements."[27] The project concluded that establishing an e-mail policy should be the main priority of an organization that uses e-mail and that the implementation of this policy should be overseen by a records manager and provide specific guidelines for acceptable use that the organization's e-mail users should be able to follow.[28]

---

[23] Minnesota State Archives (2008), "Preserving the Records of the E-Legislature." Available at http://www.mnhs.org/preserve/records/elegislature/elegislature.htm.

[24] For a complete list of electronic records projects funded by the NHPRC, including links to available Web sites, see http://www.archives.gov/nhprc/projects/electronic-records/projects.html.

[25] AHDS (2007), "About InSPECT," 8 August 2007 update version. Available at http://www.significantproperties.org.uk/about.html.

[26] Rich Dymalski and Jerry Kirkpatrick, "One County's Attempt to Move from 0 to Digital in Record Time," in *New Skills for the Digital Era: A Colloquium sponsored by National Archives and Records Administration Society of American Archivists Arizona State Library, Archives and Public Records, 31 May - 2 June 2006, Washington, DC*., Richard Pearce Moses and Susan E. Davis, eds. (Chicago, IL: Society of American Archivists, 2008), 105–122. Available at http://www.archivists.org/publications/epubs/NewSkillsForADigitalEra.pdf; Patricia Galloway, Marlan Green, Stan Gunn and Sue Soy (2002), "Coming to TERM. Designing the Texas Email Repository Model," *D-Lib Magazine* 8(9). Available at http://www.dlib.org/dlib/september02/galloway/09galloway.html.

[27] Institutional Records Management and E-mail (http://www.lboro.ac.uk/computing/irm/).

[28] See Michael Norris (2003), "Institutional Records Management and E-mail: Final Report," November 2003. Specifically, section 7.1. Available at http://www.lboro.ac.uk/computing/irm/final-report.html.

More recently, projects such as the "Digital archiving in Flemish institutions and administrations" (DAVID)[29] and "Managing the Digital University Desktop" (MDUD),[30] along with several individual studies have further explored this issue.[31] While the MDUD project has only made available a couple of presentations and an online tutorial regarding the management of e-mail, the DAVID project has elaborated on at least one issue that is at the core of InterPARES: authenticity.[32] In his report "Filing and Archiving E-mail," Filip Boudrez asserts that, to ensure the authenticity of e-mails, "all essential components [must] … be archived," including their archival context, their content (e.g., subject and attachments) and their structure (e.g., header and message field).[33] In his assessment, Boudrez offers a preservation solution based on a combination of eXtensible Markup Language (XML) and migration—a process currently being used by the Antwerp city archives.[34] These projects aside, it may be only assumed that, with the increasing reliance on e-mail, additional projects will emerge and continue to address issues and concerns related to this type of digital record.

One of the dominant current trends among digital preservation projects is to focus on developing information systems that ingest, manage and retrieve "data," "digital objects," "digital information" and/or "digital content" based on the Open Archival Information System (OAIS) model. Where InterPARES 2 based its Chain of Preservation Model on the OAIS model, adapting it to archival theory and methods,[35] other projects have followed the model more literally, incorporating also its language. This has resulted in an often confusing lexicon of digital preservation terminology. For example, the Co-operative Development of a Long-Term Digital Information Archives (KOPAL), which is currently developing tools to ensure the long-term access to electronic publications, uses the terms "digital documents" and "digital data" interchangeably in its mission statement,[36] while the "Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval" (CASPAR) project explores ways to "handle the preservation of the digital resources of many user communities."[37] The "Securing a Hybrid Environment for Research Preservation and Access Development Partner" (SHERPA DP2) project, which is devoted to preserving electronic publications, works to develop its "OAIS-based distributed preservation model … to interact with repositories holding different and varied types of digital content."[38] The recently established "Repository for Preservation of Authentic Digital Records" (RODA) project

[29] See http://www.expertisecentrumdavid.be/eng/edavid.php.
[30] See http://www.ils.unc.edu/digitaldesktop/.
[31] See also Maureen Pennock (2006), "Curating E-mails: A Life-Cycle Approach to the Management and Preservation of E-Mail Messages," July 2006, version 1.0 in *DCC Curation Manual*, S. Ross and M. Day, eds. Available at http://eprints.erpanet.org/113/01/curating-e-mails.pdf; and Marlan Green, Sue Soy, Stan Gunn and Patricia Galloway (2002), "Coming to TERM: Designing the Texas Email Repository Model," *D-Lib Magazine* 8(9). Available at http://www.dlib.org/dlib/september02/galloway/09galloway.html.
[32] It should be noted that the DAVID Project has been for many years a partner of InterPARES and there has, as a consequence, been much osmosis between the two projects.
[33] Filip Boudrez (2006), "Filing and Archiving E-mail," eDavid. Available at http://www.expertisecentrumdavid.be/docs/filingArchiving_email.pdf. Maureen Pennock also discusses the importance of authenticity and integrity in her article "Curating E-mails" (see previous note for full citation), although she discusses the concepts within the context of "digital objects" (see pp. 9–10).
[34] Ibid., 35–39.
[35] See Kenneth Thibodeau et al., "Part Three – Trusting to Time: Preserving Authentic Records in the Long Term: Preservation Task Force Report," in *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, Luciana Duranti, ed. (San Miniato, Italy: Archilab, 2005), 99–116. Online reprint version available at http://www.interpares.org/book/interpares_book_f_part3.pdf.
[36] See http://kopal.langzeitarchivierung.de/index.php.en.
[37] David Giaretta (2007), "D1202: CASPAR Guidelines," 5. Available at http://www.casparpreserves.eu/Members/cclrc/Deliverables/caspar-guidelines/at_download/file.
[38] Sherpa DP2 Project, "Sherpa DP2: Project Overview," 30 May 2007 update. Available at http://www.sherpadp.org.uk/sherpadp2.html.

seeks to "provide technical solution to digital preservation at the national level by developing processes, tools and resources capable of answering the needs of public administration institutions in terms of preservation of the digital items that are being produced."[39] Finally, the current "Digital Repository Infrastructure Vision for European Research" (DRIVER) project intends "to build the testbed for a future knowledge infrastructure of the European Research Area" by developing a digital infrastructure "for computing resources, data storage and data transport [that] will deliver the content resources, i.e. any form of scientific output, including scientific/technical reports, working papers, pre-prints, articles and original research data…" In other words, there is no limit to the types of information that these systems intend to incorporate; and they make little, if any, effort to make distinctions between data, documents or records.

