# Automated creation of analytic catalog records for born-digital journal articles

Kevin S. HAWKINS @KevinSHawkins



Digital Library
Production Service





#### Making of America Journals

aking of America (MOA) is a digital library of primary sources in American social history primarily from the antebellum period through reconstruction. The collection is particularly strong in the subject areas of education, psychology, American history, sociology, religion, and science and technology. The book collection currently contains approximately 8,500 books with 19th century imprints. For more details about the project, see About MoA.

Making of America is made possible by a grant from the Andrew W. Mellon Foundation.

| Search Journals: | Search |
|------------------|--------|
|                  |        |

Browse MoA Journals
 Other Searches in MoA

Current online journal holdings: Pages: 277.603

Journal issues: 2,457

MoA Books Home | Search MoA Books & Journals | UMDL Texts Home

Powered by DLXS

To comment or inquire about content, contact moa-feedback@umich.edu
To report errors, contact UMDL Help



| Home Search       | Browse Bookbag Help   |
|-------------------|---|
| Add to bookbag Se | arch this issue: Go Other search options                                  |
| Title:            | The American Jewess: Vol. 8, No. 5  |
| Publication Info  | Ann Arbor, Michigan: University of Michigan Library<br>2004               |
| Availability:     | These pages may be freely searched and displayed. Permission information. |
| Print source:     | The American Jewess<br>[Chicago: R. Sonneschein], [1895-                  |
| Subject terms:    | Jewish women Periodicals United States                                    |
| URL:              | http://name.umdl.umich.edu/taj1895.0008.005                               |

#### Contents

Front Matter **Advertisements** Yom Kippur Eve In A Jewish Home Before Going To Schul. **Advertisements** Zionism.

Sonneschein, Rosa; pp. 5-9

A Miniam of Januardana



Home Search Browse Bookbag Help

Format image 

Page no. [n/a] 

Page size 100% 

Search this text 

Search

Print

Serial: The American Jewess

Title: Advertisements [Volume 7, Issue: 5, September, 1898]

Collection: American Jewess Table of contents | Add to bookbag

≪ prev next ▶





### digital library eXtension service

#### About DLXS

- Overview
- Prospectus
- · Brochure
- · Members/Contacts
- DLXS examples

#### **Products & Fees**

- · Products
- Supported collections
- · XPAT

#### Technical Info

- · Requirements
- · Environment

#### Support

- · Workshops & meetings
- · Documentation
- DLXS User Group
- DLXS free listserv



inspired









quick search





Home

Search Browse

About Subscription Submission Editors Contact



The Journal of Electronic Publishing (JEP) is a forum for research and discussion about contemporary publishing practices, and the impact of those practices upon users. MORE ABOUT JEP...

Volume 15, No. 1 CURRENT ISSUE

#### Editor's Note

Rebecca Welzenbach and Shana Kimball

#### Refurbishing the Camelot of Scholarship: How to Improve the Digital Contribution of the PDF Research Article

#### John Willinsky, Alex Garnett, and Angela Pan Wong

The Portable Document Format (PDF) has become the standard and preferred form for the digital edition of scholarly journal articles. Originally created as a solution to the need to "view and print anywhere," this technology has steadily evolved since the 1990s. However, its current use among scholarly publishers has been largely restricted to making research articles print-ready, and this greatly limits the potential capacity of the PDF research article to form a greater part of a digital knowledge ecology. While this article considers historical issues of design and format in scholarly publishing, it also takes a very practical approach, providing demonstrations and examples to assist publishers and scholars in finding greater scholarly value in the way the PDF is used for journal articles. This involves but is not limited to graphic design and bibliographic linking, the deployment of metadata and research data, and the ability to combine elements of improved machine and human readability.

#### Social Media: New Editing Tools or Weapons of Mass Distraction?

#### Connect with JEP

Please send comments and questions about the Journal of Electronic Publishing to us at jep-info@umich.edu.

