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

Kevin S. HAWKINS
@KevinSHawkins
Making of America Journals

Making 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:  

• 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
Contents

Front Matter
Advertisements
Yom Kippur Eve In A Jewish Home Before Going To Schul.
Advertisements
Zionism.
    Sonneschein, Rosa; pp. 5-9
A Vision of Jerusalem
<table>
<thead>
<tr>
<th>About DLXS</th>
<th>Products &amp; Fees</th>
<th>Technical Info</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>· Overview</td>
<td>· Products</td>
<td>· Requirements</td>
<td>· Workshops &amp; meetings</td>
</tr>
<tr>
<td>· Prospectus</td>
<td>· Supported collections</td>
<td>· Environment</td>
<td>· Documentation</td>
</tr>
<tr>
<td>· Brochure</td>
<td>· XPAT</td>
<td></td>
<td>· DLXS User Group</td>
</tr>
<tr>
<td>· Members/Contacts</td>
<td></td>
<td></td>
<td>· DLXS free listserv</td>
</tr>
<tr>
<td>· DLXS examples</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
inspired
formed the basis of
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?

Agata Mrva-Montoya

Despite the exponential rise of social media use in the publishing industry, very
### Browse by Issue

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

<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editor’s Note [15.1]</td>
<td>Welzenbach, Rebecca; Kimball, Shana</td>
</tr>
<tr>
<td>Refurbishing the Camelot of Scholarship: How to Improve the Digital Contribution of the PDF Research Article</td>
<td>Willinsky, John; Garnett, Alex; Pan Wong, Angela</td>
</tr>
<tr>
<td>Publisher-Library Partnership for Accessibility: A Case Study of Scholarly Publishing for Public Audiences</td>
<td>Anderson-Wilk, Mark; Kunda, Sue</td>
</tr>
<tr>
<td>Review of Uncreative Writing: Managing Language in the Digital Age by Kenneth Goldsmith</td>
<td>Chesley, Amelia</td>
</tr>
</tbody>
</table>

[Top of Page](#)
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/10.3998/10721865.0014.106
Permissions

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 de facto 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 Prepper (1 of 8)
mPach Prepper (2 of 8)

**Journal of Electronic Publishing** > Color Variability and Body Size of Larvae of Two Epomis Species (Coleoptera, Carabidae) in Israel, with a Key to the Larval Stages

## Conversion Report

### Article Metadata

Please review the following metadata. If there are any typos or excluded fields, please correct in your source document and re-submit.

- **Article Title**: Color variability and body size of larvae of two Epomis species (Coleoptera, Carabidae)

#### Authors

- **Gil Wizen**
- **Avital Gasith**

**No contributors found**

#### Keywords

- Epomis larvae, Carabidae, color atlas, body size

#### Abstract

Species identification using the characteristics of developmental stages is challenging. However, for insect taxonomy the coloration of larval stages can be an informative feature. The use of live specimens is recommended for this because the color fades in preserved specimens. In this study we examine the

#### Pub Date


mPach Prepper (3 of 8)
Color variability and body size of larvae of two *Epomis* species (Coleoptera, Carabidae) in Israel, with a key to the larval stages

Gil Wizen
Avital Gershfeld

Abstract
Species identification using the characteristics of developmental stages is challenging. However, for insect taxonomy the coloration of larval stages can be an informative feature. The use of live specimens is recommended for this because the color fades in preserved specimens. In this study we examine the possibility of using variation in coloration and color pattern of larvae in order to distinguish between two ground beetles species *Epomis deporti* (Dejean, 1831) and *Epomis circumscriptus* (Dufresnay, 1812). We present an atlas and describe the coloration and body size of the three larval stages of the above species based on live specimens. An identification key is given for the three larval instars of the two *Epomis* species. The first instar larvae of the two *Epomis* species can be easily distinguished based on their color. From the second instar on, the variability in coloration and color patterns increases, creating an overlap in these attributes between larvae of the two species. Except for minor differences in color of the antennae and the base of the mandibles, larvae of the two species are indistinguishable at the second
Supplemental Material

Article PDF

Optional. Upload up to one PDF version of the article.

Other Material

Optional. Upload additional supplemental material, such as data sets, presentations, etc. You can add an unlimited number of files. Please provide a descriptive label for each item.

File | Label
---|---
Choose File | No file chosen
e.g. Data for figure 2, ABC conference presentation, etc.
Review

We successfully build and validated the METS package. This article is ready for submission to HathiTrust.
A Confirmation of submission

mPach submission # 0123456789

The HathiTrust has received your submission for

Color variability and body size of larvae of two Epomis species (Coleoptera, Carabidae) in Israel, with a key to the larval stages

Return to journal...
mPach Prepper (8 of 8)
Refurbishing the Camelot of Scholarship: How to Improve the Digital Contribution of the PDF Research Article

John Willinsky, Alex Garnett, and Angela Pan Wong

Volume 15, Issue 1, Summer 2012
DOI: http://dx.doi.org/10.3919/jep.2012.01.010

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 wide area networks (WAN). The PDF format quickly gained the attention of publishers as a format that could be easily used to convert legacy documents into a format that was both human-readable and machine-readable.
Journal of Electronic Publishing

Owner
Michigan Publishing

Description
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.

Our contributors and readers are publishers, scholars, librarians, journalists, students, technologists, attorneys, and others with an interest in the methods and means of contemporary publishing.

Search in this journal

Articles (369)   About This Journal

Sort by: Date Descending

Volume 16 (2013)

Volume 15 (2012)

Volume 14 (2011)

Volume 13 (2010)

Number 3 (December 2010)

Number 2 (Fall 2010)

Number 1 (Winter 2010)

The Short-Term Influence of Free Digital Versions of Books on Print Sales
by John Hilton, III; David Wiley

UP 2.0: Some Theses on the Future of Academic Publishing
by Phil Pochoda
Questions so far?
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)
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 Bibliographic Metadata Specifications, 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

<table>
<thead>
<tr>
<th>Serial record for the journal</th>
<th>Analytic record for each article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Created manually</td>
<td>Created automatically by mPach’s Prepper</td>
</tr>
</tbody>
</table>
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)

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.
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