Papers from Portico

Archiving Electronic Journals: An Overview of Portico's Approach

August 2006, Number 1

Prepared by Amy Kirchhoff & Eileen Gifford Fenton

Portico's mission is to preserve scholarly literature published in electronic form and to ensure that these materials remain accessible to future scholars, researchers, and students.

Elements of this paper first appeared in articles by Eileen Fenton published in *Ariadne Web Magazine*, Issue 47, April 2006 http://www.ariadne.ac.uk/issue47/ and in *Serials Review*, Volume 32, Issue 2, June 2006 doi:10.1016/j.serrev.2006.03.004.

Growing Digital Preservation Needs

According to the most recent (2006) Pew Internet & American Life Project Tracking Survey, "fully 73% of respondents … are internet users, up from 66%" in January 2005[1] and 47% in June 2000.[2] As users have increased so too has available content. In 2000, Lyman and Varian estimated the surface Web at approximately 50 terabytes[3] and by 2003 that estimate had increased to 170 terabytes.[4] Active websites increased from nearly 8.2 million in May 2000 to almost 41 million in May 2006.[5]

The scholarly community's use and creation of digital content has mirrored this broader trend as illustrated by a variety of data points. For example, in September 2001, the Association of Research Libraries' (ARL) *Directory of Electronic Journals, Newsletters and Academic Discussion Lists* catalogued 3,887 e-journals and newsletters.[6] By May 2006 the number of catalogued e-journals and newsletters had increased to 17,392.[7] Usage of scholarly e-content has also increased tremendously as can be seen in variety of measures. For instance, usage of JSTOR, an on-line archive of the backfiles of scholarly journals, increased from 3.9 million meaningful accesses in April 2000 to nearly 40 million meaningful accesses in April 2006.[8] The percentage of academic library budgets devoted to electronic resources is also growing and as of the 2003–04 academic year, ARL libraries were spending an average 31% of total library material expenditures on electronic resources (on average $2,718,015 per library per year).[9] Given other trends, one must conclude that the percentage is even higher today and likely continuing to grow.

---

[1] Madden, "Internet Penetration and Impact." (2006), p 3. http://www.pewinternet.org/PPF/r/182/report_display.asp

[2] Rainie et al., "More Online, Doing More." (2001), p 2. http://www.pewinternet.org/PPF/r/30/report_display.asp

[3] Lyman and Varian, "How Much Information 2000." (2000). http://www2.sims.berkeley.edu/research/projects/how-much-info/index.html

[4] Lyman and Varian, "How Much Information 2003." (2003). http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/

[5] "May 2006 Web Server Survey." (2006). http://news.netcraft.com/archives/2006/06/04/june_2006_web_server_survey.html

[6] "ARL Directory of Electronic Journals, Newsletters and Academic Discussion Lists, 7th Edition," (ARL, 2001) https://db.arl.org/edir/

[7] *New Jour Archive* [Email List Archive] (University of California at San Diego Libraries, 2006) http://gort.ucsd.edu/newjour/

[8] "JSTOR Usage Statistics," (JSTOR) http://stats.jstor.org/. A "meaningful access" is when a user accesses an intellectual item, such as viewing a volume or issue listing, viewing an issue table of contents, viewing a page, performing a search, or printing an article.

[9] Young and Kyrillidou, *ARL Statistics 2003-2004* (Washington: Association of Research Libraries, 2005).

Over the past decade, faculty perceive themselves as having become more reliant on electronic resources. A 1995 survey of faculty in chemistry, physics, sociology, psychology, English and history concluded that the respondents were beginning to use networked resources but had a "lack of trust in" e-journals.[10] A 1999-2001 survey of faculty and students in the fields of international affairs, environmental science, and political science by the Electronic Publishing Initiative at Columbia (EPIC) found that "40% [of respondents] somewhat or strongly agree that they would rather settle for what they can find online, even if it is not quite what they wanted, in order to save making the trip to the library. For 20.5% of … [the] sample, the use of electronic resources has become so pervasive that they admit to rarely even looking beyond electronic resources for information".[11] By 2000, when JSTOR conducted a survey of over 4,000 faculty in the social sciences and humanities, more than 60% of the faculty who responded to the survey considered electronic databases to be invaluable.[12] A 2003 follow-up survey of over 7,000 faculty in the social sciences and humanities and found that "use of electronic resources, and dependence on them, has increased since 2000."[13]