- Follow us on Twitter
- Subscribe to our email list
- Subscribe to our feed

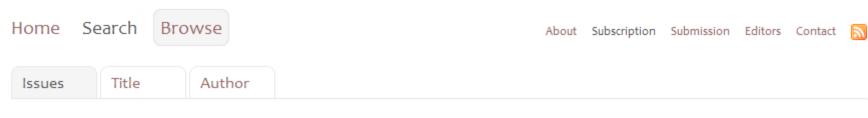
#### Recent Issues

- Volume 15, Issue 1 SUMMER 2012
- Volume 14, Issue 2 FALL 2011
- Volume 14, ISSUE 1 AUGUST 2011
- Volume 13, Issue 3 DECEMBER 2010
- Volume 13, Issue 2 FALL 2010

VIEW ALL ISSUES...

8 Agata Mrva-Montoya





### Browse by Issue

Browse by Issue > Volume 15 > Issue 1 (Summer 2012)

| Title   | Author(s)   |
|---|---|
| Editor's Note [15.1]  | Welzenbach, Rebecca; Kimball, Shana                 |
| Refurbishing the Camelot of Scholarship: How to Improve the Digital Contribution of the PDF Research Article      | Willinsky, John; Garnett, Alex; Pan Wong,<br>Angela |
| Social Media: New Editing Tools or Weapons of Mass Distraction?   | Mrva-Montoya, Agata                                 |
| <u>Publisher-Library Partnership for Accessibility: A Case Study of Scholarly Publishing for Public Audiences</u> | Anderson-Wilk, Mark; Kunda, Sue                     |
| Review of Uncreative Writing: Managing Language in the Digital Age by Kenneth Goldsmith                           | Chesley, Amelia                                     |

TOP OF PAGE ☆



Home Search Browse

About Subscription Submission Editors Contact

## NISO Z39.96 The Journal Article Tag Suite (JATS): What Happened to the NLM DTDs?

Jeffrey Beck

Volume 14, Issue 1, Summer 2011

DOI: http://dx.doi.org/10.3998/3336451.0014.106

<u>Permissions</u>



#### Abstract

In creating PubMed Central (PMC) [1], the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM) needed a common format, with a single Document Type Definition (DTD), for all content in PMC. The first version of the NLM DTD was made available to the public in early 2003, and it quickly became the defacto standard for tagging journal articles in XML even outside the NLM. As usage grew, users and potential users started asking about formalizing the article models as a standard with the National Information Standards Organization (NISO).

Work on the NISO standard began in late 2009, and the Journal Article Tag Suite was released as a Draft Standard for Trial Use as NISO Z39.96 in March 2011.

#### A Short History of the NLM DTD Project

#### PubMed Central and the pmc-1.dtd

PMC is the NLM's digital library of full-text life sciences journal literature. Currently it holds over 2 million articles from more than 250 publishers. Although PMC is also used to store articles based on research funded with NIH grants as part of the NIH Public Access project [2], the original intent of the project was to take full-text article submissions from publishers and make them available through the database. The only technical requirement at the time was that the publisher had to supply the articles in some SGML or XML format and include all images so that the articles could be displayed at PMC.

In early PMC (see Figure 1), the SGML or XML content was loaded into a database and then it was rendered into HTML from the

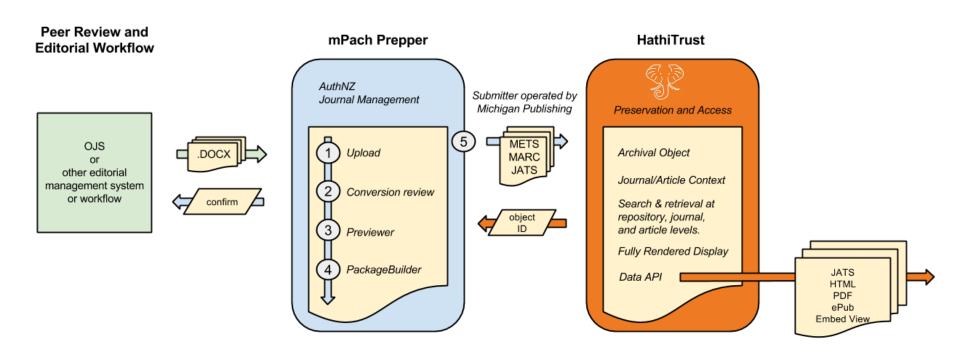
## Opportunities

- HathiTrust
  - offers a better infrastructure for development than DLXS
  - is certified by Trustworthy Repositories Audit & Certification (TRAC)
- There's growing interest among institutions in building a shared infrastructure for publishing.