Several factors have contributed to this confusion. First, as Mary Vardigan and Cole Whiteman observe, because the OAIS model was designed to preserve scientific data, the terminology used by the Consultative Committee for Space Data Systems (CCSDS)—the designers of the OAIS model—does not easily correspond to archival contexts.[40] Second, little effort has been made on the part of the system designers to thoroughly define, or at the very least, thoroughly document, the concepts and terminology used in their reports and discussion papers. Finally, many system designers appear to neglect the importance of distinguishing what objects are being preserved, especially as regards their context(s), choosing to focus, instead, almost exclusively on management and preservation of digital content at the byte level.[41] In a rather alarming statement, a recently published white paper asserts that one of the major requirements for an OAIS-based preservation storage system is that it must "[s]upport a *graceful loss* of data. Some portions of the data are likely to be lost or become corrupted over time. If some data is lost, a good preservation system must minimize the economic effect of this data loss and prevent cases where data in the system that is still intact cannot be read or interpreted."[42] Given this statement, one may only assume that within the next five or ten years, success and failure stories will emerge discussing how the lack of clear definitions of the terminology and visions of the preservations systems affected organizations when pressured to retrieve specific instances of their "digital assets."

Not all projects intend to preserve just "digital information;" there are numerous other on-going or current projects with more specific intentions. For example, "Preservation Eprint Services" (PRESERV)[43] and Portico,[44] a not-for-profit electronic archiving service initiated by JSTOR in 2002, are currently exploring methods to preserve electronic publications. Likewise, the recently completed "Exploring Collaborations to Harness Objects in a Digital Environment for Preservation" (ECHO DEPository) project sought to develop tools "for selecting and capturing materials published on the Web."[45]

Other projects have more general objectives. The "Data Preservation Alliance for the Social Science" (Data-PASS) project, for example, aims to identify, acquire and preserve "data at-risk of being lost to the social science research community," such as those data found in

[39] RODA, "About RODA." Available at http://roda.iantt.pt/en/about.
[40] See Mary Vardigan and Cole Whiteman (2007), "ICPSR meets OAIS: Applying the OAIS reference model to the social science archive context," *Archival Science* 7(1): 73–87.
[41] See also Pricilla Caplan (2007), "The Florida Digital Archive and DAITSS: A Working Preservation Repository Based on Format Migration," *International Journal on Digital Libraries* 6(4): 310.
[42] Simona Cohen, Dalit Naor, Leeat Ramati and Petra Reshef (2006), "Towards OAIS-Based Preservation Aware Storage: White Paper," IBM Haifa Labs (November 2006), 6 (emphasis in original). Available at http://www.haifa.ibm.com/projects/storage/datastores/papers/preservation_dataStores_white_paper-public.pdf.
[43] See http://preserv.eprints.org/.
[44] See http://www.portico.org/index.html.
[45] See http://www.ndiipp.uiuc.edu/.

"opinion polls, voting records, large-scale surveys, and many other social science studies,"[46] while the "Preservation and Long-term Access through Networked Services" (PLANETS) project is working to create tools and services to "ensure long-term access to [organizations'] valued digital content."[47] Where these projects have worked to create systems and repositories to hold digital content and digital objects, various other projects have focused on ways to help improve the functionality of these systems.

Recently, there has been an increase in projects devoted to the development and implementation of metadata for digital preservation. In 2005, "The Long-Term Preservation Metadata for Electronic Resources" (LMER) project released a metadata schema to assist with the long-term preservation of "electronic resources." [48] Two years later, the "Clever Recordkeeping Metadata Project" (CRKM) from Monash University concluded its work, noting that "the limitations of current recordkeeping standards and practices that are predominantly still operating within the paper paradigm inhibit realisation of the full potential of recordkeeping metadata as a business tool."[49] In 2004, the Hybrid Archives Project developed a data model that focuses on the use of metadata to allow "assets" to be shared between "an archive and the depositing institution, allowing the institution to provide sophisticated access to the resource, while the archive ensures the longer-term survival of the underlying content."[50] Finally, between 2007 and 2008 the "Service-Oriented Architecture for Preservation and Ingest of digital objects" (SOAPI) project intends to develop "an architecture and toolkit for (partially) automating preservation and ingest workflows in digital repositories."[51]

Although the original version of the OAIS model did not discuss the role of metadata, the above projects have contributed to this vital component of digital preservation. Furthermore, with the development of other metadata standards, such as the Metadata Encoding and Transmission Standard (METS),[52] it is likely that interest and research in this area will continue to grow, benefiting the long-term survival of digital records.

