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

Library Publishing Coalition Forum
May 22, 2018

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 Phillpotts (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**.
- 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, heck 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

Digital Preservation and Small Library Publishers

Wendy C. Robertson
University of Iowa
Library Publishing Forum 22 May 2018, Minneapolis, MN

Big University, Small Publisher

"Pentacrest, The University of Iowa, 1958", http://digital.lib.uiowa.edu/cdm/ref/collection/ictcs/id/100

Publications

- Conference Proceedings
- Core Competencies
- NASIG Newsletter
- NASIGuides
- NASIG Blog
- NASIG Jobs Blog

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.

Doug Ellison
@doug_ellison

Follow

Replying 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

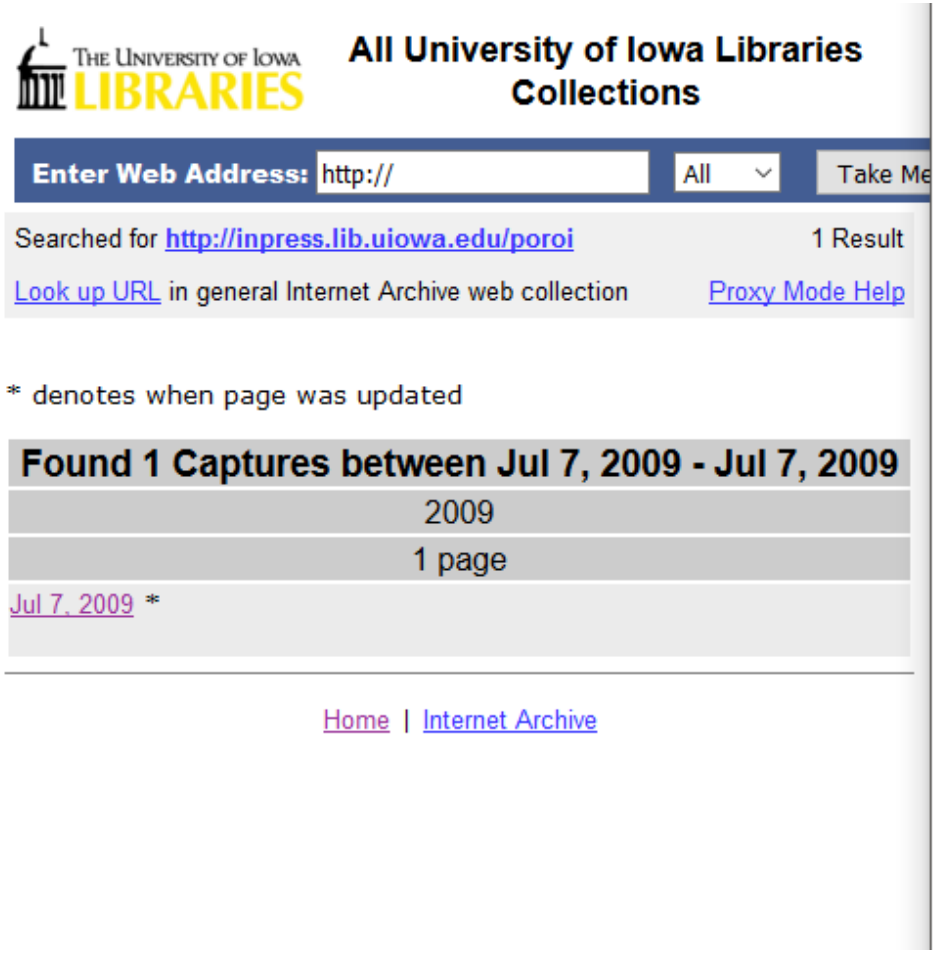
8:10 PM - 30 Apr 2018 from Alhambra, CA

428 Retweets **917** Likes

16 428 917

https://twitter.com/doug_ellison/status/991152866554400768?s=03

Archive-It (Internet Archive)