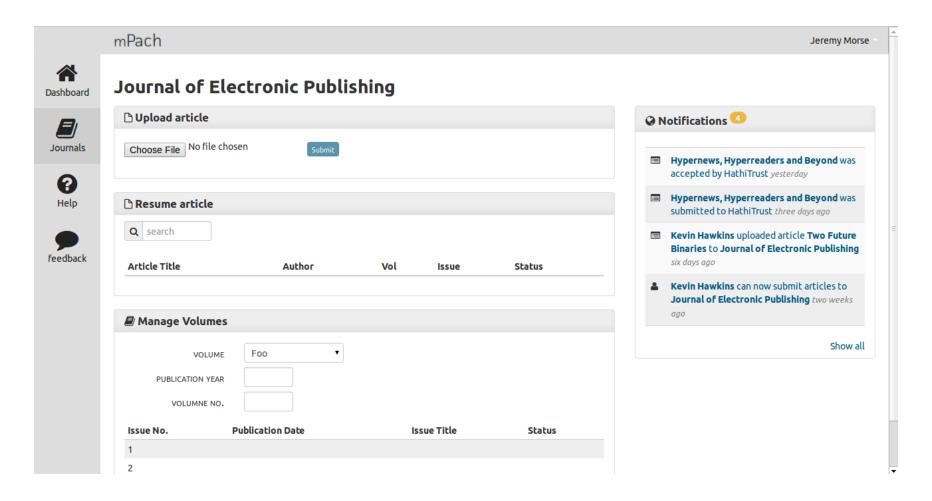
## mPach: what are we creating?

- modular platform
- tightly coupled with the HathiTrust repository
- for open-access journals
- all you need to publish and preserve an OA journal
- will integrate with Open Journal Systems (OJS)

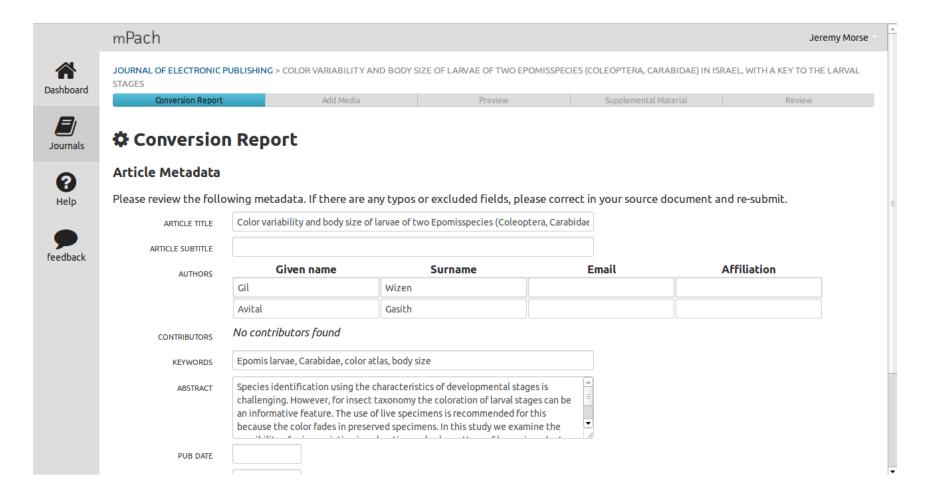
## mPach



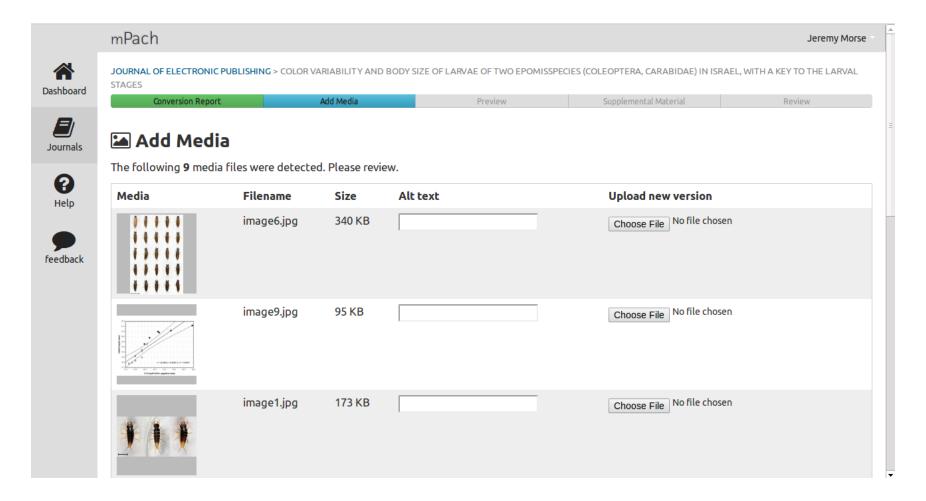
## mPach Prepper (1 of 8)