As organizations and archives continue to address the long-term preservation of digital records and information, cost and financial feasibility of implementation have also become prevailing concerns. Within the past few years, several projects have addressed this issue. Between 2005 and 2007, the "Effective Strategic model for the Preservation and Disposal of Institutional Digital Assets" (ESPIDA) project developed cost models for the preservation of digital "assets" to help potential funding organizations determine the short- and long-term value of proposed digitization projects.[53] Additionally, the University of London and the British Library created the "Life-Cycle Information for E-Literature" (LIFE) project. Currently in its second phase, the LIFE project seeks to refine the Life Generic Preservation Model it created during its first phase. This model examines the "life cycle of the management of digital materials" and provides a means by which to "calculate the costs of preserving digital information for the next 5, 10 or 100 years."[54]

[46] Data-PASS, "About the Project: Overview of the Project." Available at http://www.icpsr.umich.edu/DATAPASS/about.html.
[47] Planets, "About Planets: Objectives," 3 September 2007 version. Available at http://www.planets-project.eu/about/.
[48] Available at http://www.d-nb.de/eng/standards/lmer/lmer.htm.
[49] Monash University, "Clever Recordkeeping Metadata Project Final Report," 1. Available at http://www.infotech.monash.edu.au/research/groups/rcrg/crkm/docs/rpt-final.doc.
[50] Gareth Knight and Hamish James (2004), "Hybrid Archives Model: A Model for the Joint Curation of Digital Resources Guidebook," draft 2.0, Arts and Humanities Data Service, 4. Available at http://ahds.ac.uk/about/projects/hybrid-archives/hybrid-archives-model.pdf.
[51] SOAPI, "SOAPI: Description," 28 February 2007 version. Available at http://ahds.ac.uk/about/projects/soapi/index.htm.
[52] See http://www.loc.gov/standards/mets/.
[53] See http://www.gla.ac.uk/espida/index.shtml.
[54] LIFE, "Welcome to the LIFE Website." Available at http://www.life.ac.uk/.

Finally, the Image Spatial Data Analysis Group (iSDA), at the University of Illinois' National Center for Supercomputing Applications (NCSA), has developed a "simulation environment for understanding computational requirements of preservation and reconstruction."[55] Called "Image Provenance to Learn" (Ip2Learn), this freely-available software application, developed in collaboration with the United States National Archives and Records Administration (NARA), is designed to help users calculate "the cost of information preservation and the value of preserved information." Although Ip2Learn is designed primarily to work with geospatial digital records, especially images, it is likely that the general concept is readily adaptable to other types of digital records. Among other things, the Ip2Lean application allows users to experiment with simulated trade-off for a number of preservation variables, including variations in: information granularity, organization, representation, compression types/levels and authentication requirements; metadata gathering/harvesting methods; storage formats; retrieval requirements (e.g., local/centralized versus remote/distributed); and reconstruction processes (e.g., processes requiring human intervention versus fully automated processes). While cost models have not garnered nearly the attention that metadata or "data" preservation has, as archivists and records managers advocate the importance of digital preservation to organizations, one may only assume that additional projects will emerge and offer further insight and resources for this issue.

A final category of existing digital preservation projects involves those that act as "clearing houses" for digital preservation information generated by other research projects or preservation initiatives. Digital Preservation Europe (DPE) is a good example of this type of project. Building on the earlier successful work of ERPANET, DPE's objective is to foster "collaboration and synergies" by facilitating the "pooling of the complementary [digital preservation] expertise that exists across the academic research, cultural, public administration and industry sectors in Europe." As such, "DPE addresses the need to improve coordination, cooperation and consistency in current activities to secure effective preservation of digital materials."[56] The main interests and goals of its partners include:

a. raising the profile of digital preservation;
b. promoting the ability of Member States acting together to add value to digital preservation activities across Europe;
c. using cross-sectoral cooperation to avoid redundancy and duplication of effort;
d. ensuring auditable and certificated standards for digital preservation processes are selected and introduced;
e. facilitating skills development through training packages;
f. enabling relevant research coordination and exchange;
g. developing and promoting a research agenda roadmap; and
h. helping both citizens and specialist professionals recognise the central role that digital preservation plays in their lives and work[57]

[55] Image Spatial Data Analysis Group, "Data Preservation and Reconstruction: Understanding Computational Requirements of Preservation and Reconstruction," 4 December 2007 version. Available at http://isda.ncsa.uiuc.edu/CompTradeoffs/. See also Peter Bajcsy (2007), "Challenges in Preserving and Reconstructing Computer-Assisted Decision Processes," paper presented at the 2007 Society of American Archivists Annual Meeting, Electronic Records Section Meeting, 31 August 2007, Chicago, IL, USA. Available at http://www.ibiblio.org/saa2007/images/1/1f/DecisionProcesses_PeterBajcsy.pdf.
[56] Digital Preservation Europe, "About DPD: Mission Statement," 26 April 2006 version. Available at http://www.digitalpreservationeurope.eu/about/.
[57] Ibid.

The DELOS project, a "Network of Excellence on Digital Libraries" partially funded by the European Commission, is involved in similar activities. This four-year project, which began in January 2004 and is scheduled to conclude its work in December 2007,[58] is based on the vision that "all citizens, anywhere, anytime, should have access to Internet-connected digital devices to search all of human knowledge, regardless of barriers of time, place, culture or language." As such, "DELOS believes that, in the near future, networked virtual libraries will enable anyone from their home, school or office to access the knowledge contained in the digital collections created by traditional libraries, museums, archives, universities, governmental agencies, specialised organizations, and individuals around the world." In response to this vision, the main objective of DELOS is "to contribute to the development of the enabling technologies so that its vision for digital libraries may become reality." Given the multi-disciplinary dimension inherent in the design and implementation of digital libraries, the ultimate success of which, DELOS believes, "relies on the advancement of many enabling technologies," the main role of DELOS in this development process is to integrate the research activities that are being carried out separately under the various enabling technologies with research activities being conducted by several major European teams working within the digital library field. A key element of this activity involves "disseminating knowledge of digital library technologies to many diverse application domains, [as well as providing] specific user communities with access to advanced digital library technologies, services, testbeds, and the necessary expertise and knowledge to facilitate their take-up."[59]

## Conclusion

As this brief overview has shown, although there are a wide variety of projects devoted to the long-term preservation of digital *information*, to date, very few projects have focused on the long-term preservation of digital *records*. In fact, most projects have involved broad-scale and often ambiguously-defined approaches aimed at addressing the preservation of "digital data" and/or "digital content." It remains to be seen how the preservation systems and repositories envisioned by these projects will be affected by such broadly-defined preservation scopes and intentions. Despite the dearth of current interest in *digital records* per se, the third phase of InterPARES may be able to offer some insight to these other, more broadly-scoped projects and their products. One of the objectives of the InterPARES 3 Project is establish whether digital records preservation "theory and methods, concepts and principles apply across jurisdictions, regardless of legal/administrative, social and cultural environments; and, in...situation[s] where they do not apply, to identify why, and determine the measures that are required to ensure the preservation of digital records. In light of this objective, it is hoped that, as the Project's national and multinational TEAMs analyze and compare their findings related to specific case studies and scenarios among a wide-range of digital records users, it will be possible to determine whether preservation systems such as Data-PASS and PLANETS are adequate to meet an organization's long-term preservation needs or whether further refinement is required.

