

DIGITAL LIBRARY MATURATION

The Evolution of Digital Collections Management Programs at National Libraries

Trevor Owens

*U.S. Library of Congress
trow@loc.gov
0000-0001-8857-388X*

Jeffrey van der Hoeven

*National Library of the
Netherlands
jeffrey.vanderhoeven@kb.nl*

Andrea Goethals

*National Library of New
Zealand
Andrea.Goethals@dia.govt.nz
0000-0002-5254-9818*

Lisandro Pablo Olivares

*Biblioteca Nacional de México
lspablo@comunidad.unam.mx*

Paul Wheatley

*Digital Preservation Coalition
paul.wheatley@dpconline.org*

Michael Day

*The British Library
Michael.Day@bl.uk
0000-0002-1443-5334*

Abstract – This panel will present and discuss the development of digital collections management programs at a series of national libraries. Panelists will each give brief overviews of the development of their institutions programs, policies, and technical infrastructure and share out about upcoming and ongoing major initiatives in these areas.

Keywords – digital preservation, national libraries, policy, infrastructure, formats, scale, access

Conference Topics – Building the Capacity & Capability

I. OVERVIEW

Over the last three decades, national libraries have worked to iteratively develop and launch digital collections management programs. In that time, digitization and born digital collecting have grown from initial pilot efforts to become major components of the programs and services of these institutions. Of particular note, national libraries have responded with the development and implementation of digital legal deposit programs for published works, web archiving programs to collect and preserve digital content on the web, the development of processes for working with born digital special collections formats, and have needed to establish capacity to ensure enduring access to

content produced through mass digitization programs.

This panel brings together practitioners from leading national libraries to openly discuss various elements of the development of their digital content management programs. Below we summarize the digital collecting and management programs at each organization, current initiatives, existing challenges, and research activities.

A. *Digital Content at the U.S. Library of Congress*

The Library of Congress has been producing and managing digital collection content for nearly five decades. As early as 1982 the organization had begun digitization pilot programs. As those programs expanded in the late 1990s into the early 2000s, early initiatives rapidly grew to a wide range of digital collections programs and projects. Of particular note, the Web Archiving Program launched in 2000, the copyright eDeposit for eSerials launched in 2009, and the Cataloging-in-Publication program began to acquire eBooks in 2012. More recently, work on the digital collections has become central to organizational planning and strategic efforts.

In 2017, the Library of Congress established a new Digital Content Management unit explicitly

focused on managing general digital collections and coordinating policy development and technical infrastructure. Since that time, the organization has made major progress on standardizing policy, improving methods to process and track digital content, and improvements to methods to provide both onsite and offsite access to digital collections.

B. Digital Content at the National Library of New Zealand

In 2003, the National Library of New Zealand (NLNZ) Act was passed providing the legal mandate for NLNZ to collect and preserve digital content relating to Aotearoa New Zealand under legal deposit. In 2004 the New Zealand government provided funding for NLNZ to establish the National Digital Heritage Archive (NDHA) to meet this mandate. NDHA was put into production in 2008 and now has 575 TB of digitised and born-digital material in 224 different file formats. NLNZ has eleven staff in specialised digital preservation roles, and other roles across NLNZ to support the end-to-end workflows that contribute to preserved digital collections.

NLNZ has been collecting websites since 1999, using the Web Curator tool to selectively collect websites, and Internet Archive's services to conduct annual whole-of-domain harvests. NLNZ is also engaged in social media archiving to document particular events and individuals. Some of the recent positions at the Library reflect a new emphasis on contemporary, community-driven and participatory collecting – a Covid-19 Digital Archivist and a Curator of Contemporary Voices.

In 2020 the New Zealand government provided funds for the largest digitization project in New Zealand. The project will digitize the audiovisual collections in at-risk formats, of NLNZ, Archives New Zealand, and Ngā Taonga Sound and Vision. The scale of the project presents many challenges - a key concern is the potentially high cost of storing the digitized AV material in perpetuity.