The University of Iowa LIBRARIES

All University of Iowa Libraries Collections

Enter Web Address: http:// All Take Me

Searched for http://inpress.lib.uiowa.edu/poroi 1 Result

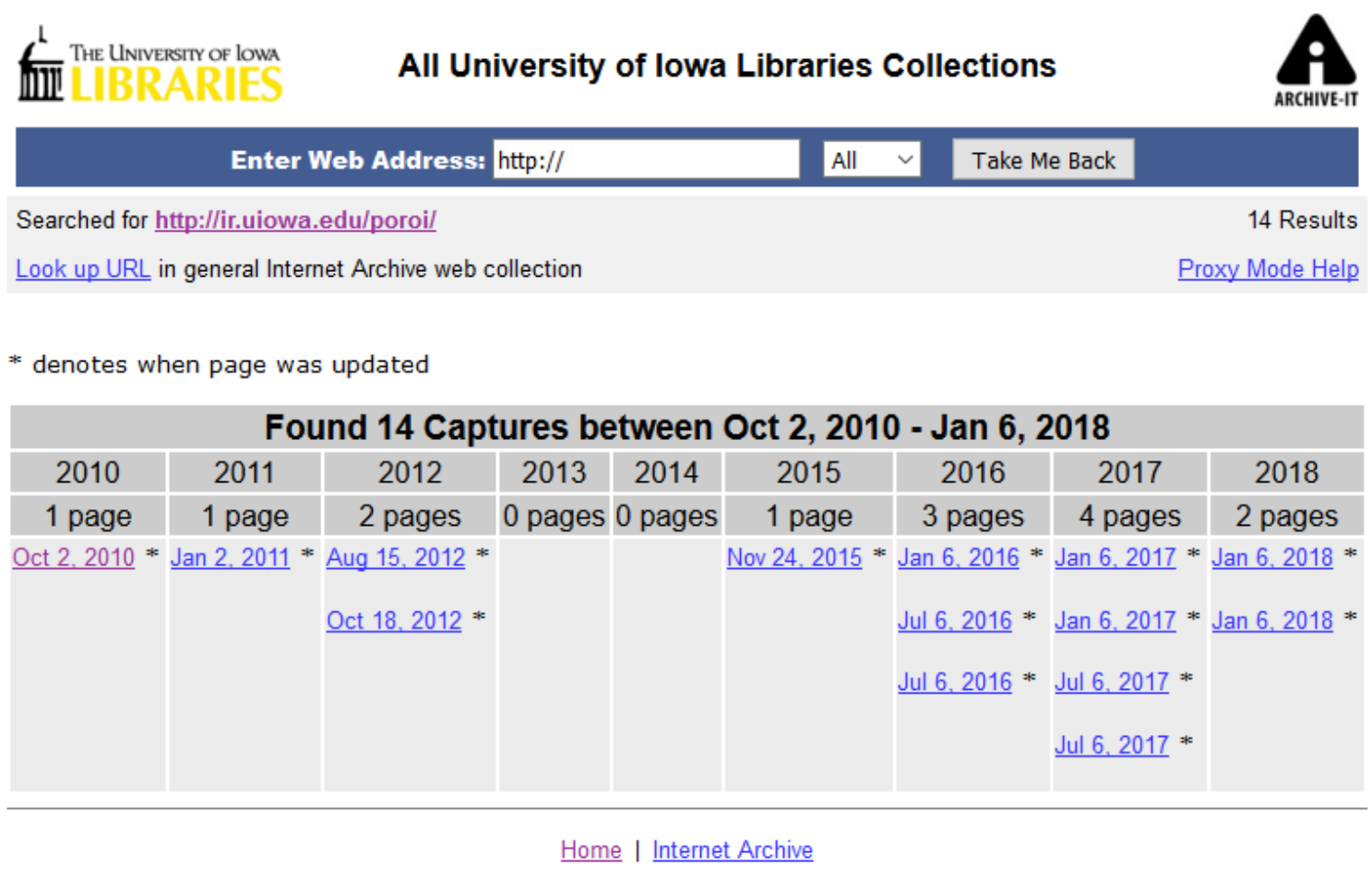
Look up URL in general Internet Archive web collection Proxy Mode Help