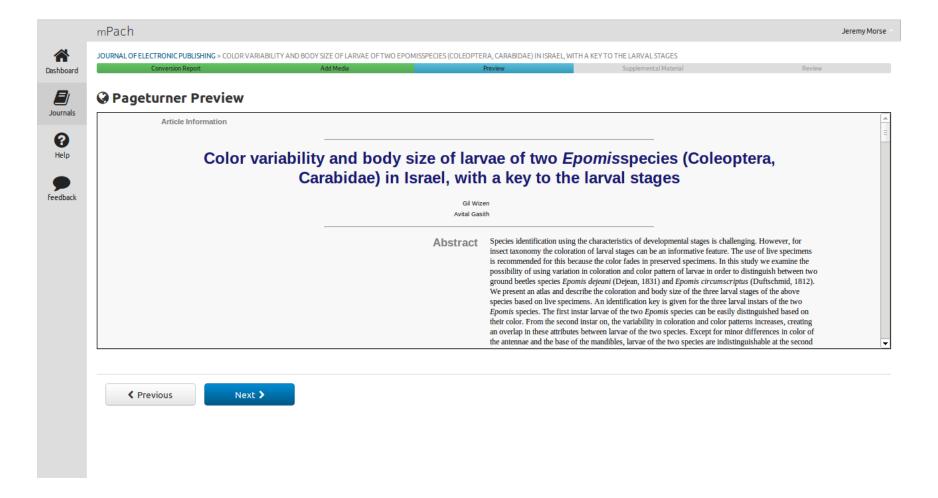
## mPach Prepper (2 of 8)



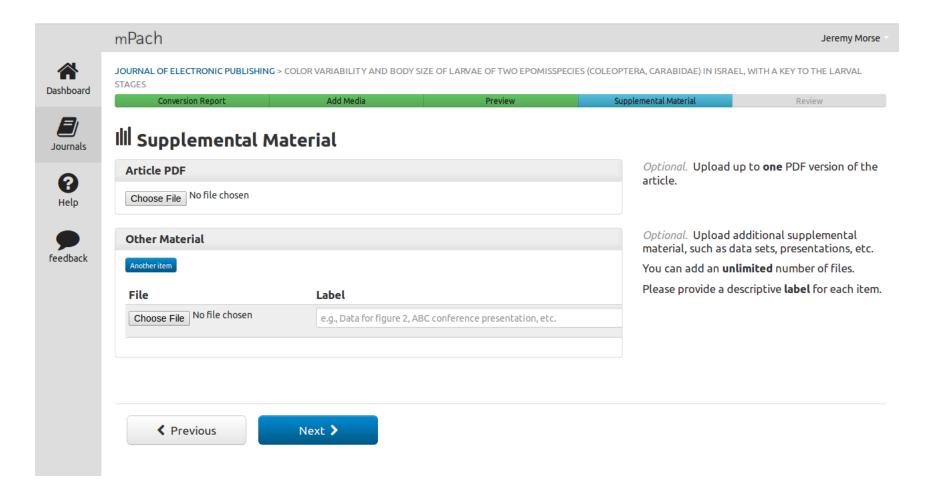
## mPach Prepper (3 of 8)



