



The Center for Research Libraries

10/14/2005

Archive Profile

Portico: An Electronic Archiving Service

by Robin Dale, Project Director, Certification of Digital Archives Project and Program Officer, Research Libraries Group, US

Organization:

Mission Statement:

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

Governance:

Portico began as the Electronic-Archiving Initiative launched by JSTOR in 2002 with a grant from The Andrew W. Mellon Foundation. The purpose of the Initiative was to build a model for a sustainable digital archive of scholarly literature published in electronic form. For more than three years project staff worked to develop the necessary archiving technology while engaging in extensive discussions with publishers and libraries about how to craft an approach that balances the needs of publishers and libraries while generating sufficient funding for the archive. In 2004, the Electronic-Archiving Initiative became a part of Ithaka, a new non-profit organization with a mission to accelerate the productive uses of information technologies for the benefit of higher education around the world. In 2005, a new electronic archiving service resulting from the Initiative's extensive collaborations with the community was launched under the name Portico and the direction of Eileen Fenton, Executive Director. Initial support for Portico is provided by JSTOR, The Andrew W. Mellon Foundation, and the Library of Congress. Ithaka continues to incubate and support Portico's development, and the Ithaka Board of Trustees provides formal governance oversight (See Appendix 1). Additional guidance is received from the Portico Advisory Committee, which is comprised of representatives from the scholarly publishing and academic library communities. [Click here to link to Portico's Organizational Chart.](#)

Funding:

Initial funding for Portico is provided by JSTOR, Ithaka, and The Andrew W. Mellon Foundation. In addition, Portico was recently awarded a grant from the Library of Congress as part of the National Digital Information Infrastructure and Preservation Program (NDIIPP) in support of further development of the Portico organizational and technological infrastructure.

As an archiving service with a long-term perspective, Portico believes it is important to develop an economic model which permits the archive to build diversified sources of revenues which can be used to cover ongoing operating costs. Support for the archive should come from the primary beneficiaries of the service - libraries and publishers - but support from other sources, such as charitable foundations and government agencies, is also expected. Portico's community-based approach builds from the assumption that libraries and publishers will share the costs of the electronic archiving service and will secure the benefits and opportunities for savings that the archive enables.

Service and Access Model:

The Portico archive model builds from two core assumptions: first, material is preserved for eventual use and access and second, a completely and perpetually “dark” archive is not desirable. From these assumptions and drawing heavily upon extensive input from libraries and publishers, the Portico electronic archiving service has been shaped as follows: Portico will act as a centralized repository. Publishers will submit source files of scholarly electronic journals to Portico, and Portico will normalize these source files to an archival format and provide long-term archival management and format migration as needed. Libraries that support the archive will have campus-wide access to archived content when very specific conditions or “trigger events” occur. Campus-wide access for participating institutions will be triggered when:

- a publisher ceases operations and no entity purchases and makes its titles accessible;
- a publisher ceases to publish a title and it is not offered by another entity;
- back issues are removed from a publisher’s site and are not available elsewhere;
- catastrophic and sustained failure disables the publisher’s delivery platform; or
- a publisher opts to rely upon Portico to meet perpetual access obligations.

In addition to access triggered by these relatively rare events, a small number of librarians at each participating institution will be granted password-controlled access to enable independent verification of the archive’s integrity. This verification access, which is granted to the entire archive, is not intended to be used as a replacement for commercial document delivery services or to fulfill inter-library loan requests.

Current Status:

Portico is securing archival agreements with both publishers and libraries for support of the archival service. Initial agreements are expected in the latter part of 2005 and early 2006. Until final publisher agreements are secured, Portico's work is focused on processing sample content provided by ten cooperating publishers.

The Portico System

Content Characteristics:

The Portico approach to archiving electronic journals is based on archiving “source files” rather than presentation or delivery files. Source files are electronic files containing graphics, text, or other material that comprise an electronic journal article, issue, or volume. Source files may differ from files presented online most typically by including more information or higher quality graphics. Publishers deliver to Portico the source files of their electronic journal editions (e.g., SGML or XML text; PDF or PostScript page images; graphics, media, and supplemental files); Portico normalizes the text files from their original proprietary format to an XML DTD based on the NLM Archival DTD and deposits the content in the Portico repository. Portico retains the deposited files for the long term; the normalized files will be migrated as needed to new formats and technologies. Portico preserves the intellectual content of the journal, including the text, images, and functionality such as internal and external linking, when included in the provided data. The exact “look and feel” of the HTML rendition and publishers’ value-add features such as e-commerce functions and personalization are not targeted for preservation.

