

**A Case Study in  
Collaborative Digital Preservation:  
The Portico Archive**

**Eileen Fenton  
Executive Director, Portico**

**Ithaka-CDL Forum  
June 9, 2009**



## Today's Topics

- The Portico Archive At a Glance
- Looking Forward: New Areas for Collaboration
  - Collection management & digital preservation
  - Preservation priorities beyond journals
  - Beyond audit: verification
  - Collaborative tools development



## Portico At A Glance: Collaborative Preservation

- Vision: Collaboratively supported infrastructure providing NFP long-term preservation of and access to scholarly literature published in electronic form.
- Goal: Facilitate a secure transition to greater reliance on e-resources at lower cost than multiple local independent efforts.
- Model: “Last resort” access for materials otherwise unavailable; provide one perpetual access mechanism, on publisher opt-in basis
- History: JSTOR/Ithaka pilot 2003-04, Portico launched in late 2005, content ingest began in early 2006.



## Portico At A Glance: Collaborative Preservation

- Collaboration includes scholarly publishers and libraries.
- Publishers range from commercial houses, to university presses to hundreds of scholarly societies and associations. Committed titles:
  - 9,170+ e-journals
  - 6,915+ e-books
- Library participation is international and comes from across the spectrum.
  - 334 domestic library participants (68%)
  - 157 international library participants (32%)
  - 13 countries

