Digital Preservation for Library Publishers: Raising Awareness

Heather Staines
Hypothesis

Wendy C. Robertson
University of Iowa

Jeremy Morse
University of Michigan

Copyright © 2018 Heather Staines, Wendy C Robertson, and Jeremy Morse

Comments
Heather Staines presented 'NASIG Digital Preservation Task Force'. Wendy C Robertson presented 'Digital Preservation and Small Library Publishers'. Jeremy Morse presented 'Preservation Policies, And Why Every Publisher Needs One'

Hosted by Iowa Research Online. For more information please contact: lib-ir@uiowa.edu.
Digital Preservation for Library Publishers

Raising Awareness

Heather Staines, Hypothesis
Wendy Robertson, University of Iowa
Jeremy Morse, University of Michigan Publishing

Library Publishing Forum 22 May 2018, Minneapolis, MN
Our Presenters:

Heather Staines, Hypothesis, is Director of Partnerships, working with publishers and platforms to enable annotation.

Wendy Robertson, University of Iowa, is responsible for helping to expand support of Electronic scholarship and e-publishing, including supporting locally published e-journals and managing the institutional repository.

Jeremy Morse, University of Michigan, manages a team of programmers in developing and supporting the software that enables the Library's publishing efforts.
NASIG Digital Preservation Task Force

Raise awareness of digital preservation.

Develop some best practices for the industry.

Marketing digital preservation to a broad range of library administrations and publishers.

Chair: Shannon Regan (NYPL)

James Phillipotts (Oxford University Press), Wendy Robertson (University of Iowa), Heather Staines (hypothes.is), Zach Van Stanley (University of Denver)

Board Liaison: Ted Westervelt
What is Digital Preservation?

“Digital preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary” (definition from the Digital Preservation Coalition). It is a suite of services and an ongoing process.
What you should know about Digital Preservation?

- Main Digital Preservation Services
- Other tips and resources
LOCKSS (Lots of Copies, Keep Stuff Safe)

A **distributed network** to preserve content **as it appears** on the publisher’s site, with regular integrity checks of the data.

Global LOCKSS Network (GLN) preserves content:

- Available online (open access and subscription-only)
- No cost, but limited space.

Private LOCKSS Network (PLN)

- Examples: MetaArchive, PKP
- One type of software or a specific geographic area.
- Any digital content, not only books and journals.
CLOCKSS (Controlled LOCKSS Network)

CLOCKSS is:

- A **dark archive** of scholarly publications
- Preserved using LOCKSS
- A “trigger event” (content not available from any publisher) makes content freely available with a [creative commons license](https://creativecommons.org).
- Publisher cost based on annual journal or ebook revenue.
- Libraries may join as a supporting members
- Membership is not required to access triggered content.
CLOCKSS for Library Publishers

- Digital Commons: Bepress Agreement in place. (32 publishers)
- OJS: No Agreement is necessary. (24 Publishers)
- Plug-ins for both platforms, so minimal set-up.
- Harvesting approach = publishers do not need to do work at their end.
- Other:
  - IUScholarworks from Indiana University Libraries,
  - Divided Society from Linen Hall Library (in Ireland)
  - Utrecht University Library Open Access Journals.
Portico

Portico is a community supported dark archive to ensure electronic content remains accessible for the long term.

- Primary access is a “trigger event.”
- Portico makes content available for use.
- Open content stays open.
- Cost to publishers is based on annual journal or ebook revenue.
- A library that publishes may choose to join both as a library and as a publisher.
Portico: Info for Library Publishers

- OJS: export plugins available for OJS 3.0 and 2.4.1+ (Currently from Github but soon PKP 3.1 plugin, so available through the OJS UI.)
- Digital Commons included in Portico
- Other systems: Portico will work with library/publisher to get the package how they need it to be; Portico can accept the metadata in a spreadsheet (template), then transform to XML
 Consortial Trusted Digital Repositories

Some libraries have partnered to create a Trusted Digital Repository (TDR) meeting specific international standards (ISO 16363).

Currently six have been certified (including CLOCKSS and Portico):

- Canadiana.org
- Chronopolis
- Hathitrust
- Scholars Portal (a consortia of 21 university libraries in Ontario, Canada)
Other tips and resources:

Keepers Registry

Index of journals preserved by one or more archiving agencies.

