



Opening our deep backfiles

Identifying open and public domain serial content in library collections

Presenters:

John Mark Ockerbloom

Rachelle R. Nelson



[JOHN:] Hello, and welcome. I'm John Mark Ockerbloom. I'm presenting today with Rachelle Nelson. We're speaking to you today from Philadelphia, on the traditional lands of the Lenape people.

We're talking today about work we're doing at the Penn Libraries to help make historic serial content in our collections freely available to the world, by having staff across our libraries research their copyrights and share our findings through linked open data.

The problem



- Lots of older serials with relevant research and cultural content, largely overlooked
 - and often hard to access, especially during a pandemic
- Many of them (as late as 1960s or even 1980s) in the public domain, but their copyright status is unclear
 - Most did not renew copyrights; some didn't even claim them



Like many research libraries, the Penn Libraries own many volumes of serials from the 20th century. Many of them don't get much use. They're not easily accessible--even before the pandemic closed our stacks to the public, many had been shipped to offsite storage. And while mass digitization has made many of the really old volumes accessible online, content after 1925 usually isn't available online without subscriptions, if it's available at all. But we know that many of them up to the 1960s or even the late 1980s are out of copyright, since most of them don't have any renewed copyrights, and some might not have claimed a copyright to begin with.

What we're doing

- Identifying serials in the public domain, or with free online issues
- Publishing what we find as structured JSON data linked to Wikidata
- Linking to free online content and related material when we find it (including when it's under ©)
- Growing the public commons of serial literature



Penn Libraries

Deep Backfile project page

In the Penn Libraries Deep Backfile project, dozens of people who work at the Penn Libraries are researching copyrights of serials in our collection, so we can identify ones that are in the public domain and can be shared freely with the world. We're publishing our findings on the Web, through web pages and JSON structured data files linked with Wikidata.

Basic information

- "Invitation to participate in a new project: Help open journals' deep backfiles" by John Mark Ockerbloom is a post from his [Everybody's Libraries](#) blog explaining the basic goals of the project, before it expanded into a Penn Libraries project for library staff.
- The [Deep Backfile data](#) page links to the main information tables being produced by the project. Penn library workers are mainly filling in the [Penn Libraries deep backfiles](#) table, but the other tables are also getting populated with information as project work progresses.
- Our guide "[Determining copyright status of serial issues](#)" describes a process for using the data we're collecting in determining whether a particular serial issue is in the public domain or under copyright in the United States.

Project documentation and training materials

<https://onlinebooks.library.upenn.edu/cce/db/>

When the COVID outbreak sent our staff home, we saw an opportunity to have them help research the copyrights of serials we own, and also to see if any of them were already free online, with or without copyrights. We created a table of over 10,000 serials we own from the mid-20th century, and a questionnaire to gather information on their copyrights. We're publishing that information as structured JSON data linked to Wikidata. We're also linking to free serial content when we find it. That's all helping make this content widely available to the world, and getting the attention it deserves.

I'll turn things over now to Rachelle Nelson to show how our process works.

A variety of roles for a variety of workers



Identify the serial you're interested in

Title

ISSN

Provide information on this serial's copyrights

Does this serial originate in the US, or some other country? (Penn's [Franklin record](#) may say.)

If not the US, from what country does it originate?

- We trained 25 workers in 14 departments to search for copyright information in online databases and resources and complete web forms on their findings
- Work can be done from home, with varying commitment levels

[RACHELLE:]

Twenty-five staff from fourteen departments signed up to work from home on this project. The questionnaire we designed for them to fill out can be completed in any web browser, without any extra software or specialized expertise needed. Staff search our catalog, the Copyright Office database, and other online resources to enter information about a serial's country of origin, its copyright renewals, its copyright notices, and any free online content. We held training sessions on Zoom showing how to work through the forms step by step. We answered questions from staff, and we made recordings and guides available for staff working on the project.

After the questionnaire, more specialized work



Copyright information

Title: *The Journal of Ecclesiastical History*
Online content: [Free online material via The Online Books Page](#)
More information: [Wikipedia article](#); [Wikidata](#)
First renewed issue: no issue renewals found in CCE or registered works database
First renewed contribution in: no contribution renewals found in CCE or registered works database

Additional note

This is a British publication. Its contents may still be copyrighted even without renewals or copyright notices, if they meet GATT copyright restoration requirements for works first published outside the US. However, the 1967-1968 volumes list a US subscription agent and US subscription price. We did not find copyright notices in them. Other volumes around the same time may be similar, but we have not examined them.