[58] In response to the impending conclusion of the DELOS project, a not-for-profit organization called the "DELOS Association" has been established "to continue as much as possible the DELOS activities." The Association's main objective "is to promote research activities in the field of digital libraries through the organization of workshops, working groups, schools, etc." (DELOS, "DELOS: An Association for Digital Libraries." Available at http://www.delos.info/index.php?option=com_content&task=view&id=614&Itemid=341).

[59] DELOS, "About DELOS." Available at http://www.delos.info/index.php?option=com_content&task=view&id=299&Itemid=26.

# Appendix 1

# Digital Preservation Projects: An Annotated Bibliography

## Preface

Serving as a companion to the report: "Digital Preservation Projects: A Report Discussing Past and Current Projects Devoted to the Long-Term Preservation of Digital Records and Digital Information," this annotated presents a snapshot, taken between late September and early November 2007, of numerous projects and organizations that have addressed or are addressing the long-term preservation of digital records or digital information. The initial goal of this bibliography was to identify only those projects that were similar in scope and intentions to InterPARES with regards to their focus or methodological approaches. Yet, as the research progressed, other prevailing trends emerged. By listing many of the major research projects and related organizations, this bibliography accounts for these trends. The bibliography does not intend to be all-encompassing. The projects and organizations listed only date back to 1999, the inaugural year of InterPARES.

The bibliography is arranged into two sections: Projects and Organizations. The first section, Projects, deals with the different methods and approaches addressed the long-term preservation of digital records and digital information. Besides the name, purpose and scope of the projects, the annotations, where applicable, also mention project deliverables, publications and reports. The section is arranged alphabetically by the English translation of the project's name. At the end of each abstract, there are one or two bold-faced words that represent the primary focus of the project (e.g., Digital Records, Metadata).

The second section, Organizations, accounts for the primary organizations devoted to the long-term preservation of digital records or digital information. The annotations provide each group's primary goal(s) and mentions any deliverables, reports, working group or Web site feature(s). Like the previous section, the organizations are arranged alphabetically by organizational name.

## Projects

*Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval (CASPAR)*
http://www.casparpreserves.eu/
Teaming with several European organizations and institutions, the CASPAR project has four primary objectives. First, it intends to build a "pioneering preservation environment" based on the OAIS reference model. Second, it wants to show that its system may handle the preservation of the digital resources. Third, the project seeks to "advance the current state of the art in digital preservation"; and finally, it will develop "technological solutions supporting the emergence of" other preservation systems and services. To date, CASPAR has created a conceptual model, a set of guidelines for understanding the inner-workings of the CASPAR project, and a "Dissemination and Use Plan." **Preservation System/Repository**

*Clever Recordkeeping Metadata Project (CRKM)*
http://www.infotech.monash.edu.au/research/groups/rcrg/crkm/
Between 2004 and 2006, the Records Continuum Research Group at Monash University in Australia sought to "develop a prototype to demonstrate how to overcome major barriers to the implementation of recordkeeping and resource discovery metadata standards, particularly in eGovernment." Additionally, it intended to provide an implementation model for the clever use of metadata in quality recordkeeping systems that capture and manage information that can support the reliability, authenticity, accessibility and usability of evidence of government decisions and activities for as long as that evidence is required." **Metadata**

*Co-operative Development of a Long-Term Digital Information Archives (KOPAL)*
http://kopal.langzeitarchivierung.de/index.php.en
Starting in July 2004, KOPAL operated for three years. The project was a joint-partnership between the German National Library, Goettingen State and University Library, Gesellschaft fuer wissenschaftliche Datenverarbeitung mbH Goettingen, and IBM Deutschland GmbH. These institutions focused on the long-term preservation of electronic publications, as well as on how to address obsolescent file formats, hardware and software, by creating its own archive for digital data—the Kopal Solution. This open-source software application is based on the OAIS model and is designed for handling and preserving digital objects. **Preservation System/Repository**

*Data Preservation Alliance for the Social Sciences (Data-PASS)*
http://www.icpsr.umich.edu/DATAPASS/
Supported by the United States Library of Congress, Data-PASS is a broad-based partnership led by the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan and other institutions throughout the United States. Established in 2004 and currently in operation, the project intends to acquire and preserve a wide range of data at-risk, including "opinion polls, voting records, large-scale surveys, and other social science studies." To date, the project has released a set of their criteria and guidelines for collecting and preserving data; these include matters relating to Articles of Collaboration, Content Selection, Appraisal, Acquisition, Metadata, Confidentiality, Data Security and a Guide to Handling Fragile Materials. **Preservation System/Repository**

*Digital Archiving in Flemish Institutions and Administrations (DAVID / eDAVID)*
http://www.expertisecentrumdavid.be/eng/edavid.php
Organized from 2000-2003, the project was funded by the Flemish government and dealt primarily with digital born records (e.g. Web sites, e-mail, databases, etc.). While its Web site continues to be updated, most of its publications and findings are in Flemish, though a few of its core papers may be found in English. These include: Boudrez, F. (2006). "Filing and archiving e-mail"; Boudrez, F. (2005). "Digital containers for shipment into the future"; and Boudrez, F. (2005). "Digital signatures and electronic records." In 2005 the project also released a handbook on digital archiving titled "Digital Archiving: The New Challenge? Legal and Archival Issues" **Digital Records (E-mail)**