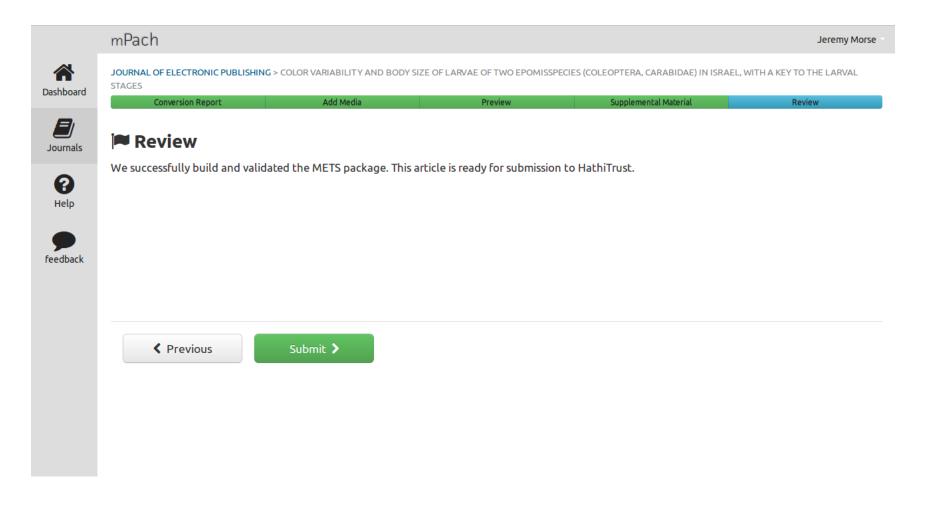
## mPach Prepper (4 of 8)



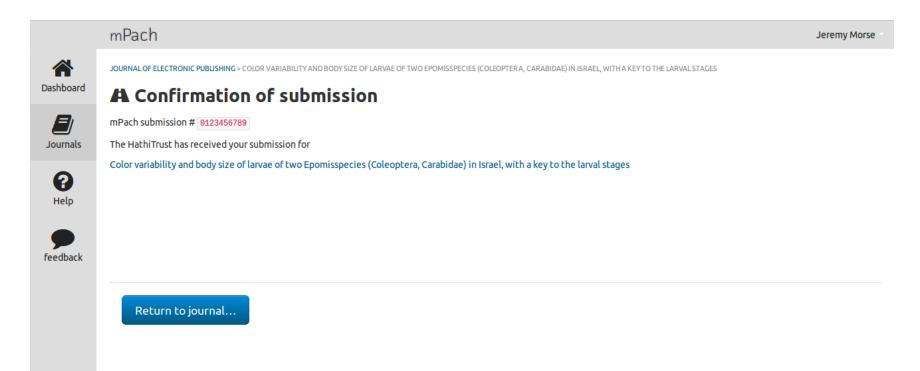
## mPach Prepper (5 of 8)



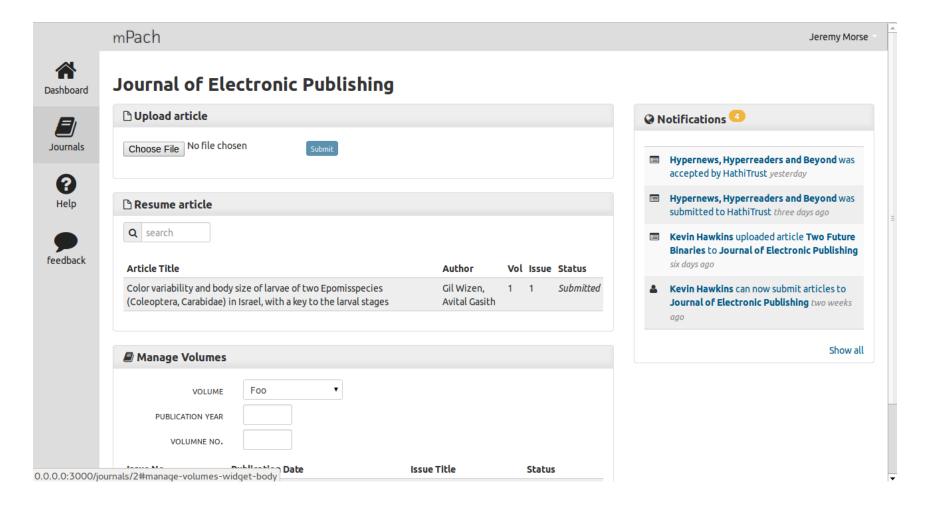
## mPach Prepper (6 of 8)