Technical Architecture:

The Portico hardware and software systems were designed by the Portico Technology Unit, currently a staff of eleven headed by Evan Owens, Chief Technology Officer. Infrastructure services (networking, hardware installation, configuration, and administration, and database administration) are provided by the Shared IT unit of Ithaka. (A component of Ithaka’s mission is “providing administrative and technological services that can be shared by affiliated entities to increase effectiveness, lower costs and allow affiliates' to focus efforts on mission-related activities.”¹)

The resulting technical infrastructure—which is in place—is developed and managed by a combination of Portico and Ithaka staff and services supplied by Princeton University. The systems hardware and masters or primary copies of all electronic journals are stored in the server room of the Princeton University Computing Center. Network connectivity for the system is also contracted through Princeton University. Local server management (hardware, software, and backups) is performed by the Ithaka Shared IT staff on behalf of Portico. Finally, access management is provided through JSTOR, leveraging its existing extensive infrastructure for delivery. Portico will export data from the archive system to the delivery platform managed by JSTOR.

Off-line copies of archived content will be stored in a combination of commercial storage providers and with strategic partners. Online replication of archival content will also be carried out either via a cooperative arrangement with JSTOR or via other partners or both. According to Owens, purchased services such as network access and storage are essentially commodities and contracting these out allows Portico staff to focus on the unique and critical archiving functionalities.

¹ Ithaka, Mission and background. Retrieved Aug. 02, 2005, from <http://www.ithaka.org/about/mission.htm>.

Scale:

All data currently processed by Portico has been submitted by publishers for piloting purposes; no archival responsibility has been assumed for these files. When formal archival agreements are in place, Portico expects to ingest the equivalent of approximately three million printed pages or about two terabytes of data by the end of 2006. The exact size of the data depends on the mix of formats provided by the publishers.

Users (Designated Communities):

Portico's mission, to ensure that electronic journals remain available to future generations of scholars, researchers, and students, clearly identifies Portico's ultimate user community as that set of individuals who at some point in the future must turn to the archive for access to importantly scholarly resources. Addressing the needs of these future users, however, must be done in collaboration with current content producers (publishers) and libraries, two constituents who have a broad and ongoing interest in the creation, management, and preservation of scholarly literature. Both of these constituents, in different ways, help Portico to prepare for future generations of scholars, researchers and students, and without the cooperation of both of these parties, Portico's community-based, cooperative approach can not succeed.

Publishers. Publishers who participate in Portico sign an archiving license, granting perpetual non-exclusive rights to Portico according to the service model discussed above. At the publisher's option, Portico may provide post-cancellation "perpetual access" to Portico supporting libraries on behalf of publisher in conformance with their customer agreements. Publishers deposit content in a timely way following publication and make an annual financial contribution to support the archive. The benefits to publishers include reducing (or eliminating) the publisher's internal archiving costs, meeting library demand for a trusted, third-party archive, and meeting library demand for perpetual access without negative impact on publisher's operations.

Libraries. Libraries who participate in Portico sign an archiving license and make a one-time and annual financial contribution. The benefits to libraries include creating a means to act upon the traditional preservation mandate, securing protection against eventual loss of access, and providing a practical way to maintain continuity of library collections.

The Portico service to libraries does not alleviate libraries' need for access to new or current scholarly electronic journals. Instead, Portico enables libraries to take advantage of reduced system-wide preservation costs that, through a cooperative mechanism, eliminates the need for each library to locally receive and archive electronic journal literature. Portico may help libraries consolidate preservation costs, reduce or eliminate their need for print subscriptions, and ensure continuity of access to important scholarly literature.

Audit Dates:

21-23 March, 2006

Appendix 1: Governance

Ithaka Board of Trustees

September 2005

William G. Bowen
Chairman, *Ithaka Board of Trustees*
President, *The Andrew W. Mellon Foundation*

Paul A. Brest
Vice Chairman, *Ithaka Board of Trustees*
President, *The William and Flora Hewlett Foundation*

Charles E. Exley, Jr.
Former Chairman and Chief Executive, *NCR Corporation*

Kenneth C. Frazier
Senior Vice President and General Counsel, *Merck & Co., Inc.*

Kevin M. Guthrie
President, *Ithaka*

W. Drake McFeely
President, *W. W. Norton & Company, Inc.*

Mamphela A. Ramphele
Managing Director, Human Development, *The World Bank*

Lawrence R. Ricciardi
Senior Advisor, *Jones Day and Lazard Frere & Company*

Charles M. Vest
Former President, *Massachusetts Institute of Technology*

Portico Advisory Committee

September 2005

John Ewing

Executive Director and Publisher, *American Mathematical Society*

Kevin M. Guthrie

President, *Ithaka*

Daniel Greenstein

Associate Vice Provost for Scholarly Information,
University Librarian, California Digital Library, *University of California*

Anne R. Kenney

Associate University Librarian for Instruction, Research, and Information Services
Cornell University Library

Clifford Lynch

Director, *Coalition for Networked Information*

Carol A. Mandel

Dean, Division of Libraries, *New York University*

David M. Pilachowski

College Librarian, *Williams College*

Rebecca Simon

Director, Journals & Digital Publishing Division and Associate Director,
University of California Press

Michael Spinella

Executive Director, *JSTOR*

Suzanne E. Thorin

University Librarian and Dean of the Library, *Syracuse University*

Mary Waltham

Founder, *Mary Waltham Publishing Consultant*

Craig Van Dyck

Vice President, Operations, Scientific, Technical, and Medical Publishing Division
John Wiley & Sons, Inc.