*"Data archiving" project (2004)*
http://www.lfcs.inf.ed.ac.uk/research/database/Archiving.html
This project focused on the preservation of scientific data, particularly complex databases. Using XML, the project's investigators developed a system that could, among other characteristics, "hold many versions in a file that is little larger the most recent version of the database." In 2006, the team released its "Scientific Data Archiver," which exemplifies this feature. For further discussion on the methodology behind the tool, readers should consult the paper "Archiving Scientific Data" by Peter Buneman, Sanjeev Khanna, Keishi Tajima, and Wang-Chiew Tan.[60] **Scientific Data**

*Digital Repository Infrastructure Vision for European Research (DRIVER)*
http://www.driver-repository.eu/
Funded by the Sixth Framework Programme, Research Networking Testbeds, from 2006 to 2007, the goal of the project was "to build the testbed for a future knowledge infrastructure of the European Research Area." During its second phase, it sought to create the "successful interoperation of both data network and knowledge repositories as integral parts of the E-infrastructure for research and education in Europe." The project has produced a couple documents presenting their architectural and functional specifications. **Preservation Resources**

*Electronic Resource and Preservation Access Network (ERPANET)*
http://www.erpanet.org/
In 2004, this project conducted 60 different case studies to examine the "perception and awareness of risk associated with information loss," "how digital preservation affects the organization," "what actions have been taken to prevent data loss" and how these actions are monitored. It also sought, through the case study findings, to determine what mechanisms are being used to determine future needs. The case studies looked at organizations containing cultural and scientific digital objects. To access the findings, readers must, at no charge, register with ERPANET. Currently, ERPANET continues to raise awareness, foster research and encourage the sharing of preservation ideas, experiences and information among individuals, organizations and institutions. In addition to its own work, the project's Web site contains a number of valuable publications and resources regarding the long-term preservation of digital information. **Preservation Resources**

[60] Available at http://www.lfcs.inf.ed.ac.uk/research/database/publications/tods04.pdf.

*Exploring Collaborations to Harness Objects in a Digital Environment for Preservation (ECHO DEPository)*

http://www.ndiipp.uiuc.edu/

The University of Illinois, the Online Computer Library Center (OCLC), National Center for Supercomputing Applications (NCSA), Tufts University's Perseus Project, and numerous other university and state libraries in the United States, joined efforts between 2004 and 2007 to determine how to best identify, archive and preserve digital resources. By the end of the project, this alliance had produced the EchoDep Hub and Spoke, an open-source digital archiving system that packages digital objects using the Metadata Encoding Transmission Standard (METS) so the objects may be archived and accessed in different systems. This project also hosts a Digital Preservation Pathfinder on its Web site, which provides an annotated list of digital preservation resources. **Metadata**

*An Effective Strategic model for the Preservation and Disposal of Institutional Digital Assets (ESPIDA)*

http://www.gla.ac.uk/espida/index.shtml

Established by the University of Glasgow, this two year project operated from 2005 to 2007. The goal of the project was to help organizations that were considering digitization projects and their potential funders determine if the proposed projects had enough value to warrant their initiation. Several publications resulted from this project, foremost being the "espida Handbook: Expressing Project Costs and Benefits in a Systematic Way for Investment in Information and IT."[61] Additionally, the project also created "Outcome Scorecards" and a "Cost Template" document that users may download from the project's Web site. **Cost Analysis**

*Fedora and the Preservation of University Records Project*

http://dca.tufts.edu/features/nhprc/index.html

Between 2003 and 2005, Yale University and Tufts University collaborated to determine the "requirements for trustworthy recordkeeping systems and preservation activities, the ingest of records into a preservation system, and the maintenance of records in a preservation system." As with the SOAPI project, this project focused on the ingest component of the OAIS model and created an Ingest Prototype Tool. Additionally, in September 2006, it released the first version of its "Requirements for Trustworthy Recordkeeping Systems and the Preservation of Electronic Records in a University Setting."[62] The project may be most notable for Fedora—an open source software application designed to help organizations and institutions manage and deliver their digital content.[63] **Digital Records; Preservation System/Repository**

*Format Curation Service (FOCUS)*

http://www.umiacs.umd.edu/research/adapt/focus/index.html

FOCUS is one of a growing number of projects that has dealt with file formats. Initiated by the Institute for Advanced Computer Study at the University of Maryland in 2004, the project addresses the long-term preservation of digital information by "developing an efficient, scalable, and secure prototype format registry that captures all the essential features" of the Global Digital Format Registry and by providing a framework for this environment. FOCUS sought to alleviate

[61] Available at https://dspace.gla.ac.uk/bitstream/1905/691/1/espida_handbook_web.pdf.
[62] Available at http://repository01.lib.tufts.edu:8080/fedora/get/tufts:UA069.004.001.00005/bdef:TuftsPDF/getPDF.
[63] For more information, see http://www.fedora.info/.

the burden of technology obsolesce by providing its users with information about any given format so that it may possibly be accessed at a later date. While it is unclear if the registry continues to be updated, the project's Web site contains a working demo of the format registry. **Format Identification**

*Hybrid Archives Project*

http://ahds.ac.uk/about/projects/hybrid-archives/index.htm

The Hybrid Archives: Towards an Extensible Model for Depositing Institutional Assets (formerly Partial Deposit) project operated from 2003 to 2004 and was initiated by the Arts and Humanities Data Service (AHDS). Focusing on metadata, the project sought to improve upon the OAI harvesting methodology by increasing its searching capability as well as its ability for preserving "disclosed assets" for the long term. The project created a data model that allows for "the institution to provide sophisticated access to the resource, while the archive ensures the longer-term survival of the underlying content." **Metadata**