- Librarians can check that titles in their collection have been preserved.
- As a publisher, check that your titles are correctly listed with preservation coverage as you expect.

For additional information, see our Guide to the Keepers Registry.
Other tips and resources (continued):

Library of Congress’ Recommended Formats Statement

Provides good guidance for publications in all formats to make sure they will last through time.

Digital Preservation Handbook (by Digital Preservation Coalition)

An internationally authoritative and practical guide for:

- Managing digital resources over time
- Issues in sustaining access to them.
Links to Digital Preservation Task Force Documents

- [Digital Preservation 101](#)
- [Guide to The Keeper's Registry](#)
- [Talking Points and Questions to ask Publishers about Digital Preservation](#)
Big University, Small Publisher

"Round Tuit" by Denise Mattox, https://www.flickr.com/photos/denisemattox/3382089636
https://creativecommons.org/licenses/by-nd/2.0/
NASIGuide: Digital Preservation 101

Prepared by the Digital Preservation Task Force

PDF

Now that scholarly publishing has transitioned to the web, it is imperative that we ensure the content lasts through time. If a publisher discontinues a title or removes content from their site, it could become unavailable to future scholars. This breaks the citation chain so others cannot verify cited works, thereby putting much of the scholarly system at risk. Libraries have long preserved scholarly output and continue to do so today. However, there are steps publishers can and should take to make sure the content they publish remains available.

“Digital preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary” (definition from the Digital Preservation Coalition). It is a suite of services and an ongoing process.

Not Preservation

Backups
Make sure you are following standard best practices for backing up your data. There should be at least three copies of the files, preferably on at least two different types of media, with at least one copy off-site (i.e. in different physical locations) or cloud-based.
Replied to @Alex_Parker @elakdawalla

MER lost files necessary to use the MER test bed. Then someone realized we had a backup. On Opportunity. On Mars. Suck it, Dropbox

8:10 PM - 30 Apr 2018 from Alhambra, CA

428 Retweets 917 Likes
## Archive-It (Internet Archive)

<table>
<thead>
<tr>
<th>Found Captures between Oct 2, 2010 - Jan 6, 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 page</td>
</tr>
</tbody>
</table>

- 2009
  - 1 page
  - July 7, 2009 *

* denotes when page was updated.
## U Iowa Titles in Global LOCKSS

<table>
<thead>
<tr>
<th>Publisher</th>
<th>Title</th>
<th>ISSN</th>
<th>eISSN</th>
<th>Present</th>
<th>Pres17</th>
<th>In prog</th>
<th>In prog</th>
<th>5 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Iowa</td>
<td>Annals at Iowa</td>
<td>0063-4827</td>
<td></td>
<td>64; 65; 66; 2005; 2006</td>
<td>75</td>
<td>2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Iowa</td>
<td>B Sides</td>
<td>2155-5687</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Iowa</td>
<td>Iowa Geological Survey Annual Report</td>
<td>2160-5270</td>
<td>2160-5300</td>
<td></td>
<td>38</td>
<td>1941</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Iowa</td>
<td>Iowa Historical Review</td>
<td>2373-1842</td>
<td></td>
<td>1; 2; 3; 4; 5</td>
<td>2007; 2011</td>
<td>7</td>
<td>2017</td>
<td></td>
</tr>
<tr>
<td>University of Iowa</td>
<td>Mathal/Mashaal: Journal of Islamic and Judaic</td>
<td>2168-538X</td>
<td></td>
<td>1; 2; 3; 4; 5</td>
<td>2011; 2012; 2013; 2015; 2016</td>
<td>2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Iowa</td>
<td>Medieval Feminist Forum</td>
<td>1536-8742</td>
<td>2151-6073</td>
<td>1; 2; 3; 4; 5</td>
<td>1980; 1980</td>
<td>53</td>
<td>2017</td>
<td></td>
</tr>
<tr>
<td>University of Iowa</td>
<td>Poroi: An Interdisciplinary Journal of Rhetoric</td>
<td>2151-2957</td>
<td></td>
<td>1; 2; 3; 4; 5</td>
<td>2001; 2003</td>
<td>13</td>
<td>2017</td>
<td></td>
</tr>
<tr>
<td>University of Iowa</td>
<td>Proceedings in Obstetrics and Gynecology</td>
<td>2154-4751</td>
<td></td>
<td>1; 2; 3; 4</td>
<td>2010; 2011</td>
<td>7</td>
<td>2017</td>
<td></td>
</tr>
<tr>
<td>University of Iowa</td>
<td>Walt Whitman Quarterly Review</td>
<td>0737-0679</td>
<td>2153-3695</td>
<td>1; 2; 3; 4; 5</td>
<td>1983; 1984; 35</td>
<td>2016; 2017</td>
<td>2017</td>
<td></td>
</tr>
</tbody>
</table>