The growing reliance on e-resources has contributed to an increasing concern about the fragility of electronic content. The first full-text search engine, Web Crawler, debuted in 1994 and indexed approximately 72,000 pages. None of the top 25 pages listed at that time exist today.[14] In October 2003, *Science* published the results of a survey which found that three months after publication, 3.8% of the internet references in three high-impact medical journals were inactive; at 15 months this percentage had risen to 10%; and it rose again to 13% at 27 months after publication.[15] Similarly, a 2005 study published in *The Serials Librarian* showed

---

[10] Budd and Connaway, "University Faculty and Networked Information: Results of a Survey," *Journal of the American Society for Information Science* 48, no. 9 (1997), p 850. http://www3.interscience.wiley.com/cgi-bin/abstract/39759/ABSTRACT

[11] *EPIC Faculty Survey* [Power Point Presentation] (Electronic Publishing Initiative at Columbia, 2003), p 17. http://www.epic.columbia.edu/eval/FacSurv.903.ppt

[12] Guthrie, "What Do Faculty Think of Electronic Resources" (JSTOR ALA Annual Conference Participants' Meeting June 17, 2001), p 17. http://www.jstor.org/about/faculty.survey.ppt

[13] Guthrie and Schonfeld, "What do Faculty Think of Electronic Resources? Findings from the 2003 Academic Research Resources Study" (CNI Task Force Meeting April 16, 2004), p 16. http://www.cni.org/tfms/2004a.spring/presentations/CNI_Guthrie_What.ppt

[14] Running, *The Top 25 Web Sites of 1994* [Weblog] (2006) http://www.downloadsquad.com/2006/04/14/the-top-25-web-sites-of-1994/

[15] Dellavalle et al., "Going, Going, Gone: Lost Internet References," *Science* 302, no. 5646 (2003) http://www.sciencemag.org/cgi/content/full/302/5646/787

that three years after publication, half of the internet references in three top communications journals did not link to active content.[16]

The concerns about the fragility of electronic resources, especially scholarly e-journals, are compounded by "the nature of the licensing regime under which these journals are now distributed. When research and academic libraries license e-journals, they do not take local possession of a copy as they did with print. Rather they use content stored on remote systems controlled by publishers… For electronic journals, the academy has as yet no functional equivalent in long-term maintenance and control over the scholarly record that 'owning a copy' provided for printed journals."[17] This statement from *Urgent Action Needed to Preserve Scholarly Electronic Journals* is echoed in the 2006 European Commission report, *Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe*, which stated that "the electronic era has brought a major paradigmatic change in the provision of access to back issues of journals: in the print era, libraries were acquiring print journals and took in charge their preservation so that they remain accessible to their user community in the long term. In the digital era, libraries and their user community are licensed online access to e-journals for a determined and limited duration."[18]

Emerging Digital Preservation Efforts

As awareness of the vulnerability of electronic scholarly resources has grown, so too have the efforts to conceptualize and develop a preservation infrastructure. From 1994-96 the Task Force on Archiving of Digital Information investigated the elements necessary to ensure "continued access indefinitely into the future of records stored in digital electronic form."[19] In 1996 the task force published its final report, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information,* outlining the technical and institutional challenges inherent in archiving digital information and the roles and responsibilities of archival organizations. In January 2002, the Consultative Committee of Space Data Systems (CCSDS) published the *Reference Model for an Open Archival Information System (OAIS),*[20] which outlines a

[16] Bugeja and Dimitrova, "The Half-Life Phenomenon: Eroding Citations in Journals," *The Serials Librarian* 49, no. 3 (2005), p 117. http://www.haworthpress.com/store/E-Text/View_EText.asp?sa=3&s=J123&v=49&i=3&fn=J123v49n03%5F10

[17] "Urgent Action Needed to Preserve Scholarly Electronic Journals." (2005). http://www.diglib.org/pubs/waters051015.htm

[18] Dewatripont et al., "Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe." (2006), p 75. http://europa.eu.int/comm/research/science-society/pdf/scientific-publication-study_en.pdf

[19] Garrett and Waters, "Preserving Digital Information: Report of the Task Force on Archiving of Digital Information." (1996), p iii. http://www.rlg.org/ArchTF/

[20] "Reference Model for an Open Archival Information System - OAIS." (2002). http://public.ccsds.org/publications/archive/650x0b1.pdf

technological and organizational model for long term preservation of digital objects. (In 2003 this reference model became an ISO standard, ISO 14721:2003.) Building on the OAIS work and *Preserving Digital Information*, the Digital Archive Attributes Working Group was convened in 2000 and was "charged with 'defining the characteristics of reliable archiving services for heterogeneous research collections.'" In 2002 the working group published *Trusted Digital Repositories: Attributes and Responsibilities*.[21] This report was extended by an RLG and National Archives and Records Administration (NARA) Task Force to develop an audit instrument, *An Audit Checklist for the Certification of Trusted Digital Libraries,* to certify digital archives.[22] As of this writing, the Center for Research Libraries (CRL) is using this tool to create certification processes for digital repositories.[23]