*Indiana University Electronic Records Project*

http://www.libraries.iub.edu/index.php?pageId=3313

While the Indiana University Electronic Records Project did not focus solely on the long-term preservation of digital records, its findings and conclusions provide important insight to the status of digital records in a couple environments, mainly the education and financial sectors. From 2000 to 2002 the project performed several case studies and discovered that information systems were ill-designed and not properly suitable as recordkeeping systems. Besides several White Papers, the project developed "A Methodology for Evaluating Existing Information Systems as Recordkeeping Systems," "Functional Requirements for Recordkeeping Systems," and a set of "Recordkeeping Metadata Specifications." **Digital Records**

*Life-Cycle Information for E-Literature (LIFE)*

http://www.life.ac.uk/

Initiated in 2005, the University College London (UCL) Library Services and the British Library are currently operating in the second phase of the project. Where the first phase created the Life Generic Preservation Model to help establish a "methodology to model the digital lifecycle and calculate the costs of preserving digital information for the next 5, 10 or 100 years," from March 2007 to April 2008, LIFE$^2$ will test this model by conducting a series of case studies. As of November 2007, the project has released a revision of its model for user feedback. **Cost Analysis**

*JSTOR/Harvard Object Validation Environment (JHOVE)*

http://hul.harvard.edu/jhove/

From 2004 to 2006, the JSTOR/Harvard Object Validation Environment, established by the partnership between JSTOR and Harvard University, was one of the first projects that sought to create a "tool for format-specific digital object identification, validation, and characterization." By the project's conclusion it had created java-based application usable on Unix, Windows or OS X platforms. The project's Web site offers a tutorial for installing and using JHOVE, its schema and various modules. As of 2006, the project had submitted a proposal for the implementation of JHOVE2 to improve on its previous work. **Format Identification; Practical Resources**

*Long-Term Preservation Metadata for Electronic Resources (LMER)*

http://www.d-nb.de/eng/standards/lmer/lmer.htm

In 2005, the German National Library sought to fill a gap in strategies devoted to the long-term preservation of electronic documents. LMER saw a void in the creation of technical metadata and, as a result, based on a model of the National Library of New Zealand, developed its own XML Schema. The schema is available on the project's Web site in both English and German versions. **Metadata**

*Managing the Digital University Desktop (MDUD)*

http://www.ils.unc.edu/digitaldesktop/

For three years, 2003-2006, the University of North Carolina at Chapel Hill explored and documented how "faculty, administrators, and staff use and manage files and records from electronic mail and other desktop applications." The project concluded that within any given office, there were a number of different behaviours, attitudes and practices that address and handle e-mail. To help combat these differing views and procedures, MDUD—consisting of an online tutorial accompanied by several PowerPoint presentations—focused on the management of digital records and e-mail. **E-mail**

*Network of Expertise in Long-Term Storage of Digital Resources (NESTOR)*

http://www.langzeitarchivierung.de/

Concentrating on the preservation of digital resources, various archives, libraries and museums in Germany have combined their efforts to ensure the long-term survival of electronic journals, reference databases and online publications. As of 2007, the project was in the second of its proposed three-year term. Due to its large scope and various audiences, a number of papers have resulted from the project, most of which are in German. **Electronic Publications**

*Portico*

http://www.portico.org/index.html

Portico is one of a growing number of projects exploring ways to preserve electronic publications. Funded by the Andrew Mellon Foundation, JSTOR and the United States Library of Congress, the project strives to ensure that scholars, students and researchers continue to have access to "scholarly literature published in electronic form." The Web site lists numerous articles by Portico contributors that discuss the project and the importance for saving this form of digital information. **Electronic Publications**

*Preserving Georgia's Historical Data*

http://sos.georgia.gov/archives/who_are_we/rims/digital_History/default.htm

By 2008, the Georgia State Archives (U.S.) will have fully implemented a digital archives system that has been three-years in the making. The goal of this system is to allow digital content to be preserved and managed by the Archives but also be accessible by the public. Using a combination of migration, emulation and other "analog preservation alternatives (namely microfilm)," the project also relies on the use of established best practices and partnerships to ensure the preservation of the digital information. The project's Web site contains a white paper providing some more details about the Digital Archives of Georgia, as well as some "case studies" that discuss the intended procedures for implementation. **Preservation System/ Repository**

*Preserving the Records of E-Legislature*

http://www.mnhs.org/preserve/records/elegislature/elegislature.htm

For two years, 2005-2007, the Minnesota Historical Society (MHS) (U.S.), the Office of the Revisor of Statutes, and the Legislative Reference Library orchestrated this project to find a means for preserving the electronic records of the Minnesota legislature. While the project's final report will not be available until early 2008, the project also sought to deliver a preservation plan and set of "Best Practices;" presumably, these will also be made available in 2008. Readers may also be interested in several earlier publications produced by the MHS.

In July 2002, the Minnesota Historical Society released version 4 of its "Trustworthy Information Systems Handbook."[64] Created by the Trustworthy Information Systems (TIS) Project, this handbook establishes a set of criteria and offers a methodology for evaluating government information systems for trustworthiness. More recently, in March 2004, the MHS released version 4 of its "Electronic Records Management Guidelines."[65] These guidelines "provide information on a variety of topics, such as file formats and file naming, electronic records management strategies, electronic document management systems, digital media, storage facilities and procedures, e-mail and web content management, and electronic and digital signatures." The second section of the Guidelines deals primarily with the long-term preservation of digital records. **Digital Records**

*Preservation and Long-term Access through Networked Services (PLANETS)*

http://www.planets-project.eu/

During the next several years, PLANETS will be constructing "practical services and tools to help ensure long-term access to our digital cultural and scientific assets." Started in June 2006, the project has already released an Interoperability Framework[66] and is currently developing software that uses emulation and migration strategies for digital preservation. The project consists of various European national libraries and archives, leading research universities and technology companies and is funded by the European Union under the Sixth Framework Programme. **Preservation Resources**