## mPach Prepper (7 of 8)



## mPach Prepper (8 of 8)







« Back to Journal of Electronic Publishing collection



#### Journal of Electronic **Publishing**

Vol. 15, No. 1 (Summer 2012)

About this journal

#### About this Article

Refurbishing the Camelot of Scholarship: How to Improve the Digital Contribution of the PDF Research Article

John Willinsky Alex Garnett Angela Pan Wong

View full catalog record

Copyright: (cc) BY

#### **Get this Article**

Download (PDF)



### Refurbishing the Camelot of Scholarship: How to Improve the Digital Contribution of the PDF Research Article

:≣+

Search in this text

Find

John Willinsky, Alex Garnett, and Angela Pan Wong

Volume 15, Issue 1, Summer 2012

DOI: http://dx.doi.org/10.3998/3336451.0015.102

(cc) av Permissions

This paper was refereed by the Journal of Electronic Publishing's peer reviewers.

#### Abstract

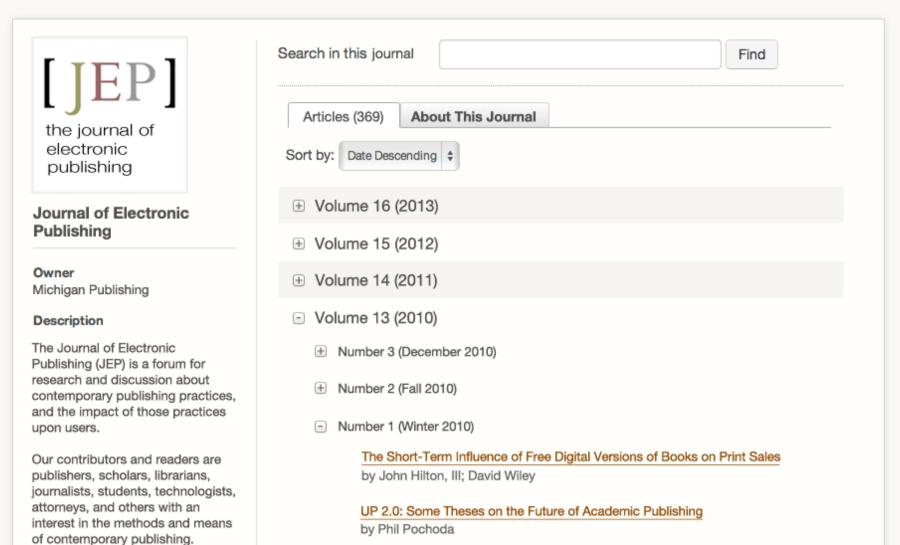
The Portable Document Format (PDF) has become the standard and preferred form for the digital edition of scholarly journal articles. Originally created as a solution to the need to "view and print anywhere," this technology has steadily evolved since the 1990s. However, its current use among scholarly publishers has been largely restricted to making research articles print-ready, and this greatly limits the potential capacity of the PDF research article to form a greater part of a digital knowledge ecology. While this article considers historical issues of design and format in scholarly publishing, it also takes a very practical approach, providing demonstrations and examples to assist publishers and scholars in finding greater scholarly value in the way the PDF is used for journal articles. This involves but is not limited to graphic design and bibliographic linking, the deployment of metadata and research data, and the ability to combine elements of improved machine and human readability.

#### Introduction

The Portable Document Format (PDF) was released by Adobe Systems in 1993 to facilitate the electronic distribution of documents. It was created to assist the circulation of digital documents among the newly networked computers that were spreading through offices, whether in local area networks (LAN) or







Questions so far?



**FULL-TEXT CATALOG** 

Search words about the items

All Fields

LOG IN \*

Advanced catalog search Search tips

#### ▼ Full view only

#### Similar Items

#### Orpheus: myths of the world.

By: Colum, Padraic, 1881-1972. Published: (1930)

#### Creatures,

By: Colum, Padraic, 1881-1972. Published: (1927)

#### At the gateways of the day. With illustrations

By: Colum, Padraic, 1881-1972.

Published: (1924)

#### The boy who knew what the birds said /

By: Colum, Padraic, 1881-1972.

Published: (1920)

#### The island of the mighty; being the hero stories of Celtic Britain,

By: Colum, Padraic, 1881-1972.

