

PRESERVING EBOOKS: PAST, PRESENT AND FUTURE

A Series of National Library Perspectives

Trevor Owens

*Library of Congress
USA*

trow@loc.gov

<https://orcid.org/0000-0001-8857-388X>

Maureen Pennock

*British Library
UK*

maureen.pennock@bl.uk

<https://orcid.org/0000-0002-7521-8536>

Faye Lemay

*Library & Archives Canada
Canada*

faye.lemay@canada.ca

Tobias Steinke

*Deutsche Nationalbibliothek
Germany*

t.steinke@dnb.de

<https://orcid.org/0000-0002-3999-1687>

Abstract – This panel will present and discuss different eBook workflows and challenges from four national libraries, considering a range of issues from technical complexities to evolution of the content type and changes in the publishing/collecting landscape.

Keywords – digital preservation, ebooks, ingest, formats, scale, access

Conference Topics – The Cutting Edge: Technical Infrastructure & Implementation; Exploring New Horizons

I. OVERVIEW

eBooks are the backbone of many a National Library collection, constituting a substantial proportion of the digital content our readers expect to be able to access and consult. Our digital preservation activities reflect this, with established infrastructures and workflows for eBook acquisition, ingest, management and access, all at scale. Yet the eBook as a content type is evolving, and user expectations for access are evolving alongside. Dealing with this requires both a responsive framework and an eye on the horizon.

This panel brings together experts from leading national libraries to openly discuss various elements of their respective eBook preservation activities and research programs, and explore where similarities and differences may lie. Below we summarize the eBook collections at each organization, existing challenges, and research activities.

A. *eBooks at the British Library*

Since 2013 The British Library has collected eBooks under the UK's Non-Print Legal Deposit (NPLD) Regulations. Our preferred formats are EPUB and PDF

though we also have a small number of MOBI files. There are around 400,000 NPLD eBooks in the collection with access rates at around 5,500 per month. We also have a substantial number of digitized books published under commercial partnerships with Google and Microsoft. Going forwards, we have an interest in Open Access eBooks published outside of the UK and eBooks published as mobile apps.

Current challenges include ensuring an uninterrupted supply to readers during a forthcoming repository migration, and delivering access to all six UK Legal Deposit Libraries in line with regulation requirements for single sequential access. Active research areas include collection and preservation of mobile apps and evolution of the EPUB format.

B. *eBooks at the Library of Congress*

The U.S Library of Congress has acquired eBooks through a wide range of different programs and initiatives. For years, the institution has received and acquired eBooks through its Cataloging in Publication Program, special relief agreements for copyright deposit, web archiving, and other routine transfer methods for acquisition.

In support of the digital collecting plan, staff across the institution are currently working to expand these efforts and to pilot acquiring, preserving, and delivering selected open access eBooks. The majority of this content is in PDF and EPUB formats, but the institution has copies of eBooks in a much wider range of formats as well. As outlined in the Library of Congress Digital Strategy, it is necessary to plan for work around eBooks in terms of exponential collection growth. To that end,

a key area of focus for the institution is working to scale up and enhance workflows and processes.

C. *eBooks at the Deutsche Nationalbibliothek*

The German National Library has currently around 1 million eBooks in the formats PDF and EPUB, equating to approx. 16% of all collected digital publications (excluding digitized objects). The German legal deposit collection has included eBooks since 2006. eBooks are ingested in the digital preservation system of the German National Library. All eBooks are analyzed and validated, resulting in generation of a risk analysis 'ingest level'. Checks include tests on copy protection especially in PDF files. There is a separate repository for giving access.

In an ongoing internal project all aspects of the digital workflows are currently being optimized for a better performance. This includes using a common workflow engine, replacing the repository for access with something more fitting and consolidating the different workflows for digital objects including eBooks.

D. *eBooks at Library & Archives Canada*

LAC has been acquiring eBooks of various different formats since the 1990's. Digital legal deposit legislation came into effect in 2006, though participation in the legal deposit program varies with commercial/retail publishers and scholarly communities lagging behind government and self-published content.

The current technical platform for eBook acquisition is based on a pilot project created in 1994. In 2018, LAC embarked on an initiative to modernize its systems and, as part of that, procured Preservica as a DAM and a Digital Preservation Solution. New information package specifications for published heritage collections are currently being developed for use within Preservica. In addition, LAC's Published Acquisitions sector is working to implement a collection gap analysis and monitoring framework in order to measure and expand participation in the Legal Deposit program. Another key activity is the development of a seamless platform for publishers and authors to transfer digital content and metadata to LAC. One of the desirable outcomes is to ensure that streamlined workflows from acquisition to preservation are developed.

II. PANEL STRUCTURE

Following short introductions on the state of the practice to acquire, preserve, and deliver eBooks at each institution, panelists will then move on to discuss a range of questions such as:

- How does your organization staff and support eBook acquisition, preservation and access?
- How have you embedded preservation support into your end to end workflows?

- Do you have preferred formats for eBook preservation; if so, what are they and why?
- What are the biggest challenges you have encountered in collecting, preserving and providing access to eBooks?
- What changes have you seen in your eBook collection over the past decade and how have you responded?
- How are you monitoring the publishing landscape for more changes going forwards?

Panelists will discuss answers in advance of the session to ensure answers are representative of the variety in our approaches, thus ensuring we provide sufficient conflicting perspectives to create interesting discussion. Attendees will be encouraged to ask additional questions of the panelists during an open-ended Q&A session.

III. PANELISTS

Maureen Pennock is Head of Digital Preservation at the British Library. She sits on the Digital Preservation Coalition Board of Directors and co-chairs the DPC Special Interest Group for Digital Preservation in National Libraries, Archives and Museums. She is also Chair of the UK Legal Deposit Libraries' Digital Preservation Committee and a member of the UNESCO PERSIST initiative.

Dr. Trevor Owens serves as the first Head of Digital Content Management at the U.S. Library of Congress. In addition, he teaches graduate seminars in digital history for American University's History Department and graduate seminars and digital preservation for the University of Maryland's College of Information, where he is also a Research Affiliate with the Digital Curation Innovation Center

Tobias Steinke works at the German National Library on the conceptual development of digital preservation and is responsible for the web archiving project of the library. He has been involved in several national and international projects about digital preservation and standardization.

Faye Lemay has been the Manager of Digital Preservation at Library and Archives Canada for nearly a decade and has been the driving force in the development and deployment of a comprehensive digital preservation program. She oversees the long-term preservation of Canada's digital documentary heritage comprised of published heritage, government records and private archives.

The panel will be moderated by Paul Wheatley, Head of Research & Practice at the Digital Preservation Coalition. Paul is an experienced panelist and moderator with many years of experience working with digital collections and in digital preservation.