*preservation Eprint SERVices (PRESERV)*

http://preserv.eprints.org/

PRSERVE is another project devoted to the preservation of electronic publications. Established and operated by the Southampton University (School of Electronics and Computer Science), The National Archives, The British Library, Oxford University and the Intelligence, Agents, Multimedia (IAM) Group and funded by JISC, its initial phase operated from 2005 to 2007. The project designed PRONOM-ROAR, a Web-based preservation service that identifies file formats for over two hundred repositories.[67] From 2007 to 2008, PRESERV 2 will continue to investigate and improve the means by which PRONOM-ROAR identifies and preserves digital information within the different repositories. **Preservation System/Repository; Electronic Publications**

[64] Available at http://www.mnhs.org/preserve/records/tis/tableofcontents.html.
[65] Available at http://www.mnhs.org/preserve/records/electronicrecords/erguidelinestoc.html.
[66] Available at http://www.planets-project.eu/software/Planets_IF2345-D3_ReleaseReport-Final_Website.pdf.
[67] For more information, see http://trac.eprints.org/projects/iar/wiki/Profile.

*Repository for Preservation Authentic Digital Records (RODA)*

http://roda.iantt.pt/en/

Organized and operated by the National Archive Institute of Portugal (IAN/TT), this project seeks to "provide technical solutions to digital preservation at the national level." For the next several years, RODA will develop tools for incorporating digital records and digital objects of Portugal's public organizations into the National Archive, as well as tools for managing and preserving these digital records and objects. Due to the project's recent founding, no documentation is available on its Web site. **Practical Resources**

*Securing a Hybrid Environment for Research Preservation and Access (SHERPA)*

http://www.sherpa.ac.uk/index.html

From 2005 to 2006, SHERPA designed an open-access institutional repository to assist organizations in the preservation of scholarly publications. Following the conclusion of this project, the Arts and Humanities Data Service (AHDS) implemented two other projects to further develop this work. SHERPA DP (Securing a Hybrid Environment for Research Preservation and Access Development Partner), 2005-2007, created a "shared preservation environment" based on the OAIS model. In 2005, SHERPA DP produced a set of requirements for a Disaggregated Service and an OAIS Complaint Model for Disaggregated Services. Currently, SHERPA DP2 will expand on this research. One of the objectives for SHERPA DP2 will be the creation of a Handbook regarding the preservation of electronic prints and publications. **Preservation System/Repository; Metadata**

*Service-Oriented Architecture for Preservation and Ingest of digital objects (SOAPI)*

http://ahds.ac.uk/about/projects/soapi/index.htm

SOAPI is another project concentrating its efforts on the ingest component of digital repositories. Between 2007 and 2008, the Arts and Humanities Data Service (AHDS) will "develop an architecture and toolkit" to help automate the preservation and ingest processes in a digital repository. To date, the project has not produced any deliverables. **Preservation System/ Repository**

*Significant Properties (InSPECT)*

http://www.significantproperties.org.uk/

From April 2007 to September 2008, the Arts and Humanities Data Service (AHDS), will examine the concept of "significant properties." Defined as "the characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects," the project wants to determine which elements of digital objects are most vital to ensure their long-term preservation. To accomplish this, InSPECT will look at different types of electronic information, such as raster images, e-mails, structured text and digital audio. The project is building off the concept of "essence" developed by the National Archives of Australia. At the time of this bibliography, none of the project's findings had been published. **Digital Object Identification (theoretical)**

*Victorian Electronic Records Strategy (VERS)*

http://www.prov.vic.gov.au/vers/vers/default.htm

VERS was one of the first projects devoted to the preservation of digital records. Developed by the Public Record Office Victoria (PROV) in the late 1990s, it sought to provide leadership and direction in the management of digital records. In 1998, it produced its "Final Report," which

concluded that, to preserve a digital record, metadata must be associated with the record, ideally using XML that allows for "layers" of metadata to be created and attached to the record. The report also stressed the importance of creating archival systems that would incorporate, manage and recover the records. **Digital Records**

## Organizations

*Committee on Data for Science and Technology (CODATA)*

http://www.codata.org/

CODATA strives to "improve the quality, reliability, management and accessibility of data of importance to all fields of science and technology." Particularly, the organization serves as a resource for scientists and engineers so that they may access and share information regarding "data activities." Within CODATA, there was the Task Group on Preservation of and Access to Scientific and Technical Data in Developing Countries,[68] which last met in 2005. While the task force produced no definitive projects, it did hold a workshop in Lisbon in 2003, which sought to "provide an international forum to exchange information about data archiving policies and practices across different scientific, institutional, and national contexts." For a summary of the workshop see "Selection, Appraisal, and Retention of Digital Scientific Data: Highlights of an ERPANET/CODATA Workshop."[69]

*Digital Curation Centre (DCC)*

http://www.dcc.ac.uk/

The purpose of the DCC "is to provide a national focus for research and development into curation issues and to promote expertise and good practice, both national and international, for the management of all research outputs in digital format." The organization meets annually and holds periodic workshops. Its Web site contains a variety of information including a listing of publications and presentations as well as current standards and tools. The DCC is also a co-contributor to the Digital Repository Audit Method Based on Risk Assessment (DRAMBA)[70] and is currently developing its "Digital Curation Manual."[71] This manual "is designed to assist data creators, curators and re-users to better understand the challenges they face and the roles they play in creating, managing and preserving their digital information over time." It is to be updated and added to on a regular basis, as articles continue to be written. Currently, there are over forty-five "installments" in the making with eight chapters currently available. Chapters include: E-Mail, Learning Object Metadata, Appraisal and Selection, Preservation Metadata, Investment in an Intangible Asset, Archival Metadata, and Open Source for Digital Curation.