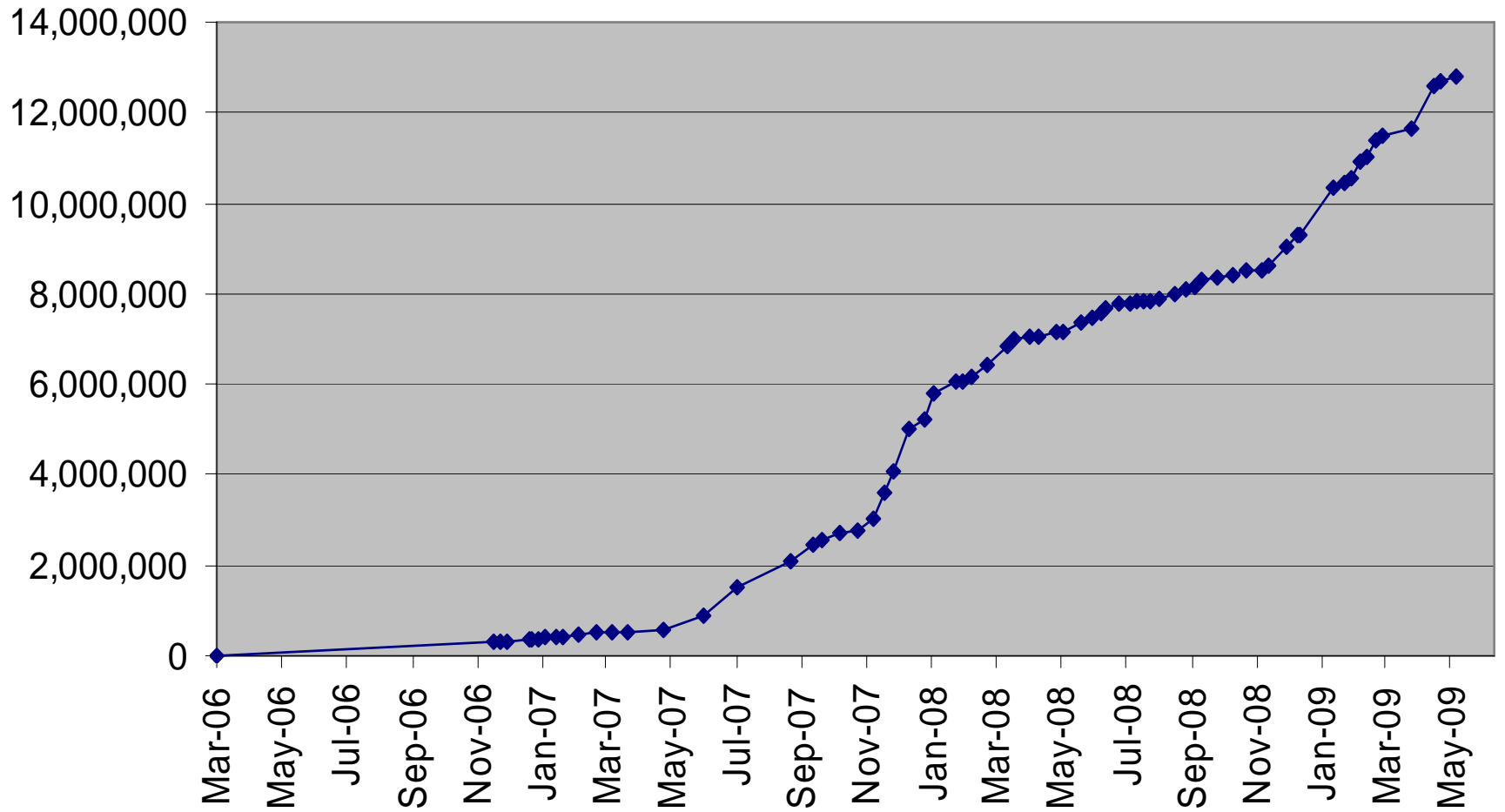


## Portico At A Glance: The State of the Archive

- “Archival Units” (articles) preserved: 15.7 million
- Files preserved: 151.8 million
- Titles represented in the archive: 7,154
- Trigger events supported: 3
- Perpetual access claims supported: 1
- Audit:
  - Participants: 4,605 test accesses of articles, TOCs, or searches, etc.
  - CRL: 2006 test audit via TRAC; 2009 first formal TRAC audit underway
  - Portico: self-assessment via DRAMBORA scheduled for Q3 2009



### Articles In the Archive



## Looking Forward: New Areas for Collaboration

- Collection management & digital preservation
- Preservation priorities beyond e-journals
- Beyond audit: Completeness verification
- Collaborative tools development



## Looking Forward: New Areas for Collaboration

- Collection management & digital preservation
  - Libraries and Portico compare holdings to support print collection management and space allocation decisions.
  - Holdings Comparison Reports: compare Portico archive to a library's collection
    - Title (ISSN) committed to the archive
    - Volumes / issues in the archive
    - Publisher supplying content
    - Portico named as perpetual access source
  - Can identify gaps in collections or preservation strategies



## Looking Forward: New Areas for Collaboration

- Preservation priorities: moving beyond e-journals
  - E-books
    - UC (and others) joined Portico-led e-book preservation study in 2007-2008.
    - Found e-books were an emerging priority and preservation technically feasible within Portico's infrastructure
    - In 2009, e-books rapidly rising in preservation priority for both publishers and for libraries
  - Digitized historical collections ("d-collections")
    - Most frequently named preservation priority
    - Typically a "big library" concern; may not cross full library spectrum



## Looking Forward: New Areas for Collaboration

- Beyond audit: completeness verification
  - Core question: Is the preserved journal run complete?
  - Multi-level checks are required.
    - The title level: Are all committed titles listed in the preservation agreement accounted for?
      - Portico verifies.
    - The volume and issue level: Are all expected volumes and issues present?
      - Collaborative verification needed.
    - The article: Are all referenced elements preserved?
      - Portico verifies.



## Looking Forward: New Areas for Collaboration

- Are all expected volumes and issues present?
  - Scale requires automated checks done in cooperation with partners.
  - Reliance on publishers' verification that transmissions are complete may yield circular checks.
  - A single library's holdings or check-in records may be incomplete or "noisy".
  - Comparison of multiple library holdings to archival holdings may yield "best effort" completeness verification.



## Looking Forward: New Areas for Collaboration

- Collaborative tools development: JHOVE2
  - Project with CDL, Stanford, Portico to enhance file format validation tool.
  - Builds on the work of JHOVE, the JSTOR-Harvard Object Validation Environment
  - Portico responsible for the PDF module and gathering input through JHOVE user group meetings.



Questions?

Comments?

Suggestions?



Eileen Fenton  
eileen.fenton@portico.org  
www.portico.org

