

Preservation of
New Content Types
in the
Portico Archive

Evan Owens
VP, Content Management



Preservation of New Content Types

- Portico preservation capabilities
 - Generic infrastructure for ingest
 - Package interpretation and restructuring
 - Apply specific preservation policies (e.g., normalize)
 - Generate preservation metadata
 - High throughput: 1 million files per day and increasing
- Solutions must follow from the nature of content:
 - Significant properties of the digital objects themselves
 - Contextual factors:
 - Content genre, conventions, standards, expectations
 - Circumstance of the creation of the content
- Portico's e-journal approach is a content-type specific preservation strategy
- Preservation Research
 - E-Book Study
 - Digitized Content
 - JHOVE2 Project



Digitized Historical Collections

- Characteristics:
 - Proprietary structures and packaging
 - But much simpler than E-Journals
- Components
 - Scanned Images
 - Pages and snippets
 - Structural Information
 - Sequence of the images
 - Links to/from metadata
 - OCR data
 - Metadata: Item and Collection
 - Formal: MARC, DC, MODS, proprietary
 - Informal: Essays, about pages, etc.
- Preservation Strategy:
 - Document and preserve all digital components
 - Identify, test, validate all content
 - Create DC metadata for archival discovery
 - Defer costly product development until after trigger event
 - E.g., support for snippets, other high value add features



Preservation Strategies for E-Books

- Several Different Challenges
 - But mostly not publisher specific
- Types of E-Books
 - True E-Books ("E-Books")
 - OEB, EPUB and other E-Book reader formats
 - Highly standardized across publishers
 - But standards continue to evolve!
 - Journal Platform E-Books ("J-Books")
 - Same as E-Journals
 - Digitized Books ("D-Books")
 - Same as Digitized Collections
 - HTML Books ("H-Books")
 - Rare; like web harvesting
 - Web-hosted PDF Books ("P-Books")
 - Minor logistical problems with aligning object and metadata
- Preservation Strategy
 - Varies by Type