*DELOS*

http://www.delos.info/

DELOS, an initiative funded by the European Commission's Information Society Technologies 5th Framework Programme (IST-FP5), seeks to promote digital library research and development in Europe. The initiative launched its Preservation Cluster (WP6) focusing on digital preservation. Within this initiative was the "Preservation Cluster (WP6)," which dealt with digital preservation.

[68] For more information, see http://www.tgdc-codata.org.cn/.
[69] Available at http://www.jstage.jst.go.jp/article/dsj/3/0/227/_pdf.
[70] For more information, see http://www.repositoryaudit.eu/.
[71] Available at http://www.dcc.ac.uk/resource/curation-manual/.

This cluster seeks to "eliminate the duplication of effort between research activities… examine core issues that will deliver essential guidelines, methods, and tools to enable the construction of preservation functionality within digital library activities… establish[ment] of testbeds and validation metrics… [and] to relate the digital preservation research agenda more directly to the development of exploitable product opportunities and to develop links with the industrial sectors." The Preservation Cluster has released several reports and papers, which include: "A Framework for Documenting the Behaviour and Functionality of Digital Objects and Preservation Strategies" (2006),[72] "DELOS Digital Preservation Testbed for Testing and Evaluating Digital Preservation Solutions" (2006)[73] and "Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation" (2003).[74]

*Digital Preservation Coalition (DPC)*

http://www.dpconline.org/

Established in 2001, the Digital Preservation Coalition sought "to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base." In 2002, the organization released its "Preservation Management of Digital Materials: A Handbook," initially compiled by Neil Beagrie and Maggie Jones. According to the DPC, the handbook "provides an internationally authoritative and practical guide to the subject of managing digital resources over time and the issues in sustaining access to them." While the book is no longer in print, it is freely available online and continues to be updated by the DPC.[75] Readers may also want to see Neil Beagrie's 2005 article "Digital Preservation: Best Practice and its Dissemination,"[76] where he discusses the Handbook in more depth. Additionally, this handbook contains a "Decision Tree for the Selection of Materials for Long-Term Retention." Modified and updated in March 2006 by Deborah Woodyard-Robinson, this section is an interactive feature that may be used to construct or test an organization's policy for the long-term preservation of digital materials.

*Digital Preservation Europe (DPE)*

http://www.digitalpreservationeurope.eu/

Digital Preservation Europe seeks to build on the work and successes of ERPANET. Like ERPANET, DPE aims to foster "collaboration and synergies between many existing national initiatives across the European Research Area" and to address "the need to improve coordination, cooperation and consistency in current activities to secure effective preservation of digital materials." The DPE and its members have produced numerous reports and papers. Most recently it released "Report on the Legal Framework on Repository Infrastructure Impacting on Cooperation Across Member States" (November 2007),[77] which focuses on copyright issues, and a draft of its "Digital Preservation Research Roadmap,"[78] which discusses the research agendas of numerous digital preservation projects. Finally, the DPE Web site contains a useful "Resource" section listing numerous other international organizations and digital preservation projects.

[72] Available at http://www.dpc.delos.info/private/output/DELOS_WP6_d641_final__vienna.pdf.
[73] Available at http://www.delos.info/index.php?option=com_browse_odl&collid=14604808712280030000045.
[74] Available at http://eprints.erpanet.org/archive/00000048/01/Digitalarchiving.pdf.
[75] Available at http://www.dpconline.org/graphics/handbook/.
[76] Available at http://www.ariadne.ac.uk/issue43/beagrie/.
[77] Available at http://www.digitalpreservationeurope.eu/publications/reports/PO3_4_LegalFramework.pdf.
[78] Available at http://www.digitalpreservationeurope.eu/publications/reports/dpe_research_roadmap_D72.pdf.

*Library of Congress (United States)*

http://www.digitalpreservation.gov/

While the Library of Congress's Web site lacks the quantity of research papers and reports that may be found on other organizational sites, it offers a variety of resources that readers may want to examine. The Library of Congress has teamed with numerous organizations and institutions and provides funding for digital preservation projects; this site lists these partnerships and the funded projects. Additionally, the Library of Congress established the Section 108 Working Group, which continues to examine copyright issues.[79] Finally, there is a small section called "Preserving Your Digital Memories" that provides some basic advice and recommendations for the preservation of e-mails, computer files, storage disks and digital photographs.[80]

*Ministerial Network for Valorising Activities in digitization, eContentplus (MINERVA EC)*

http://www.minervaeurope.org/

Continuing the work of MINERVA (2002-2006), MINERVA EC forms the core of an international network of institutions and individuals designed to "facilitate the creation of added value products and services at European level, to improve awareness of the state-of-the-art in the sector, to contribute to the overcome of fragmentation and duplication of digitisation activities of cultural and scientific content and to maximise cooperation among the Member States." While MINERVA EC's focus is not necessarily on the long-term preservation of digital information, its Work Package 6 will explore means to establish standards and best practices for saving cultural content.[81] The organization has created a "Best Practice Handbook" for the implementation of digitization projects.[82] Finally, the MINERVA Web site contains a large list of different "Digitization Guidelines" that readers may want to consider when creating a digitization project.[83]

*Preserving Access to Digital Information (PADI)*

http://www.nla.gov.au/padi/

The National Library of Australia has created a Web site for gathering and organizing a large assortment of information regarding digital preservation. The resources are divided by resource type (e.g., Events, Projects, Strategies, etc.) and then sub-divided by type of material (e.g., Articles, Reports, Books, etc.). The site attempts to track current trends, publications and projects, but many of the sections have not been updated since 1999 or 2000, while only a few have had new material added as recently as 2005.

[79] See http://www.section108.gov/.
[80] See http://www.digitalpreservation.gov/you/digitalmemories.html.
[81] For more information, see http://www.minervaeurope.org/structure/wg/bp.htm.
[82] Available at http://www.minervaeurope.org/structure/workinggroups/goodpract/document/bestpracticehandbookv1.1.pdf.
[83] Available at http://www.minervaeurope.org/guidelines.htm.