Published: (1929)

#### The road round Ireland.

By: Colum, Padraic, 1881-1972. Published: (1926)

#### The road round Ireland.

By: Colum, Padraic, 1881-1972.

Published: (1930)

#### ← Back to Search Results

Showing record 1 of 548 Next record



Export to Endnote



The forge in the forest / by Padraic Colum; with pictures by Boris Artzybasheff.

| Main Author:            | Colum, Padraic, 1881-1972.                     |
|-------------------------|--|
| Other Authors:          | Artzybasheff, Boris, 1899-1965., Dohm, Janice, |
| Language(s):            | English  |
| Published:              | New York: Macmillan Co., 1925.                 |
| Physical Description:   | 148 p. : ill. ; 20 cm.                         |
| Locate a Print Version: | Find in a library                              |

#### Viewability:

■ Full view

(original from University of Michigan)

#### Similar Items

### Scheme of transliteration of the School of Russian studies.

By: University of Liverpool. School of Russian Studies. Published: (1913)

## Russian review, a quarterly review of Russian history, politics, economics, and literature.

By: Baring, Maurice, 1874-1945

Published: (1912)

#### The Russian review.

By: Harper, Samuel N. 1882-1943.

Published: (1912)

#### The Russian review.

By: Pasvolsky, Leo, 1893-1953.

Published: (1916)

#### The Russian review.

Published: (1916)

#### The Russian review.

Published: (1916)

#### The Russian review.

By: Von Mohrenschildt, Dimitri Sergius,

1902-2002.

Published: (1941)

#### Russian review.

Published: (1945)

#### ← Back to Search Results

Showing record 1 of 546 Next record →



**Export to Endnote** 



#### The Russian review.

| Language(s):            | English   |
|-------------------------|---|
| Published:              | London: T. Nelson, 1912-1914.   |
| Subjects:               | Russian literature > History and criticism > Periodicals.  Russia > History > Periodicals.  Russia > Politics and government > Periodicals. |
| Note:                   | Issued by: the School of Russian Studies in the University of Liverpool.  |
| Physical Description:   | 3 v. : ill. ; 25 cm.  |
| Locate a Print Version: | Find in a library   |

#### Viewability:

| ■ Full view | <u>v.1 1912</u> | (original from Cornell University) |
|-------------|-----------------|------------------------------------|
| ■ Full view | <u>v.2 1913</u> | (original from Cornell University) |
| ■ Full view | v.3 1914        | (original from Cornell University) |

# HathiTrust's Bibliographic Metadata Specifications

When a HathiTrust partner institution provides a digital object for inclusion in HathiTrust, it must provide a catalog record in MARCXML format using fields as defined in the <u>Bibliographic</u> <u>Metadata Specifications</u>, an extension of MARC 21 minimal-level requirements.

What is the repository unit (barcode equivalent) for born-digital journals?

an individual article

### But ...

There is also metadata that relates to the journal as a whole, such as:

- title of the journal
- name of the publisher
- place of publication

What to do with these?

# mPach's solution: creating two kinds of records conforming to HathiTrust's Bibliographic Metadata Specifications

| Serial record for the journal | Analytic record for each article         |
|-------------------------------|--|
| Created manually              | Created automatically by mPach's Prepper |

# Workflow for manual creation of serial records (1/2)

When a new journal comes along that will use mPach:

- 1. Journal editor fills out a form that asks for:
  - journal title
  - any alternative titles or abbreviations
  - any previous titles
  - any ISSNs related to the journal
  - a short description of the scope of the journal

# Workflow for manual creation of serial records (2/2)

 A serials cataloger will check to see if the HathiTrust catalog already contains a record for the journal (or for any previous titles). They will be modified, a new record will be created, or both—linking to the journal's homepage.

#### Similar Items

### Scheme of transliteration of the School of Russian studies.

By: University of Liverpool. School of Russian Studies. Published: (1913)

## Russian review, a quarterly review of Russian history, politics, economics, and literature.

By: Baring, Maurice, 1874-1945

Published: (1912)

#### The Russian review.

By: Harper, Samuel N. 1882-1943.

Published: (1912)

#### The Russian review.

By: Pasvolsky, Leo, 1893-1953.