* denotes when page was updated

Found 1 Captures between Jul 7, 2009 - Jul 7, 2009

2009
1 page
Jul 7, 2009 *

Home | Internet Archive



The University of Iowa LIBRARIES

All University of Iowa Libraries Collections

ARCHIVE-IT

Enter Web Address: http:// All Take Me Back

Searched for http://ir.uiowa.edu/poroi/ 14 Results

Look up URL in general Internet Archive web collection Proxy Mode Help

* denotes when page was updated

Found 14 Captures between Oct 2, 2010 - Jan 6, 2018

2010	2011	2012	2013	2014	2015	2016	2017	2018
1 page	1 page	2 pages	0 pages	0 pages	1 page	3 pages	4 pages	2 pages
Oct 2, 2010 *	Jan 2, 2011 *	Aug 15, 2012 *			Nov 24, 2015 *	Jan 6, 2016 *	Jan 6, 2017 *	Jan 6, 2018 *
		Oct 18, 2012 *				Jul 6, 2016 *	Jan 6, 2017 *	Jan 6, 2018 *
						Jul 6, 2016 *	Jul 6, 2017 *	
							Jul 6, 2017 *	

Home | Internet Archive

https://wayback.archive-it.org/org-120/*/http://inpress.lib.uiowa.edu/poroi
https://wayback.archive-it.org/org-120/*/http://ir.uiowa.edu/poroi/

Ulowa Titles in Global LOCKSS

1	Publisher	Title	ISSN	eISSN	Preserv	Preserv	In Prog	In Prog	s Years
12648	University of Iowa	Annals at Iowa	0003-4827		64; 65; 66;	2005; 2006	76	2017	
12649	University of Iowa	B Sides	2155-6687						
12650	University of Iowa	Books at Iowa	0006-7474	2378-4830			50; 51; 52;	1989; 1989; 1990; 1990; 1991; 1991; 1992; 1992; 1993; 1993; 1994; 1994; 1995; 1995; 1	
12651	University of Iowa	Dada/Surrealism	0084-9537	2372-6725	19; 20	2013; 2015	1; 2; 3; 4; 5	1971; 1972; 1973; 1974; 1975; 1976; 1977; 1978; 1979; 1982; 1983; 1984; 1985; 1986; 1	
12652	University of Iowa	Electronic Journal of Africana Bibliography	1092-9576		1; 2; 3; 4; 5	1997; 1997; 1997; 1999; 1999; 2000; 2001; 2002; 2004; 2006; 2009; 2009; 2012; 2014; 2015; 2016			
12653	University of Iowa	Iowa Geological Survey Annual Report	2160-5270	2160-5300			38	1941	
12654	University of Iowa	Iowa Historical Review	2373-1842		1; 2; 3; 4; 5	2007; 2011	7	2017	
12655	University of Iowa	Iowa Journal of Cultural Studies	2168-569X	2168-5738	1993; 1994	1993; 1994	17	2016	
12656	University of Iowa	Iowa Journal of Literary Studies	0743-2747	2325-8845	1; 2; 3; 4; 5	1977; 1978; 1981; 1983; 1984; 1985; 1986; 1987; 1988; 1989; 1991			
12657	University of Iowa	Marilyn Zurmuehlen Working Papers in Art E	2326-7070	2326-7062	2005; 2006	2005; 2006	2017	2017	
12658	University of Iowa	Mathal/Mashal: Journal of Islamic and Judaic	2168-538X		1; 2; 3; 4; 5	2011; 2012; 2013; 2015; 2016			
12659	University of Iowa	Medieval Feminist Forum	1536-8742	2151-6073	1; 2; 3; 4; 5	1986; 1986	53	2017	
12660	University of Iowa	Poroi: An Interdisciplinary Journal of Rhetori	2151-2957		1; 2; 3; 4; 5	2001; 2003	13	2017	
12661	University of Iowa	Proceedings in Obstetrics and Gynecology	2154-4751		1; 2; 3; 4; 5	2010; 2011	7	2017	
12662	University of Iowa	Syllecta Classica	1040-3612	2160-5157	1; 2; 3; 4; 5	1989; 1990; 1992; 199	1989; 1990; 1992; 1993; 1994; 1996; 1996; 1997; 1998; 1999; 2000; 2001; 2002; 2003; 2		
12663	University of Iowa	Walt Whitman Quarterly Review	0737-0679	2153-3695	1; 2; 3; 4; 5	1983; 1984	34; 35	2016; 2017	