[Link to CSV file: http://www.lockss.org/keepers/keepers-LOCKSS-report.csv]
<table>
<thead>
<tr>
<th>Title</th>
<th>ISSN</th>
<th>Publisher</th>
<th>Current extent of archiving</th>
<th>Archiving agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Annual report for ... with accompanying papers</td>
<td>2160-5270 (Print); 2160-5500 (Online)</td>
<td>University of Iowa</td>
<td>In Progress: 38</td>
<td>Global LOCKSS Network</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Published for the Iowa Geological Survey</td>
<td>Archived: v.14; v.12 (1901); v.16 (1905); v.11 (1900); v.7 (1896); v.8 (1897); v.4; v.2; v.3 (1893); v.4 (1894); v.5 (1895); v.6 (1896); v.7; v.8 (1897); v.9 (1898); v.10 (1899); v.11 (1900); v.12 (1901); v.13 (1902); v.14 Sup. (1903); v.15 (1904); v.16 (1905); v.17 (1906); v.19 (1908); v.20 (1909); v.21 (1910-1911); v.22 Biblio. (1912); v.23 (1912); v.24 (1913); v.25 (1914); v.26 (1915); v.27 (1916); v.28 (1917-1918); v.29 (1919-1920); v.30 (1921-1922); v.18 (1905); v.25 1914; v.26 1912; v.12 1902; v.11 1901; v.3 1895; v.2 1893; v.21 1911; 1917-1918; v.1 1902; v.3 1893; v.4 1894; v.5 1895; v.6 1896; v.7 1899; v.8 1897; v.9 1898; v.10 1899; v.11 1900; v.12 1901; v.13 1902; v.14 1903; v.15 1904; v.16 1905; v.17 1906; v.18 1907; v.24 1913; v.26 1915; v.27 1916; v.28 1917/18; v.29 1919/20; v.30 1921/22; v.19 1908; v.20 1909; v.21 1910/11; v.22 1912</td>
<td>HathiTrust</td>
</tr>
</tbody>
</table>
Joining Portico as a Publisher

### Fees

**E-journals**

Annual fees for e-journals are based on total journal revenues (print and electronic subscriptions, licensing, and advertising, as well as article processing charges and other revenue streams for OA content) as follows:

<table>
<thead>
<tr>
<th>Annual Journal Revenue</th>
<th>Annual Archive Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;$200 million</td>
<td>$81,960</td>
</tr>
<tr>
<td>$100-200 million</td>
<td>$54,635</td>
</tr>
<tr>
<td>$50-100 million</td>
<td>$27,320</td>
</tr>
<tr>
<td>$10-50 million</td>
<td>$16,440</td>
</tr>
<tr>
<td>$5-10 million</td>
<td>$5,460</td>
</tr>
<tr>
<td>$1-5 million</td>
<td>$2,735</td>
</tr>
<tr>
<td>$500k-1 million</td>
<td>$1,085</td>
</tr>
<tr>
<td>$250-500k</td>
<td>$500</td>
</tr>
<tr>
<td>$0-250k</td>
<td>$250</td>
</tr>
</tbody>
</table>
# Portico Required Metadata

