

Digital Archiving: Realities of Implementing an Archiving Policy

Tim Ingoldsby



Digital Archiving
Assoc of Earth Science Editors
San Diego – 10 November 2006



Presentation Overview

- What is a digital archive and why is it important
 - What librarians are saying about the need for digital archives
- Reminder of terms contained in AIP's archival policy
- What approaches to archiving are available
 - Selection of Portico™ to become AIP's trusted third-party archive service provider



What Is A Digital Archive?

- Collection of all files required to exactly replicate the archival record of scholarship as contained in a journal
 - May include text, graphics files (in multiple resolutions), page image files, multimedia files (movies, audio clips, etc.), and even original manuscripts
 - Also should include metadata that describes each file and provides metrics about file content that can be used to verify completeness and validity of each archive component
- Infrastructure (and plan) for maintaining the archive over years, decades, and centuries
 - As standards change, software applications become obsolete, and formats evolve, archival components may need to be transformed to new formats



Why Archive?

- Teaching and research rely upon access to scholarship of past generations.
- Responsibility for preservation of print resources lay primarily with libraries, but today's scholarship is increasingly published electronically.
- Both publishers and libraries face the problem of the long-term preservation of electronic information resources for future access.
- Peer-reviewed scholarly electronic publications urgently require a robust preservation mechanism for long-term preservation and accessibility.



What Librarians Are Saying

- Will not switch to online only access unless confident that digital archive will be preserved
 - Librarians do not trust publishers – even society publishers – to carry out this long-term commitment
- Changes in scholarship (increasing use of movies, audio, incorporation of datasets, etc.) will eventually make the online version the *de facto* version of record (and some agreements with institutions already make this explicit)



AIP Response

- AIP Policy, *Archiving and Use of AIP Electronic Information*
 - Enacted in 1998, updated in 2003
 - [Archiving and Use of AIP Electronic Information](#)
- Policy implementation has been a challenge
 - All files are being protected via multiple backup copies, off-site storage, etc.
 - Development of true archive is an expensive development project



Possible Archiving Approaches

- Author self-archiving
- Institutional Repositories
- Publisher self-archiving
- Replication of print archiving – LOCKSS/CLOCKSS
- National Libraries
- Trusted Third Parties – Portico

Scitation

What is Portico?

- A service that provides a permanent archive of born-electronic scholarly journals to ensure that in 20, 50, or 100 years a valid, reliable copy of the work exists
- An archive open to all peer-reviewed journals that may have a print version in addition to an electronic version or may be available only in electronic form

Scitation

AIP and Portico Responsibilities

- AIP deposits content in a timely way following publication and Portico converts it into neutral archival format (Journal Archiving and Interchange XML DTD, NLM standard) and assumes responsibility for future format migration(s)
 - Portico provides digital copy of neutral archive version to AIP for its own use/re-use
- Portico loads converted content to its archival system and distributes preservation copies throughout the world
 - Participating libraries allowed periodic access only to confirm that all content is being appropriately deposited
- Archival access provided to users at participating institutions only under specific trigger events
 - Publisher ceases doing business (and title not transferred or acquiring publisher does not continue Portico relationship)
 - Publisher terminates title
 - Publisher ceases offering access to back issues
 - Publisher online service unavailable for 90 days due to catastrophic failure

Scitation

How does Portico sustain the archive?

- Three funding sources
 - Participating Libraries
 - Participating Publishers
 - Contributions
 - Charitable Foundation
 - Government Agencies
- Preservation strategy – reliance on preservation standards
- Preservation methodology – *migration*
 - Advisory committee of librarians, publishers, technologists

Scitation

How Well Is Portico Doing?

- 21 Publishers have signed up, representing 4,600+ journals
 - Amer. Mathematical Society, SIAM, Univ. Chicago Press, BioOne, Oxford Univ. Press, and now AIP
 - Elsevier, Wiley, Informa (Taylor & Francis) Sage, Annual Reviews
- More than 100 libraries have signed up
 - Portico expects to have 300 libraries by the end of 2006
- \$3 million grant from Library of Congress
- \$5 million grant from J-STOR
 - Additional funds from Mellon Foundation

Scitation

Current Status

- Agreement signed 10 July
- Samples of files to be deposited have been reviewed by Portico and “ingest” programs have been developed
- Operational status by 15 November 2006
 - Initial plan is to deposit all files published since 1 January 1999

Scitation

Future Plans

- AIP will offer management services to all Scitation Publishing Partners who elect to join us in Portico archive service
 - Packaging and deposit
 - Management of process
- AIP will participate in Portico oversight



Questions or Comments?

Email: tingoldsby@aip.org
www.scitation.org