http://www.lockss.org/keepers/keepers-LOCKSS-report.csv

Search results

2160-5270 Search

1 hit found
Current search: **2160-5270**

See more results: « »

	Title	ISSN	Publisher	Current extent of archiving	Archiving agency
1	Annual report for ... with accompanying papers ***Variant title:*** *Annual report ... with accompanying papers*	2160-5270 (Print); 2160-5300 (Online)	University of Iowa	In Progress: 38	Global LOCKSS Network
			Published for the Iowa Geological Survey	Archived: v.14; v.1; v.12 (1901); v.16 (1905); v.11 (1900); v.7 (1896); v.8 (1897); V4; 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 (1908); v.25 1914; v.23 1912; v.12 1902; v.11 1901; v.3 1895; v.2 1893; v.21 1911; 1917-1918; v.1 1892; v.3 1893; v.4 1894; v.5 1895; v.6 1896; v.7 1896; 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	HathiTrust

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:

Annual journals revenue	Annual archive contribution
>$200 million	$81,960
$100-200 million	$54,635
$50-100 million	$27,320
$10-50 million	$16,440
$5-10 million	$5,460
$1-5 million	$2,735
$500k-1 million	$1,085
$250-500k	$500
$0-250k	$250

https://www.portico.org/join/how-publishers-can-join/

Portico Required Metadata

- Journal Title
- Journal P-ISSN (if available)
- Journal E-ISSN
- Article ID (DOI if available)
- PDF File Name
- Article Title
- Authors
- Publication Date
- Volume
- Issue
- Supplement (if relevant)
- First Page
- Last Page
- Copyright Statement

The Library of Congress in 1902, public domain
https://en.wikipedia.org/wiki/Library_of_Congress#/media/File:The_Library_of_Congress,_Washington-LCCN2008678216.jpg

Tulips, 2 May 2018, Wendy Robertson, CC-BY

Review your files to make sure all is well

Keep Calm and Carry On poster, 1939, public domain, wartimeposters.co.uk

UTF-8 Disguised as iso-8859-1

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

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

```
<keyword>ijtihÄ□d</keyword>
<keyword>fiqh</keyword>
<keyword>madhhab</keyword>
<keyword>Ê¿ulamÄ□Ê¾</keyword>
<keyword>ijmÄ□Ê¿</keyword>
<keyword>sharÄ«Ê¿a.</keyword>
```

```
<keyword>ijtihād</keyword>
<keyword>fiqh</keyword>
<keyword>madhhab</keyword>
<keyword>ʿulamāʾ</keyword>
<keyword>ijmāʿ</keyword>
<keyword>sharīʿa.</keyword>
```