<table>
<thead>
<tr>
<th>Journal Title</th>
<th>Publication Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal P-ISSN (if available)</td>
<td>Volume</td>
</tr>
<tr>
<td>Journal E-ISSN</td>
<td>Issue</td>
</tr>
<tr>
<td>Article ID (DOI if available)</td>
<td>Supplement (if relevant)</td>
</tr>
<tr>
<td>PDF File Name</td>
<td>First Page</td>
</tr>
<tr>
<td>Article Title</td>
<td>Last Page</td>
</tr>
<tr>
<td>Authors</td>
<td>Copyright Statement</td>
</tr>
</tbody>
</table>
Review your files to make sure all is well
UTF-8 Disguised as iso-8859-1

<?xml version='1.0' encoding='iso-8859-1' ?>
<documents>
<document>
<title>The Moroccan Jurist al-Khamlīshī: Can a Woman Become a Legislator (Mujtahid)?</title>
</document>
</documents>

<?xml version='1.0' encoding='utf-8' ?>
<documents>
<document>
<title>The Moroccan Jurist al-Khamlīshī: Can a Woman Become a Legislator (Mujtahid)?</title>
</document>
</documents>

<keyword>ijtihād</keyword>  <keyword>ijtihād</keyword>
<keyword>fiqh</keyword>     <keyword>fiqh</keyword>
<keyword>madhhab</keyword>  <keyword>madhhab</keyword>
<keyword>'ulamāʾ</keyword>  <keyword>'ulamāʾ</keyword>
<keyword>ijmāʾ</keyword>    <keyword>ijmāʾ</keyword>
<keyword>sharīʿa</keyword>  <keyword>sharīʿa</keyword>
Privacy

An Ethical Framework for Library Publishing

Version 0.5 (Draft for Comment)

Editors

Jason Boczar, University of South Florida (Ethical Framework Task Force chair); Charlotte Roh, University of San Francisco (2017–18 LPC Fellow); Melanie Schlosser, Library Publishing Coalition
Journal Publishing Tag Library

July 2017

National Center for Biotechnology Information (NCBI)
National Library of Medicine (NLM)

Developer:
Mulberry Technologies, Inc.
17 West Jefferson Street, Suite 207
Rockville, MD 20850
Phone: 301/315-9631
Fax: 301/315-8285

<article xmlns:ali="http://www.niso.org/schemas/all/1.0/" dtd-version="1.2di" xml:lang="en"
xmlns:oasis="http://www.niso.org/standards/z39-96/ns/oasis-exchange/table">
  <front>
    <journal-meta>
      <!-Not in our metadata - add if necessary -->
      <!- <journal-id journal-id-type="pmc"/> <journal-id journal-id-type="pubmed"/>-->
      <journal-id journal-id-type="publisher">
        <xsl:value-of select="substring-before(submission-path, '/')"/>
      </journal-id>
    </journal-meta>

    <!-- if multiple titles in same series or subtitle, select pub_title lines -->
    <journal-title-group>
      <journal-title><xsl:value-of select="normalize-space(publication-title)"/>
      <xsl:variable name="jtitle">
        <xsl:value-of
          select="normalize-space(fields/field[@name = 'pub_title'])"
        />
      </xsl:variable>
      <journal-title><xsl:value-of select="substring-before($jtitle, ': ')"/>
      <journal-subtitle><xsl:value-of select="substring-after($jtitle, ': ')"/>
      <xsl:variable name="jtitlegroup">
        <!--If the URL is not an abbreviated title, comment this out-->
        <!--<xsl:variable name="jtitlegroup">
          <xsl:value-of select="upper-case(substring-before(submission-path,...
      </journal-subtitle>
      </journal-title-group>
    </front>
</article>
The full text is freely available immediately; there is no embargo period. Authors will receive monthly emails providing the count of downloads of their articles. There are no article submission charges nor article processing charges (APC) to publish.

**Open Access Statement:** This is an open access journal which means that all content is freely available without charge to the user or their institution. The journal allows readers to read, download, print, search, or link to the full texts of its articles, without asking prior permission from the publisher or the author. Any other use, including distributing non-commercially, must be approved by the author.

*Iowa Historical Review* is preserved in the [Global LOCKSS Network](http://ir.uiowa.edu/iowa-historical-review/about.html).

Complete metadata for all articles in *Iowa Historical Review* is available via OAI [http://ir.uiowa.edu/do/oai/?verb=ListRecords&metadataPrefix=gdc&set=publication:iowa-historical-review](http://ir.uiowa.edu/do/oai/?verb=ListRecords&metadataPrefix=gdc&set=publication:iowa-historical-review). See our policy regarding use and reuse of our metadata.