Published: (1916)

#### The Russian review.

Published: (1916)

#### The Russian review.

Published: (1916)

#### The Russian review.

By: Von Mohrenschildt, Dimitri Sergius,

1902-2002. Published: (1941)

### Russian review.

Published: (1945)

#### ← Back to Search Results

Showing record 1 of 546 Next record →



**Export to Endnote** 



#### The Russian review.

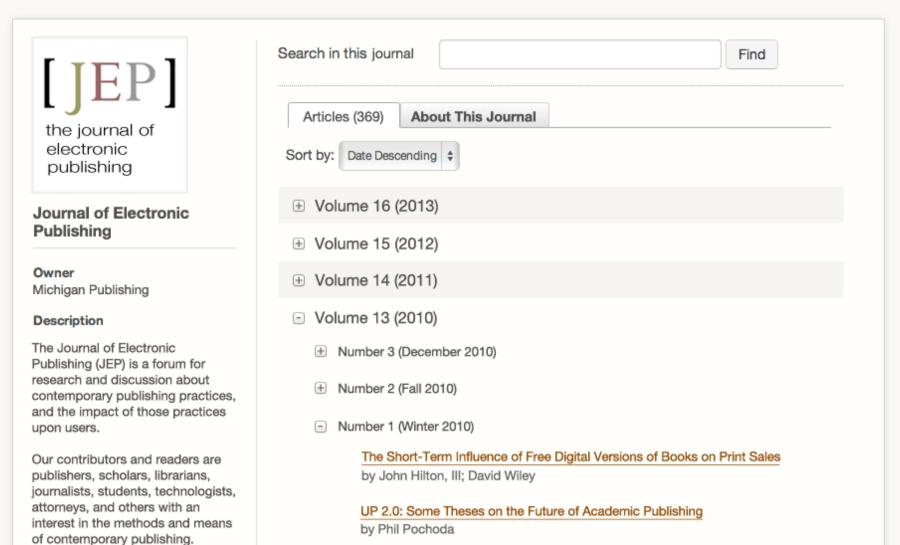
| Language(s):            | English   |
|-------------------------|---|
| Published:              | London: T. Nelson, 1912-1914.   |
| Subjects:               | Russian literature > History and criticism > Periodicals.  Russia > History > Periodicals.  Russia > Politics and government > Periodicals. |
| Note:                   | Issued by: the School of Russian Studies in the University of Liverpool.  |
| Physical Description:   | 3 v. : ill. ; 25 cm.  |
| Locate a Print Version: | Find in a library   |

#### Viewability:

| ■ Full view | v.1 1912        | (original from Cornell University)     |
|-------------|-----------------|--|
| ■ Full view | <u>v.2 1913</u> | (original from Cornell University)     |
| E Full view | v.3 1914        | (original from Cornell University)     |
| Full view   | v. 4 (2014) -   | (original from University of Michigan) |







# So can users only discover articles by way of the journal homepage?

### Nope!

The analytic records for each article will also be in the HathiTrust catalog, so you can find articles directly (if, for example, you search the catalog for a known article title).

# Automatic creation of article records (1/2)

To review, the user (e.g., the journal editor) uses mPach's Prepper to prepare an article for ingest into HathiTrust.

A combination of paragraph styles in Microsoft Word and manually entered metadata in Prepper ensures that the bibliographic metadata is properly encoded in JATS XML.

# Automatic creation of article records (2/2)

So because we'll have data that is correctly structured and actually correct, we will be able to map from JATS XML to the fields required to create an analytic MARCXML record for the article.

Each analytic record will be created automatically at the time that an article is ingested.

Our crosswalk, developed with significant assistance from Steven Holloway at ATLA, was donated to the JATS community on the JATS wiki.

# But how good are these records? Do they follow AACR2 or RDA?

### Not in the following ways:

- Records will not have titles of articles transcribed according to AACR2/RDA; instead, they will be in the record as displayed in the article.
- Names will be handled as the mPach user spelled them and divided them into forenames and surnames.
- We haven't bothered with choosing a main entry: all access points are added entries.

For anyone interested, I have an annotated handout of a working document showing how the analytic and serial records will relate to each other and the other components of mPach.

Questions?

http://www.lib.umich.edu/mpach