

Emily Nabasny  
Digital Preservation  
Kara Van Malssen  
October 1, 2013

## **Portico**

As researchers become more dependent on digital resources, the need to ensure the availability of these materials long-term becomes more urgent. Portico grew out of an initiative started in 2002 by JSTOR as a solution for the preservation of electronic scholarly journals. The initiative transitioned to ITHAKA in 2005, renamed Portico, and has since become a sustainable digital archive that can serve the academic community, while allowing publishers and libraries to feel safe as they began to rely more and more on digital content. Portico is a digital archive whose mission is to preserve scholarly content in electronic form through source file normalization and format migration to ensure that they remain available for future generations of scholars, students, and researchers.

As more publishers and researchers become reliant on electronic content, the continually growing problem of long-term access becomes ever more relevant. What would happen if a publisher goes out of business? Or if they no longer offer back issues? That is what Portico seeks to solve. Portico is well over 8 years into this project and their digital preservation approach combining long-term commitment from organizations with long-term content management is working very well. Their main goal of addressing the need of a permanent archive of electronic scholarly content means that they are continually building their archive and looking to preserve more and more content. This service, overseen by a Board of Trustees who represent the academic community (university presidents, librarians, publishers, and faculty members), along with an advisory committee of librarians and publishers, ensures the review and guidance of Portico's policies, and the use of best practices in the continuing digital preservation of the content. Portico utilizes a community-based, cooperative approach to address the challenges of digital preservation and achieve the best results.

Content of the project includes not only born digital content such as e-books and e-journals, but content published in print and electronic form, and digitized historical collections (d-collections) as well. The scope of what is being preserved spans 849 libraries from over 14 countries, 219 publishers, and over 27 million archival units are currently being preserved, and thousands more are added to the archive each day. Over 227,000 thousand e-book titles, 16,000 e-journal titles, as well as 72 digital collections are in the archive. (A detailed list of the content in the archive can be found online.<sup>1</sup>) Once content is deposited in the archive, it remains there.

As the preservation of this content can present many technological challenges, content is preserved through a migration strategy based on digital format. Portico has 2 key points to this strategy: the first, to identify preservation metadata at the initial point of the preservation process, and the second, to only migrate content at the point when it is necessary. This strategy is detailed in a 5-step preservation approach implemented when

---

<sup>1</sup> Portico Archive Content: <http://www.portico.org/digital-preservation/the-archive-content-access/content-in-the-archive>

taking in new content. The 5-steps of preservation are: 1) Preservation Planning, 2) Receipt & Inventory Management, 3) Processing & Archival Deposit, 4) Monitoring & Management, and 5) Content Delivery.

When Portico receives the source files from the publishers, the preservation process automatically begins, analyzing the files and putting them into an archival package separate from the original file (step 1: Preservation planning). Portico operations staff convert the full-text files into archival format based on the Journal Archiving and Interchange DTD created by the National Library of Medicine (NLM). As it is source files, and not delivery files, that the archive takes in, high-quality HTML, PDFs, SGML or XML text files, images, and supplemental data such as video and audio is being preserved and made available. The package and a preservation plan are developed based on the content and the needs of the provider and libraries. New tools necessary to the execution of the plan may also be created if they are necessary to: retrieve or load content, repackage the content, or migrate files in certain formats. If files cannot be converted, the files undergo byte preservation. The content is then manually or automatically transferred by way of: portable media, File Transfer Protocol (FTP), Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) or Open Archives Initiative Object Reuse Exchange (OAI-ORE), or through Portico's custom software, and loaded into the storage area (step 2: Receipt & Inventory Management).

As the content is processed and deposited in the archive (step 3: Processing & Archival Deposit), preservation metadata is created following the OAIS model, and includes all DTDs, schemas, and documentation necessary to understand the content. The metadata also includes: validating files for the formats, checksums, technical and descriptive metadata pulled from the source file, and a record of all actions taken on the content. Portico works to preserve the intellectual content of the files as well as the resources themselves by using a modified version of METS to track actions taken on the source file.

After ingest, and over time, the operations staff also monitors the condition of the content using sample data (step 4: Monitoring & Management). During audits content can be replicated, checked for problems, repaired or replaced if corrupted, audited, updated as per preservation needs, or migrated to new formats. Portico's archive is also replicated offline with multiple copies. The final stage of Portico's preservation plan (step 5: Content Delivery) grants access to preserved content to Portico supporters when such a time arrives that titles are no longer available from the publisher or another source. This circumstance known as a "trigger event," also includes when a publisher: stops operating, ceases publishing a title, no longer offers back issues, or experiences failure of a delivery platform. Access is also available to libraries when a claim has been filed and okayed by the publisher.

A major design of Portico that makes its approach interesting is that Portico is both a "dark" archive and a "light" archive. "Dark," meaning that when content is still available outside of Portico it cannot be accessed through the archive. The archive becomes "light" when participants audit the archive, and when content is made available as a result of "trigger event", which renders the content unavailable anywhere else. (There is a list on Portico's website of triggered content.<sup>2</sup>) Participating publishers and libraries are given

---

<sup>2</sup> Portico. *Triggered Content*. 2013. <http://www.portico.org/digital-preservation/the-archive-content-access/access-to-archived-content>. (ITHAKA 2005-2013)

password access to the archived content for the purpose of auditing, verification, and use. Non-participants can never access Portico's content. As a result of this design, Portico secures the right to obtain, normalize, migrate, and exhibit content for preservation purposes.