Over the past decade national libraries and other groups have established programs to both implement digital archives and encourage the development of archival practices for digital repositories, including:

- The Preserving Access to Digital Information (PADI) initiative of the National Library of Australia.
- The e-Depot and digital preservation research of the Koninklijke Bibliotheek (KB), the national library of the Netherlands.
- The Digital Preservation Coalition (DPC) in the UK.
- The National Digital Information Infrastructure and Preservation Program (NDIIPP) of the US Library of Congress.

A Brief History of Portico

In response to the growing need for solutions focused on long-term preservation of scholarly journals, Portico was launched in 2005. Portico's mission is to preserve scholarly literature published in electronic form and to ensure that these materials remain accessible to future scholars, researchers and students. Portico began as the Electronic-Archiving Initiative, a project launched by JSTOR in 2002 with a grant from The Andrew W. Mellon Foundation. The Initiative was built upon the Mellon Foundation's 1999 E-journal Archiving Program which funded a number of programs to investigate the technological infrastructure requirements and economic

[21] "Trusted Digital Repositories: Attributes and Responsibilities." (2002). http://www.rlg.org/longterm/repositories.pdf

[22] "Audit Checklist for the Certification of Trusted Digital Repositories: Draft for Public Comment." (2005). http://www.rlg.org/en/page.php?Page_ID=20769

[23] *CRL Auditing and Certification of Digital Archives* [Website] (Center for Research Libraries, 2005) http://www.crl.edu/content.asp?l1=13&l2=58&l3=142

model options for sustaining an e-journal archive.[24] The initial focus of the Initiative was designing and prototyping content handling and archival systems, crafting potential archive service models, testing these possible service models with libraries and publishers, and drafting a business model capable of supporting a long-term archival effort. For more than two years, project staff engaged in extensive discussions with publishers and libraries to develop a sustainable business model and technological approach that would balance the needs of all interested communities.

During 2004 the project was transitioned to Ithaka, a not-for-profit organization aiming to help new initiatives become established and sustainable as part of its broader mission to accelerate the productive uses of information technologies for the benefit of higher education around the world. Once established at Ithaka, Portico initiated a series of conversations with a wide-ranging, informal network of librarians from more than fifty academic institutions of all types and sizes. Portico also formally engaged ten publishers that agreed to participate in the discovery phase of the project including small scholarly societies, a university press, and large commercial publishers.[25]

Our early work was informed by several key assumptions regarding characteristics critical to a long-term archival effort. These assumptions were drawn from a variety of sources including the Mellon Foundation's E-journal Archiving Program, JSTOR's operational experience serving as a third-party archive since 1995, and the May 2002 RLG/OCLC report, *Trusted Digital Repositories: Attributes and Responsibilities*.[26] Building from this collective base of experience, we assumed that a reliable, third-party, long-term archive would require at least five basic elements:

1) an institutional mission with preservation at its core,
2) an economic model capable of sustaining the archival effort,
3) a robust and evolving technological infrastructure sufficient to meet the complexities of electronic resources,
4) relationships with libraries, responsible for the preservation of their collections, and
5) relationships with publishers, creators of the content that is to be preserved.

---

[24] *Archiving Electronic Journals* [Website] (Digital Library Federation, 2003) http://www.diglib.org/preserve/ejp.htm

[25] Publishers participating in the Portico pilot phase included: the American Economic Association, American Mathematical Society, American Political Science Association, Association of Computing Machinery, Blackwell, Ecological Society of America, National Academy of Sciences (PNAS), Royal Society, University of Chicago Press, and John Wiley & Sons.

[26] "Trusted Digital Repositories: Attributes and Responsibilities."

During the early months of the Portico project we tested these assumptions working with the ten publishers who were part of our discovery phase and a wide base of libraries. The publishers submitted a broad selection of sample e-journal data that enabled us to assess the technological challenges facing an e-journal archive. They also shared their own perspective on the e-journal archiving challenge and the concerns and needs which they faced as content publishers. Our publisher conversations were mirrored by those we conducted in the library community. Librarians shared their preservation concerns, expectations of an e-journal archive, and thoughts about various economic model possibilities.

This community engagement and our work to build a prototype repository and define a sustainable business model for a third-party archiving service affirmed our beginning assumptions but also yielded several important lessons.