Privacy

An Ethical Framework for Library Publishing

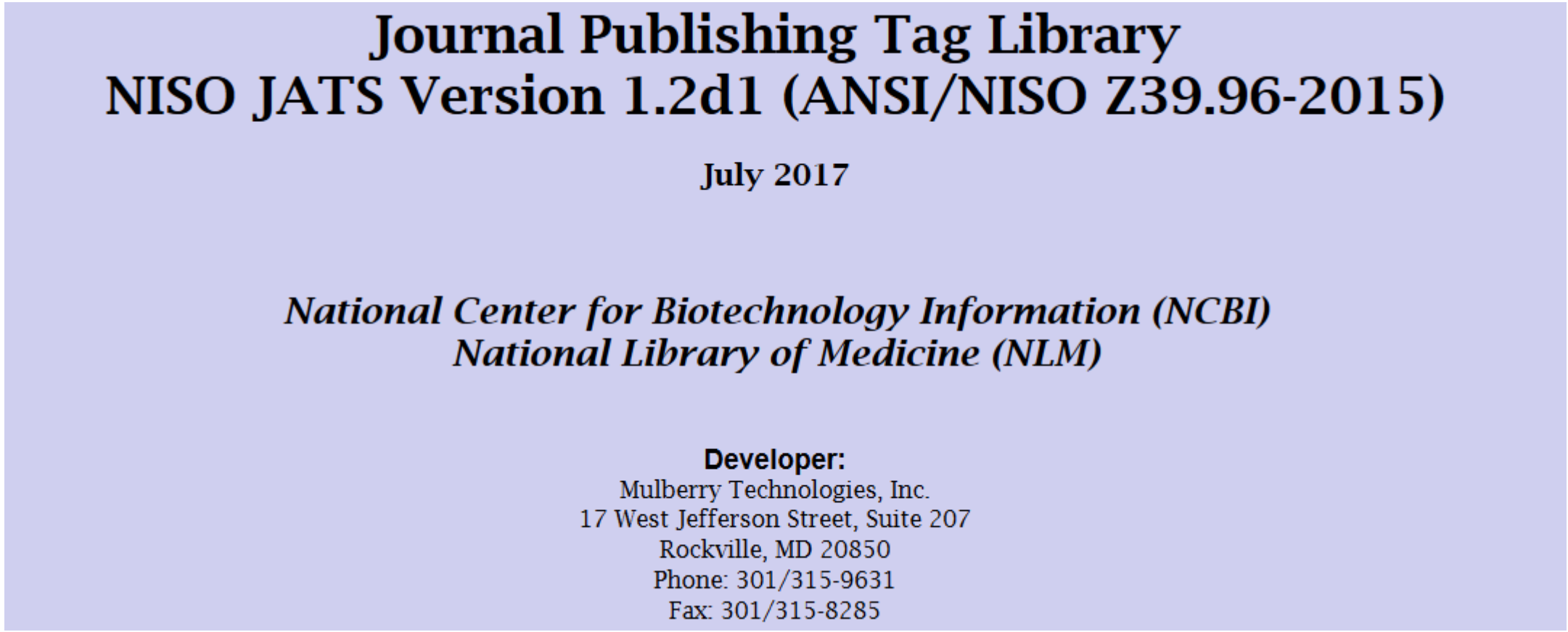
Version 0.5 (Draft for Comment)

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

https://docs.google.com/document/d/1BSRVblTZeRLaaxHl58aeuD0Dqs_imJTP2MsTyhum8HA/edit?usp=sharing

JATS



Journal Publishing Tag Library
NISO JATS Version 1.2d1 (ANSI/NISO Z39.96-2015)

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

https://jats.nlm.nih.gov/publishing/tag-library/1.2d1/index.html

```
<article xmlns:ali="http://www.niso.org/schemas/ali/1.0/" dtd-version="1.2d1" xml:lang="en"
    xmlns:mml="http://www.w3.org/1998/Math/MathML" xmlns:xlink="http://www.w3.org/1999/xlink"
    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 journal-id-type="pubmed"></journal-id>-->
            <journal-id journal-id-type="publisher">
                <xsl:value-of select="substring-before(submission-path, '/')"/>
            </journal-id>

            <!-- 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)"/></journal-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($jtitle,': ')"/></journal-title>
                <journal-subtitle><xsl:value-of select="substring-after($jtitle,': ')"/></journal-subtitle>-->

                <!--If the URL is not an abbreviated title, comment this out-->
                <!--<abbrev-journal-title abbrev-type="publisher"><xsl:value-of select="upper-case(substring-before(submission-path
            </journal-title-group>
```

https://github.com/wendycr/digital-commons/blob/master/digital-commons-backup2JATS.xsl

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.