In 2010 the Center for Research Libraries (CRL) certified Portico as a trustworthy digital repository following an audit they conducted in 2009. The audit concluded that ...the practices and services described in Portico's public communications and published documentation are generally sound and appropriate to the content being archived and the needs of the CRL community. Moreover the CRL Certification Advisory Panel expects that in the future, Portico will continue to be able to deliver content that is understandable and usable by its designated user community.<sup>3</sup>

Libraries and publishers around the world know that they can rely on Portico when triggered events lead to loss of access. It is interesting and admirable to see so many participants who not only support the work that Portico is doing, but they understand the reasons and have already benefited from it. At Drew University, the library was hesitant to have exclusively electronic journals for uses because they were concerned with availability, but now they use Portico and rely on its archive to fulfill their users needs.<sup>4</sup> Beyond only using Portico as an archive, Jeffrey Horrell of Dartmouth College uses Portico both as an archive, and as an aid in discussions with "faculty and students about storage decisions for print knowing that we may not need to maintain certain materials or they may be stored off-site."<sup>5</sup> Portico is assisting not only in the preservation of content but to the global discussion of archiving as well.

Overall Portico remains an important and worthwhile option for the digital preservation of scholarly content. Their mission to continually store, migrate, and provide access to these materials gives libraries, publishers, and scholars the confidence to be more comfortable when choosing to rely on digital content. Portico's storage and archive practices are carefully thought out and executed. The actions taken on content continually reveal Portico's focus on the necessity that is long-term preservation over immediate access. They, and their participants, recognize the larger benefits of what this project is achieving, while also acknowledging the need for more conversation and action on digital preservation.

---

<sup>3</sup> Center for Research Libraries. "Center for Research Libraries: Report on Portico Audit Findings." *Center for Research Libraries*. January 10, 2010. <http://www.crl.edu/sites/default/files/attachments/pages/CRL%20Report%20on%20Portico%20Audit%202010.pdf> (accessed 2013).

<sup>4</sup> Connors, Linda. *Looking to the Future*. 2008. <http://www.portico.org/digital-preservation/who-participates-in-portico/what-participants-say-about-portico/looking-to-the-future> (accessed 2013).

<sup>5</sup> Horrell, Jeffrey L. *Envisioning Strategic Partnership*. <http://www.portico.org/digital-preservation/who-participates-in-portico/what-participants-say-about-portico/envisioning-strategic-partnership> (accessed 2013).

## Bibliography/Web-ography

Aryis, Paul. *What Participants Say*. <http://www.portico.org/digital-preservation/who-participates-in-portico/what-participants-say-about-portico> (accessed 2013).

Center for Research Libraries. "Center for Research Libraries: Report on Portico Audit Findings." *Center for Research Libraries*. January 10, 2010. <http://www.crl.edu/sites/default/files/attachments/pages/CRL%20Report%20on%20Portico%20Audit%202010.pdf> (accessed 2013).

Connors, Linda. *Looking to the Future*. 2008. <http://www.portico.org/digital-preservation/who-participates-in-portico/what-participants-say-about-portico/looking-to-the-future> (accessed 2013).

Fenton, Eileen. "Portico: A Collaborative Approach to Preservation." *Digital Preservation Europe*. 2007. [http://www.digitalpreservationeurope.eu/publications/portico\\_briefing\\_paper\\_def.pdf](http://www.digitalpreservationeurope.eu/publications/portico_briefing_paper_def.pdf) (accessed 2013).

Fitzpatrick, Kathleen. *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*. New York: New York University Press, 2011.

Horrell, Jeffrey L. *Envisioning Strategic Partnership*. <http://www.portico.org/digital-preservation/who-participates-in-portico/what-participants-say-about-portico/envisioning-strategic-partnership> (accessed 2013).

Kirchhoff, Amy, and Eileen Gifford Fenton. "Archiving Electronic Journals: An Overview of Portico's Approach." *Portico - A Digital Preservation and Electronic Archiving Service*. August 2006. <http://www.portico.org/digital-preservation/wp-content/uploads/2009/12/PapersFromPortico.1.Overview.pdf> (accessed September 19, 2013).

Morrissey, Shelia, et al. "Portico: A Case Study in the Use of the Journal Archiving and Interchange Tag Set for the Long Term Preservation of Scholarly Journals." *National Center for Biotechnology Information*. 2010. <http://www.ncbi.nlm.nih.gov/books/NBK47087/>.

Portico. *Portico- A Digital Preservation and Electronic Archiving Service*. 2005-2013. <http://www.portico.org/digital-preservation/> (accessed September 20, 2013).

—. *Triggered Content*. 2013. <http://www.portico.org/digital-preservation/the-archive-content-access/access-to-archived-content> (accessed 2013).

Royal Society Publishing. *Digital Preservation - Our Commitment to Archiving*. 2013.  
<http://royalsocietypublishing.org/site/librarians/DigitalPreservation.xhtml> (accessed  
September 2013).