- First, the preservation of e-journals presents many significant technological challenges. E-journals vary in format and data quality across the industry and often within the practice of individual publishers and even individual titles. Any long-term preservation infrastructure, including software, hardware and staff with appropriate expertise, must respond to the challenges of this diversity.
- Second, the sheer scale of the digital material requiring preservation dictates that libraries and publishers build multiple mechanisms for cooperative effort. The cooperation of the scholarly publishing community is essential because in many cases libraries lease rather than own e-journals. Securing this cooperation is complicated by the fact that preservation is not a core activity for most publishers.
- Third, access to archived literature is a key issue for both publishers and libraries, but they view this issue from very different perspectives. Publishers are eager to ensure that access to archived literature does not reduce the value of their current product offerings. Libraries are focused on assuring that access to core literature for their campuses will be reliable and timely.
- Finally, all the parties we engaged recognized the importance of building a robust economic model capable of sustaining a long term archive.

Building from these assumptions and findings, the Portico electronic archiving service was launched in 2005.

The Portico Model

The Portico archiving service is open to a scholarly publisher's complete list of journals, including those titles which may be published in electronic format only, those published jointly in print and electronic formats and those digitized from an original print publication. Portico's

archival approach for e-journals is that of managed preservation focused on the long-term, iterative migration of publishers' e-journal source files (as discussed in greater detail below).

Financial Support

Diversified financial support is critical to Portico's long-term preservation effort. In general, two kinds of support are needed: (1) funds for initial development of technological infrastructure and early operations and (2) ongoing funds that can support the operation of the archive over the life of the preserved materials. Portico has secured grant support from JSTOR, the Library of Congress, and The Andrew W. Mellon Foundation to cover the costs of initial development with additional support from Ithaka. As the Portico archive grows, the organization will cover its operating costs from diversified sources, in order to avoid dependence upon any single revenue stream. The chief beneficiaries of the archive, publishers and academic institutions, will provide the primary sources of funding. Charitable foundations and government agencies are also expected to provide periodic grant support.

Publishers are asked to make an annual contribution to the archive to support the ongoing costs to receive, normalize, store, and migrate article source files, The contributions are based on publishers' total journals revenues (including subscription, advertising, and licensing) and range from $250 to $75,000 per year. Library participants are also asked to make an annual payment to support the ongoing work of the archive. Library Annual Archive Support payments are tiered from $1,500 to $24,000 per year and vary according to a library's total library materials expenditure, reflecting Portico's value in preserving the growing portion of library collections that is digital content. Details of the current library and publisher contribution levels are available on the Portico website www.portico.org.

Access to the Archive

Campus-wide access to archived content will be granted to libraries supporting Portico when specific conditions are met that cause titles to no longer be available from the publisher **or** any other source. These specific conditions are often called trigger events and include:

- A publisher stops operations, or
- A publisher ceases to publish a title, or
- A publisher no longer offers back issues, or
- Upon catastrophic and sustained failure of a publisher's delivery platform.

Portico also provides a reliable means to secure perpetual access, if participating publishers choose to designate Portico as a provider of post-cancellation access.

In addition to access being triggered by the above events, up to four librarians at each participating institution will be granted password-controlled access to the archive for audit and verification purposes only. Publishers are also granted password-controlled access to their own content in the archive.

Archival Approach

In addition to developing an access and economic model, Portico has developed a set of guiding principles which shape our approach to preservation of e-content. These principles have emerged from the scholarly community's current understanding of digital preservation issues and our iterative discussions with the library and publishing communities.

- The integrity of the scholarly record must be preserved.

  The Portico archive accepts content in the format in which it was originally published. Original and migrated versions of component source files are preserved, and once deposited, archived content will be preserved in the archive permanently.

- Portico's focus is archiving scholarly content.

  Portico's role is to preserve the intellectual content of scholarly resources beginning with electronic scholarly journals. We are not focused on the preservation of publisher business systems, related content production data or delivery platforms.

- Source files can reliably capture the intellectual content of electronic scholarly journals.

  Portico preserves scholarly journals by capturing and migrating the source files which comprise the e-journals. These files include the master copy of the article and the components used in the web and print renditions of the journal. Source files may include PDF files used in the print and online display; SGML or XML full text files; SGML or XML header text files of bibliographic metadata; multiple images in different resolutions for each figure in the article; images for the equations or tables in the article; and other supplemental data such as video, audio, and datasets.

  We recognize that some information may not be captured in source files. This typically includes material that exists only in the publisher's business systems or is generated through the publisher's delivery infrastructure such as covers and other front matter, print or on-line advertisements, and print or on-line table of contents. Supplemental material which exists outside the boundaries of the scholarly journal

and is not maintained by the publisher may also not be available for source file preservation.