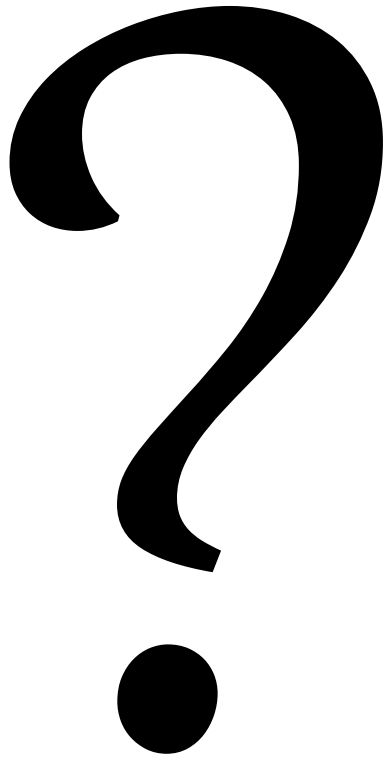
Open Access Statement: This is an open access journal which means that all content is freely available without charge to the user or their institution. The journal allows readers to read, download, print, search, or link to the full texts of its articles, without asking prior permission from the publisher 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.

Complete metadata for all articles in *Iowa Historical Review* is available via OAI http://ir.uiowa.edu/do/oai/?verb=ListRecords&metadataPrefix=qdc&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 (Iowa Research Online).

Future



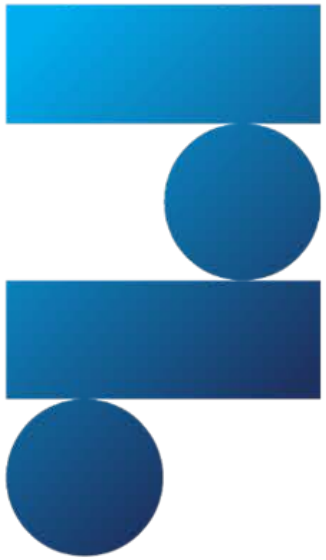
Library Publishing Curriculum

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/1uNMuDDQPe6rF20cmue2U6dTzsCGTctXW



fulcrum

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

fulcrum

Preservation Strategies

Each of these suggests a different level of commitment

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

- Bitwise
- Migration
- Emulation

fulcrum

Bitwise Preservation

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

fulcrum

Migration-based Preservation

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

fulcrum

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

fulcrum

Other Preservation Strategies

- Static snapshots
  - Lossy migration, or cheap emulation?
- Distribution
  - Geographic
  - Institutional

fulcrum

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.

fulcrum

“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

fulcrum

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?

fulcrum

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

fulcrum

The Application Layer

- “Look and feel”
  - For more dynamic content
  - Design changes over time
- Navigation elements
- The software itself

fulcrum

Can we preserve it all?

Probably not!

This is expensive, and resources are finite.

fulcrum

What we preserve,
and what we don’t,
should be a matter of choice,
not happenstance.

fulcrum

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!

fulcrum

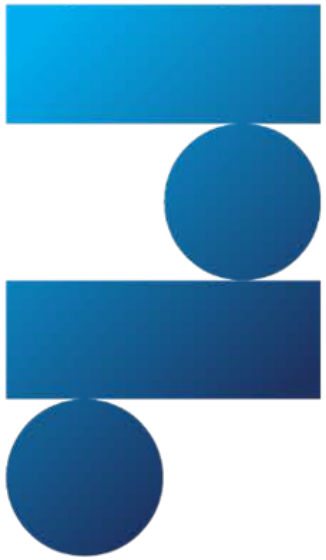
Preservation Statement should say...

- 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!

fulcrum



fulcrum

https://fulcrum.org

@fulcrumpub

fulcrum-info@umich.edu