Page information

Page responsibility: John Mark Ockerbloom (ockerblo (at) pobox (dot) upenn (dot) edu)
Acknowledgement: Thanks to Felice Gollotti for copyright research on this title.

- Verify staff answers
- Create JSON records from the answers, generating web pages like this...

After a staff member fills out the questionnaire for a particular serial, their response gets put in a queue for others to review. This is where more specialized work comes in. Right now, John double-checks the answers given, edits them as needed, and then runs a script to create a JSON file with a unique identifier. The JSON file encodes copyright information in a form that can be processed by machines. It's also used to create serial copyright web pages like the one you see here. We make sure the file acknowledges the people who worked on it.

...and create records for free online issues if any...



The Online Books Page

presents serial archive listings for

The Journal of Ecclesiastical History

The Journal of Ecclesiastical History is a journal on the history of Christianity published by Cambridge University Press. (There is a [Wikipedia article about this serial](#).)

Publication History

The Journal of Ecclesiastical History began in 1950. No issue or contribution copyright renewals were found for this serial. ([More details](#)) It is still published today.

Persistent Archives of Complete Issues

- **1967-1968:** HathiTrust has [volumes 18 and 19](#) freely readable online. (The volumes may be misdated by HathiTrust; however, we found no copyright notice in the volumes, and did find US subscription agents and prices mentioned in the front matter. Works of that era that were initially or simultaneously published in the United States required a copyright notice to secure copyright.)

Official Site / Current Material

- Cambridge University Press's [Journal of Ecclesiastical History web page](#) has more information about this journal, and provides access to its full run. Most articles requires subscription or payment to access at this site, but a small number are available open access.

This is a record of a major [serial archive](#). This page is maintained for [The Online Books Page](#). (See [our criteria for listing serial archives](#).) This page has no affiliation with the serial or its publisher.

If our workers found free online content for the serial, either from our existing catalog records, or from searching online, other workers then create records that link to that content. Those records are discoverable in Penn's Online Books Page.

...and link to Wikidata (creating items if needed)



The screenshot displays the Wikidata interface for the item 'The Journal of Ecclesiastical History' (Q7743575). The page title is 'The Journal of Ecclesiastical History (Q7743575)'. Below the title, it is identified as a 'journal'. The 'Statements' section lists two types of statements:

Statement	Value	References	Action
instance of	history journal	0 references	+ add reference
instance of	academic journal	0 references	+ add reference

We also make sure that there's a Wikidata item for the serial. If there isn't one already, we create one. When we add the unique identifier for the serial's JSON file to its Wikidata item, a link gets automatically created to our copyright information page on that serial. We now have several librarians working on these serial records in Wikidata, and we also have other librarians working on methods to automatically add Wikidata records to serials not already represented there.

Wikidata IDs & ISSNs pull all this data together



0731-3667	reading.	1957-1964	Q98971576	None	None known	us
0097-4250	The Journal of documentary reproduction.	1938-1942	Q99325541	None	None known	Contact us
0022-0418	The Journal of documentation.	1945-	Q6295097	None*	None known	Contact us
0022-0469	The journal of ecclesiastical history.	1950-	Q7743575	None*	1967-1968	Contact us
0022-0477	The Journal of ecology.		Q766513	None*	1913- -recent	Contact us

We also rely on a lot of careful work done, and data recorded, by serials catalogers in tracing and cataloging the serials that we document in this project

The Wikidata identifier and the ISSN are what tie all of our data together. They help ensure that we can link our copyright information about a serial to our catalog records, and they make it clear what serial we're talking about when we describe one. As we've increasingly relied on ISSNs in this project, we've found that many of the ISSNs in our catalog were incorrect or canceled. We've been fixing them as we find them. We can do that with our own catalog records, and with master records in WorldCat. We've had more difficulty with our Alma community zone records, because we're not allowed to edit ISSNs there.

We've also run into limits on how much authoritative information we can get about particular ISSNs because we lack an ISSN Portal subscription.

Overall, though, we've been fortunate to draw on a long history of careful work that serial catalogers have done to describe and trace the history of serials, and to locate online content. We're very grateful to be able to build on that work.

We can document more than just what Penn owns



Penn Libraries



Deep backfile data

This page gives access to information on sets of serials that are or potentially could be freely available online. We are actively seeking more information on these sets, and welcome your data and expressions of interest. See our [project page](#) for more information.