- Preservation of e-journal content can be achieved through migration.

  Portico has opted for a strategy of migration by transitioning component source files from one file format to another as technology changes and file formats become obsolete. Portico supplements and supports the migration policy with byte preservation by archiving the original source files along with any migrated versions.

  Our preservation concern and attention is directed at understanding the source file components that create the articles and not on the technology for delivery of that content. Portico focuses on file format management. We monitor the information technology and archival communities to understand the community acceptance of the specific formats of files within the Portico archive. As necessary, we will migrate files in the archive to new formats in order to keep the contents of the archive current with technological advances.

  Portico's first preservation action is to normalize the publishers' SGML or XML header or full-text files to the Journal Archiving and Interchange DTD[27] created by the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM). This normalization process limits the number of formats we must monitor and also allows us to develop a uniform HTML rendition of all the content in the archive. The original SGML or XML files and the NLM DTD based XML files are all preserved in the archive. Following the OAIS model, the archive also includes all DTDs, schemas and accompanying documentation necessary to understand the archived content.[28]

- Reliance upon accepted standards and participation in community project development enhances archival reliability. Below are some of the standards and projects that have influenced Portico's archival practice:

[27] *Journal Archiving and Interchange DTD* [Website] (National Center for Biotechnology Information of the National Library of Medicine, 2004) http://dtd.nlm.nih.gov/

[28] "Reference Model for an Open Archival Information System - OAIS."

- Digital Item Declaration Language (DIDL) of MPEG-21: DIDL is the packaging format language of MPEG-21, a developing multimedia framework.[29] The Portico content model is based upon the DIDL packaging concepts.
- Metadata Encoding and Transmission Standard (METS): METS is an XML schema for encoding descriptive, administrative, and structural metadata regarding objects within a digital library.[30] Portico uses a modified version of METS to represent the articles in the archive and to track the preservation actions taken on the source files of each article.
- Journal Archiving and Interchange DTD created by the National Library of Medicine (NLM DTD): The NLM DTD provides a common format in which publishers and archives can exchange and preserve journal content. The intent of the DTD suite is "to preserve the intellectual content of journals independent of the form it which" it was originally delivered.[31] Every publisher full-text or header file provided to Portico is transformed into an XML file based on the NLM DTD.
- Reference Model for an Open Archival Information System (OAIS): OAIS is a recommendation on the requirements that should inform an archive which provides permanent preservation of digital information.[32] The Portico archive was designed to be OAIS compliant.
- Global Digital Format Registry (GDFR): GDFR is a project to create a distributed format registry to store, discover, and deliver information about digital formats.[33] The Portico format registry is based upon the GDFR work and will be replaced by GDFR when appropriate.
- JSTOR/Harvard Object Validation Environment (JHOVE): JHOVE is a collaborative project to develop an extensible framework for file format validation.[34] JHOVE is used to validate the format of the files in the archive and we are developing additional modules as needed.

---

[29] *MPEG-21 Part 2: Digital Item Declaration Language (DIDL)* [Website] (Cover Pages, 2004) http://xml.coverpages.org/mpeg21-didl.html

[30] *METS: Metadata Encoding & Transmission Standard* [Website] (Library of Congress) http://www.loc.gov/standards/mets/

[31] *Journal Archiving and Interchange DTD*

[32] "Reference Model for an Open Archival Information System - OAIS."

[33] *Global Digital Format Registry* [Website] (Harvard University Library, 2005) http://hul.harvard.edu/gdfr/

[34] *JHOVE - JSTOR/Harvard Object Validation Environment* [Website] (Harvard University Library, 2006) http://hul.harvard.edu/jhove/. Portico, when it was the JSTOR Electronic-Archiving Initiative, partnered with Harvard to create this open source tool.

- Preservation Metadata: Implementation Strategies (PREMIS): The PREMIS working group developed a data dictionary and XML schema for preservation metadata.[35] Portico staff participated in the PREMIS working group and it influenced the preservation metadata captured in the Portico operations process.

Ongoing Processes and Preservation

Portico receives the source files from the publishers and through an automated process assembles these files into an archival package that is independent of the original publisher's format. This package is then ingested into the archive. The ingestion process is driven by a set of publisher specific rules and tools built through our analysis of the publisher's source files and discussions with the publisher. (Much of this analysis happens before we begin processing data, however it is revisited as needed when newly received content does not match the current rule set.) The ingest process verifies the structure and relationships of the source files, identifies the formats of the source files, and validates the compliance of the files to their format specifications. The SGML or XML header or full-text files are normalized to the NLM DTD in this ingest process. The process then creates the Portico required preservation metadata to track the actions taken on the data. The entire ingest process is managed by operations staff who also perform manual quality assurance on sample data. The last step of the ingest process is to deposit the article into the Portico archive.