*Iowa Historical Review* is hosted by the [University of Iowa Libraries' Institutional Repository](http://ir.uiowa.edu/iowa-historical-review/about.html) (Iowa Research Online).
Future
Library Publishing Curriculum

Sustainability Module:

The Sustainability Module addresses core concerns related to library publishing as a socio-technical system. In order for library publishing endeavors to establish longevity and long-term success, professionals need to address a range of concerns related to people, policies, communities, technologies, and financials.

Authors: Laurie N. Taylor, Digital Scholarship Librarian at the University of Florida; Brian W. Keith, Associate Dean for Administration and Faculty Affairs at the George A. Smathers Libraries, University of Florida

https://drive.google.com/drive/folders/1uNMuDDQPe6rF20cmue2U6dTzsCGTKtXW
Preservation Policies
And Why Every Publisher Needs One

Jeremy Morse
Michigan Publishing
Library Publishing Forum 2018
Preservation

- It’s not just a technology issue
  - ... though we tend to talk about it that way
- It’s a commitment of resources over the long term
  - Time
  - Attention
  - Active management
- It’s a series of decisions
Preservation Strategies

Each of these suggests a different level of commitment

… and might vary by content type, or site by site.

- Bitwise
- Migration
- Emulation
Bitwise Preservation

- Preserving the data exactly as it is
- No promise of future usability or viability
- “Garbage in, garbage out”
Migration-based Preservation

- Reformat the content over time
- Guards against obsolescence
- Easiest with open standards
- Targets discrete components (i.e. individual files)
Emulation-based Preservation

- Preserving the software that serves the content
  - So the content can continue to be used “as is”
- Put it in some kind of wrapper that then gets migrated
  - Emulation
  - Virtualization
  - Containerization
Other Preservation Strategies

- Static snapshots
  - Lossy migration, or cheap emulation?

- Distribution
  - Geographic
  - Institutional
We usually talk about *how we preserve stuff*...
Have we thought about *what* exactly we’re preserving?
Preservation Policy

Making deliberate, transparent choices about what aspects of a publication you’re preserving…

… and what aspects you’re not.

Make sure our choices are explicit,

… not implicit and, perhaps, unintended.
“The Version of Record”

What constitutes it? Whatever contributes directly to the scholarly argument

- **Text**
- **Embedded assets**
  - figures
  - tables
  - maybe video, etc.
- **Metadata**
  - For the work as a whole
  - For the components
- **Relationships among the parts**
  - Structural metadata
  - Linked data
Supplemental Materials

- Research datasets
- Supporting documents compiled during research
- Material that was cut for length
  - Additional figures
  - Longer versions of audio, video
  - Additional appendices
- Author annotations
  - Branded hypothes.is layer?
The Engagement Layer

aka AboutWare, Brochureware

- Things that might be in a print journal’s frontmatter
  - Editorial Board
  - Submission guidelines
  - Etc.
- Public comments and annotation
- Open review notes
- Marketing and promotional materials
The Application Layer

- “Look and feel”
  - For more dynamic content
  - Design changes over time
- Navigation elements
- The software itself
Can we preserve it all?

Probably not!

This is expensive, and resources are finite.
What we preserve, and what we don’t, should be a matter of choice, not happenstance.
Preservation Statement

● To answer the questions:
  ○ What will future scholars find when they follow a citation?
  ○ How faithful will a future version of this publication be?

● In the interest of
  ○ Transparency
    ■ Akin to a Privacy Statement or Accessibility Statement
  ○ Trustworthiness
    ■ Our constituencies expect libraries to take care of this problem!
Preservation Statement should say...

- What we’re preserving
  - In other words, what we're contributing to the Permanent Scholarly Record
- What we’re not preserving
  - The things we consider ephemeral
- What level of commitment we’re making
  - Preservation level (bitwise, migration, etc.)
    - Maybe we just take a snapshot of the look & feel once a year?
  - Period of commitment
    - And a stated period for revising the policy

I’ll be posting ours on www.fulcrum.org

I’d like to see yours!