Choose from the following sets:

- **Cambridge University Press**
 - [Cambridge University Press journals free and potential public domain backfile](#)
- **Elsevier**
 - [ScienceDirect Available Journals free and potential public domain content](#)
- **JSTOR**
 - [JSTOR free and potential public domain backfile](#) (predecessor and successor titles consolidated)
 - [JSTOR free and potential public domain backfile](#) (predecessor and successor titles separated)
- **Library holdings**
 - [Penn Libraries free-online and potential public domain deep backfiles](#)
- **Oxford University Press**

[JOHN:]

In our current project, we're concentrating on serials owned by Penn. But others may be particularly interested in different serials. This page links to lists of serials in popular subscription packages that may have public domain content. It also links to lists of serials written about in Wikipedia. Anyone can research or ask about serials that are of interest to them in these sets. And as we fill in information on serials in Penn's table, or in another table shown here, that information also appears in other tables that include the same serial.

How this work can benefit you



- You and your scholars can **consult our knowledge base** to find serial publications you can digitize, freely quote, or otherwise reuse.

- We're talking to some nonprofit mass-digitization projects about using our data to **further open already digitized serial content**

Penn Libraries deep backfiles

Serials in Penn Libraries' holdings that have free online and/or potentially public domain content. Publications without a reported ISSN are not included in this table. Some serial subtitles do not yet appear in this table. Select the title link of any serial to see its full title (and more details about what we have) in Penn's Franklin catalog.

Table generated March 30, 2021 using provider data downloaded May 22, 2020. [A table key follows.](#) ([Download as a spreadsheet.](#))

ISSN	Title	Coverage	Wikidata	First renewal	Free issues	More info?
0094-3320	13th moon		Q95584323	All	1973-2003	Contact us
1054-7193	291	1915-1916	Q372444	N/A	1915-1916	Contact us
0365-6152	L'Aéronautique	1919-1940	Q95981116	None*	1919-1940	Contact us
0567-3356	AACTE study series /	1953-1965	Q96000171	None*	1954-jj-1963	Contact us
0001-0197	AARN news letter	1948-	Q27706670	None*	None known	Contact us

So if you're interested in using or digitizing content from 20th century serials, or if some of your users are, you can consult our knowledge base to see whether that content is still subject to copyright. Our project page includes a decision guide you can consult on how to use the information we've gathered, and on how to look up additional information if you need to.

In the future, the serials we're researching might simply become open online for everybody. We've been talking with HathiTrust about a pilot program to review and open up volumes for some of these serials after 1925, based on our data and recommended procedures.

How you can build on this work



- You can work on the serials we're working on as well
 - Researching copyrights, making wikidata records...
 - Or digitizing based on the data
- You can work on serials of special interest to you
 - Individually (whether or not they're part of one of our sets)
 - Or possibly define a set of your own to research (talk to us)

The screenshot shows the Penn Libraries website with a dark blue header containing the Penn Libraries logo and name. Below the header, the title "Deep backfile data" is centered. A paragraph of text explains that the page provides information on sets of serials that are or potentially could be freely available online, and that the site is actively seeking more information on these sets. Below this text, a section titled "Choose from the following sets:" lists several categories of backfile data, including Cambridge University Press, Elsevier, JSTOR, and library holdings, with sub-items for each category.

You can also build on the work we're doing. We still have thousands of serials in our collections to research, and we welcome participation from people at other libraries, as well as our own workers.

If you're interested in a serial not on the Penn Libraries list, you might find it on one of the other Deep Backfile lists we have on our site. Or you could make up your own list of serials, and research the titles in it like we do. If your list is likely to be of wide interest, you can talk to us about possibly having it featured on our site as well.

Where you can find out more



John Mark Ockerbloom
ockerblo@upenn.edu



Rachelle R. Nelson
nelsonrr@upenn.edu

Project page: <https://onlinebooks.library.upenn.edu/cce/db/>

Thank you!

We've been able to do as much as we have on this project thanks to dozens of people at the Penn Libraries who have given their time and talent to it while we've all been trying our best to ride out the COVID crisis. If you go to our project page, at the URL on this slide, you'll find the credits for their work, as well as links to all the information, documentation, and training we've produced. We hope you've found this presentation informative, and we'd be glad to talk with you if you'd like to learn more about it or get involved yourself. Thank you!