Every article within the archive and all its associated files are considered an archival unit and are represented by a human-readable XML file in a modified version of the METS schema, known as Portico METS or "PMETS".[36] This PMETS file references all the component source files necessary to render the article. The entire archive can be recreated from the component source files and the PMETS files without reference to any proprietary data structures other than the human-readable XML PMETS files. For security purposes, Portico has developed a quarterly schedule to replicate the archive across multiple media, within multiple archive management systems, and in separate geographic locations.

The operational processing of the source files ensures that we understand the state of specific files (e.g. which files are well-formed and which corrupt) and captures their relationships in a uniform manner across all publishers and titles. We shall audit the archive on a regular basis and monitor the acceptance of specific file formats in the community. Portico preservation and

[35] *PREMIS (Preservation Metadata: Implementation Strategies) Working Group* [Website] (OCLC, 2005) http://www.oclc.org/research/projects/pmwg/

[36] *METS: Metadata Encoding & Transmission Standard*

format policies will be regularly reviewed and updated as needed and the content will be migrated when appropriate. As shown in Figure 1: Portico Preservation Process, the ingest and normalization of source files and the analysis and updating of archival policies are ongoing tasks that inform our preservation activities.

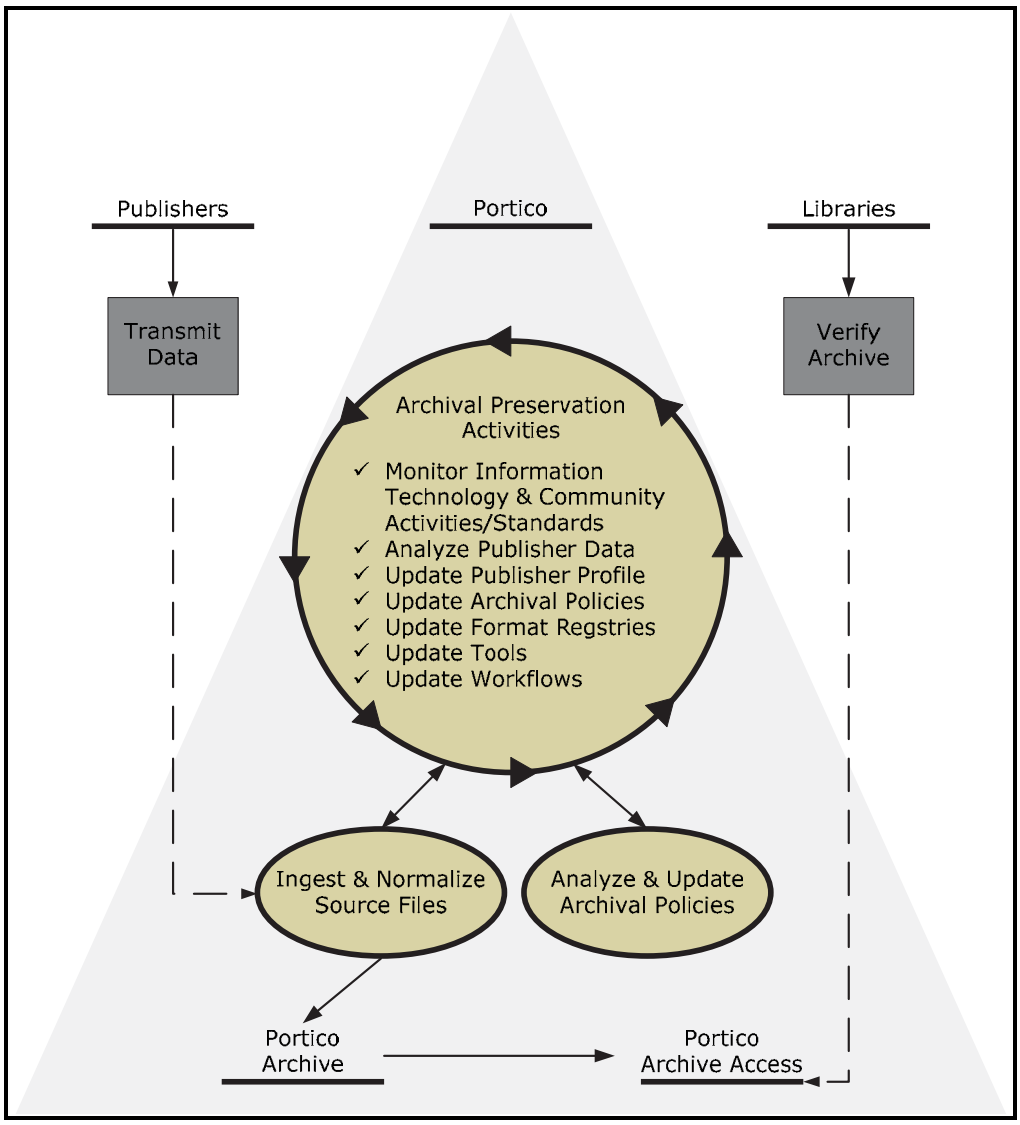

Figure 1: Portico Preservation Process

A more detailed examination of the Portico ingest and archiving process will be forthcoming in the second paper of this series. We also expect to share information on Portico's technical design and architecture in one or more forthcoming technical papers.

Referenced Organizations and Projects

- The Andrew W. Mellon Foundation http://www.mellon.org
- Association of Research Libraries http://www.arl.org/
- Center for Research Libraries (CLR) http://www.crl.edu/
- The Consultative Committee for Space Data Systems (CCSDS) http://public.ccsds.org/
- Creative Archiving at Michigan and Leeds Emulating the Old On the New (CAMiLEON) http://www.si.umich.edu/CAMILEON
- Digital Library Federation (DLF) http://www.diglib.org/
- Digital Preservation Coalitions (DPC) http://www.dpconline.org/
- Electronic Publishing Initiative at Columbia (EPIC) http://www.epic.columbia.edu/
- International Organization for Standardization (ISO) http://www.iso.org/
- Ithaka http://www.ithaka.org
- JSTOR http://www.jstor.org
