

Online Journal Publishing Best Practices:
From Production to Preservation

Allen Press Editorial Seminar
September 2006

Evan Owens
Chief Technology Officer
Portico
evan.owens@portico.org



Why Publish Journals?

- Communications
 - From today's authors to today's reader
 - From today's authors to tomorrow's readers
- Mediation and Quality Control
 - Peer review
 - Editorial and production values
- Fixity and Permanence
 - "Record of scholarship"
 - "Version of record"
 - "Build on the record of science"
- Components in a System of Communication
 - Not an isolated artifact
 - Hence need for "best practices"



From Print to Online: What has Changed?

- The unit of access to information
 - Article rather than issue
 - Direct links into articles from outside (e.g., DOI)
- The delivery format(s)
 - Multiple renditions or version (HTML, PDF, print and online)
- The stability of the object
 - Fragile electronic objects (cf. acid-free paper)
 - Can break, can be changed inadvertently
 - Sometimes dynamically constructed
 - Risk of technological obsolescence
 - Formats and software
- The stability of the entire system
 - Often single delivery system or vendor



Online Best Practices: Citation and Linking

- Content should be citable and linkable
 - Example: ISSNs and DOIs being assigned to older digitized print
- Links should be persistent and actionable
 - Article DOIs required
 - DOI-based reference linking expected
- New identification schemes are okay
 - Example: APS article numbers replacing page numbers print and online
 - Example: BePress numbered articles
- Having no identification scheme is not okay!
 - Example: a humanities journal with uncitable and unlinkable online content
- Put essential information on every article
 - Example: a book review where the name of the reviewer was only found on the table of contents; bad idea in print, fatal online



Online Best Practices: Versions and Revisions

- Versions
 - Bad: inadvertent differences between renditions (HTML, PDF, print, XML/SGML)
 - Example: “Abstract” or “Summary” in generated text
 - Okay: intentional explicit differences between versions
 - Help the reader understand what is going on
 - Example: captions that indicate color online, b&w in print
 - Example: highlight additional content available online
- Revisions / Updates
 - Journals are publications of record
 - Changes after publication are dangerous
 - Silent changes undermine trust in the journal publishing system
 - A formal policy regarding revision of online content is essential
 - Readers need to know if something is changed
 - Example: erratum appended to original article



Online Best Practices: Issue-level information

- Sometimes not available online
 - Covers
 - Front matter
 - Back matter
 - Advertisements
- Sometimes updated rather than versioned
 - Example: editorial board membership
- Print journal issues captured the historical context
- No standard approach yet in online publications
- Some publishers posting PDF for front / back matter



Online Best Practices: Quality Control

- Mistakes happen
 - In print or online!
- Good content management practices
 - Written documentation, naming conventions, coding conventions
- Format validation
 - Production content: PDF, XML, graphics
 - Author-supplied supplemental content: various formats
- Change control
 - Tracking updates to content
 - Distributing updated content to business partners
- Handing content to a third party is a useful exercise
 - A surrogate for handing content off over time
 - Portico's work with content from leading publishers



What is Digital Preservation?

- Ensuring long-term viability of electronic objects
- 20, 50, 100 years from now, can we
 - read the files?
 - understand the structure of the files?
 - be sure that we have an authentic copy of the work?
- Layers
 - Physical Layer: storage media
 - Logical Layer: file formats, structured data
 - Conceptual/Intellectual Layer: the “work”
- Approaches to preservation:
 - Emulate (or maintain) the original technology
 - Migrate (and/or normalize) to currently supported formats
 - Byte preserve for future digital archeologists



What Is Portico?

- Portico is a not-for-profit organization with a mission and singular focus to provide a permanent archive of electronic scholarly journals
- Portico is a centralized archive that is open to all peer-reviewed journals
- Portico is a community-based, cooperative approach to the digital preservation challenge
- Portico's Mission Statement:

To preserve scholarly literature published in electronic form and to ensure that these materials remain available to future generations of scholars, researchers, and students



Portico's History

- In 2002, JSTOR initiated a project known as the Electronic-Archiving Initiative, the precursor to Portico
- Began with 2-year pilot phase (2003-2005) working with 10 publishers to gain experience in content formats and technological challenges
- Portico was launched in 2005 by JSTOR and Ithaka, with support from The Andrew W. Mellon Foundation
- Portico receives \$3 million grant award from the Library of Congress Nation Digital Information Infrastructure and Preservation Program (NDIIPP) in October 2005
- First publishers signed 4th quarter 2005
- First libraries signed 1st quarter 2006
- Operations “live” 1/06 with initial content ingested



Portico's Approach to E-Journal Archiving

- Publishers deliver to Portico the “source files” of electronic journals (SGML, XML, PDF, etc.)
- Portico converts the files from their original proprietary format to an archival format based on the NLM Archival DTD
- Portico deposits the content in the Portico repository
- Portico retains the source files for the long term; the normalized files will be migrated as needed to new formats
- Portico preserves the intellectual content of the journal, including the text, images, and limited functionality such as internal linking, but “look and feel” and publishers’ value-add features are not preserved



Portico's Access Model

- Portico offers access to archived content only to those libraries supporting the archive financially
- Access is offered only when specific trigger event conditions prevail and when titles are no longer available from the publisher or other sources
 - When a publisher ceases operations and titles are no longer available from any other source
 - When a publisher ceases to publish and offer a title and it is not offered by another publisher or entity
 - When back issues are removed from a publisher's web site and are not available elsewhere
 - Upon catastrophic failure by publisher's delivery platform for a sustained period of time
- When and if trigger event resolves, archive closes again



Sources of Support

- Primary support for the archive comes from the chief beneficiary of the archive – libraries - but publishers, charitable foundations, and government agencies will also offer support
- Anticipated sources of funding:
 - Libraries – 75%
 - Foundations/Government support – 20%
 - Publishers – 5%



Portico Participating Publishers

- American Anthropological Association
 - American Institute of Physics
 - American Mathematical Society
 - Annual Reviews
 - Berkeley Electronic Press
 - BioOne
 - Copernicus Publications
 - Elsevier
 - Institute of Physics Publishing
 - NOW Publishing
 - Oxford University Press
 - SAGE Publications
 - SIAM
 - Symposium Journals (UK)
 - United Kingdom Serials Group
 - University of Chicago Press
 - Wiley
- (See website for latest updated list)*



Benefits of Archiving

- Facilitates the community's transition to reliance upon electronic resources
- Enables system wide savings through reduced processing and storage of print resources with shared infrastructure of "virtual stacks"
- Provides a means of assuring access to e-resources over the long term and protects against gaps in library collections
- Provides a practical mechanism to address "perpetual access" needs
- Ensures the longevity of electronic content through digital preservation



Contact Information

Toni Tracy, Director of Publisher Relations

toni.tracy@portico.org

publisher.relations@portico.org

Evan Owens, Chief Technology Officer

evan.owens@portico.org

Our web site: www.portico.org