- Koninlijke Bibliotheek (The National Library of the Netherlands also known as "the KB") http://www.kb.nl/index-en.html
- The Library of Congress http://www.loc.gov/
- The National Archives (NARA) http://www.archives.gov/
- National Center for Biotechnology Information (NCBI) http://www.ncbi.nih.gov/
- National Digital Information Infrastructure and Preservation Program (NDIIPP at The Library of Congress) http://www.digitalpreservation.gov/
- National Library of Medicine (NLM) http://www.nlm.nih.gov/
- Online Computer Library Center OCLC http://www.oclc.org
- Pew Internet & American Life Project http://www.pewinternet.org/
- Portico http://www.portico.org
- Preserving Access to Digital Information (PADI at the National Library of Australia) http://www.nla.gov.au/padi/
- The Research Libraries Group (RLG) http://www.rlg.org

Bibliography

Archiving Electronic Journals. [Website] Digital Library Federation (2003). http://www.diglib.org/preserve/ejp.htm (Accessed May 18, 2006).

ARL Directory of Electronic Journals, Newsletters and Academic Discussion Lists, 7th Edition. ARL (2001). https://db.arl.org/edir/ (Accessed May 22, 2006).

Audit Checklist for the Certification of Trusted Digital Repositories: Draft for Public Comment. Mountain View, CA:RLG and NARA, 2005. http://www.rlg.org/en/page.php?Page_ID=20769 (Accessed May 22, 2006).

Budd, John M., and Lynn Silipigni Connaway. "University Faculty and Networked Information: Results of a Survey." *Journal of the American Society for Information Science* 48, no. 9 (1997): 843-52. http://www3.interscience.wiley.com/cgi-bin/abstract/39759/ABSTRACT (Accessed June 9, 2006).

Bugeja, Michael, and Daniela V. Dimitrova. "The Half-Life Phenomenon: Eroding Citations in Journals." *The Serials Librarian* 49, no. 3 (2005): 115-23. http://www.haworthpress.com/store/E-Text/View_EText.asp?sa=3&s=J123&v=49&i=3&fn=J123v49n03%5F10 (Accessed May 23, 2006).

CRL Auditing and Certification of Digital Archives. [Website] Center for Research Libraries (2005). http://www.crl.edu/content.asp?l1=13&l2=58&l3=142 (Accessed June 19, 2006).

Dellavalle, Robert P., Eric J. Hester, Lauren F. Heilig, Amanda L. Drake, Jeff W. Kuntzman, Marla Graber, and Lisa M. Schilling. "Going, Going, Gone: Lost Internet References." *Science* 302, no. 5646 (2003): 787 - 88. http://www.sciencemag.org/cgi/content/full/302/5646/787 (Accessed May 22, 2006).

Dewatripont, Mathias, Victor Ginsburgh, Patrick Legros, Alexis Walckiers, Jean-Pierre Devroey, Marianne Dujardin, Francoise Vandooren, Pierre Dubois, Jerome Foncel, Marc Ivaldi, and Marie-Dominique Heusse. *Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe*. European Commission, 2006. http://europa.eu.int/comm/research/science-society/pdf/scientific-publication-study_en.pdf (Accessed May 22, 2006).

EPIC Faculty Survey. [Power Point Presentation] Electronic Publishing Initiative at Columbia (2003). http://www.epic.columbia.edu/eval/FacSurv.903.ppt (Accessed May 19, 2006).

Garrett, John, and Don Waters. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. Task Force on Archiving of Digital Information, 1996. http://www.rlg.org/ArchTF/ (Accessed May 2, 2006).

Global Digital Format Registry. [Website] Harvard University Library (2005). http://hul.harvard.edu/gdfr/ (Accessed May 2, 2006).

Guthrie, Kevin. "What Do Faculty Think of Electronic Resources." Presented at the JSTOR ALA Annual Conference Participants' Meeting, June 17, 2001. http://www.jstor.org/about/faculty.survey.ppt (Accessed May 23, 2006).

Guthrie, Kevin, and Roger Schonfeld. "What do Faculty Think of Electronic Resources? Findings from the 2003 Academic Research Resources Study." Presented at the CNI Task Force Meeting, Alexandria, Virginia, April 16, 2004. http://www.cni.org/tfms/2004a.spring/presentations/CNI_Guthrie_What.ppt (Accessed May 23, 2006).

JHOVE - JSTOR/Harvard Object Validation Environment. [Website] Harvard University Library (2006). http://hul.harvard.edu/jhove/ (Accessed May 2, 2006).

Journal Archiving and Interchange DTD. [Website] National Center for Biotechnology Information of the National Library of Medicine (2004). http://dtd.nlm.nih.gov/ (Accessed May 18, 2006).

JSTOR Usage Statistics. JSTOR. http://stats.jstor.org/ (Accessed May 22, 2006).

Lyman, Peter, and Hal Varian. *How Much Information 2000*. University of California, 2000. http://www2.sims.berkeley.edu/research/projects/how-much-info/index.html (Accessed May 22, 2006).

———. *How Much Information 2003*. University of California, 2003. http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/ (Accessed May 22, 2006).

Madden, Mary. *Internet Penetration and Impact*. Pew Internet & American Life Project, 2006. http://www.pewinternet.org/PPF/r/182/report_display.asp (Accessed June 21, 2006).

May 2006 Web Server Survey. Netcraft, 2006. http://news.netcraft.com/archives/2006/06/04/june_2006_web_server_survey.html (Accessed May 22, 2006).

METS: Metadata Encoding & Transmission Standard. [Website] Library of Congress. http://www.loc.gov/standards/mets/ (Accessed May 18 2006).

MPEG-21 Part 2: Digital Item Declaration Language (DIDL). In *Technology Reports*, [Website] Cover Pages (2004). http://xml.coverpages.org/mpeg21-didl.html (Accessed June 9, 2006).

New Jour Archive. [Email List Archive] University of California at San Diego Libraries (2006). http://gort.ucsd.edu/newjour/ (Accessed June 21, 2006).

PREMIS (Preservation Metadata: Implementation Strategies) Working Group. [Website] OCLC (2005). http://www.oclc.org/research/projects/pmwg/ (Accessed June 19, 2006).

Rainie, Lee, Dan Packel, Susannah Fox, John Horrigan, Amanda Lenhart, Tom Spooner, Oliver Lewis, and Cornelia Carter. *More Online, Doing More*. Pew Internet & American Life Project, 2001. http://www.pewinternet.org/PPF/r/30/report_display.asp (Accessed June 21, 2006).

Reference Model for an Open Archival Information System - OAIS. National Aeronautics and Space Administration - Consultative Committee for Space Data Systems, 2002. http://public.ccsds.org/publications/archive/650x0b1.pdf (Accessed May 2, 2006).

Running, Jordan. The Top 25 Web Sites of 1994. In *downloadsquad*, [Weblog] (2006). http://www.downloadsquad.com/2006/04/14/the-top-25-web-sites-of-1994/ (Accessed May 22, 2006).

Trusted Digital Repositories: Attributes and Responsibilities. Mountain View, CA:RLG-OCLC, 2002. http://www.rlg.org/longterm/repositories.pdf (Accessed May 19, 2006).

Urgent Action Needed to Preserve Scholarly Electronic Journals. Digital Library Federation, 2005. http://www.diglib.org/pubs/waters051015.htm (Accessed May 1, 2006).

Young, Mark, and Martha Kyrillidou. *ARL Statistics 2003-2004*. Washington: Association of Research Libraries, 2005.