C. Digital Content at the National Library of the Netherlands (KBNL)

Over the last decades, the KBNL's digital collection grew substantially in volume and diversity. This is due to large-scale digitization of its paper collections and a paradigm shift in publishing towards digital only such as ebooks and web-based publishing. This required a new approach to acquisition and preservation in to ensure this valuable content is kept safe for future generations.

In 2018, the library kickstarted a new content strategy [1] in which the primary focus is on digital first, continuing digitization programmes and a strong incentive for archiving the Dutch web by expanding our current harvesting activities with a full domain crawl.

Since then, the library is making significant efforts to transform its acquisition, processing and preservation activities into mainstream digital, including an upgrade of the underlying IT infrastructure [1]. To reach a higher coverage of published ebooks we aligned our acquisition process with those of large distributors in the Netherlands to receive a digital copy directly from this process. The big advantage is that it streamlines the content from many different (smaller) publisher while having little impact on the publication process itself. We plan to extend this approach for other types of content as well, such as newspapers.

To ensure our digital holdings cover scholarly publications as good as possible, we are an active member of the online ISSN IC Keepers registry [2] which keeps track of the holdings of digital archives across the world.

D. Digital Content at Biblioteca Nacional de México

The National Library of Mexico has about six thousand titles of collections, archival and documentary funds, and iconography, in digital format, from the 15th century to the present day, that are part of the bibliographic memory of Mexico. It also houses more than 9 million images, of more than 940 titles of periodic publications of historical character from 1722 to 2006. In 2002 it started digitization programs of books and newspaper resulting in two web platforms "Biblioteca Nacional Digital de México" and "Hemeroteca Nacional Digital de México" that enable users to search and download digital content.

In 2017 the National Library of Mexico started implementing digital preservation in all digital collection management processes, creating strategies and policies, adapting workflows, reorganizing infrastructure, developing customized software, and creating study groups, that ease the management of its digital collections. Digital legal deposit is the next challenge, so the library is working on two fronts, one internal, defining reception criteria and software specifications; and the other

external, promoting federal legislation that facilitates and promotes the deposit of digital materials, emphasizing digital preservation.

E. *Digital Content at the British Library*

The British Library has been collecting and creating digital content for several decades. The digitisation of the Library's collections commenced in the early 1990s with the Initiatives for Access programme, and this has evolved today into an extensive portfolio of digitisation activities. The acquisition of digital publications is another major activity, and one that has evolved in-line with changes in the wider publishing environment. In the 2000s, for example, the Library developed a system for the voluntary deposit of e-journals. It was also one of the six organisations that established a UK Web Archive in 2005. A major change came in 2013, when the Legal Deposit (Non-Print Works) Regulations gave the British Library and the five other UK legal deposit libraries the right to collect digital publications through Legal Deposit. This meant, for example, that the UK Web Archive could scale-up to conduct annual crawls of the UK web domain as well as the more-regular harvesting of selected websites. It also led to the development of ingest workflows for other publication types, including: e-books, e-journals, geospatial data, and music scores. The Library also collects the born-digital (or hybrid) archives of individuals and organisations as part of its Personal Digital Archives programme.

The collection management and preservation of the British Library's digital collections is the responsibility of a large number of people from all across the organisation, including curatorial, technical, and operational colleagues. The Library also has a very-active digital scholarship team, which works with the research community to promote the use of digital collections and data. The Library's digital preservation team works with many others across the organisation to help ensure that all of its collections can be properly preserved for the long term. The team has a particular interest at the moment in the project implementing of a new repository system, the analysis of collections and file formats, and the development of frameworks for preservation planning.

II. PANEL STRUCTURE

Following short introductions on digital collections management programs at each institution, panelists will then move on to discuss a range of questions moderated by Paul Wheatley such as:

- How does your organization staff and support digital collections acquisition, preservation and access?
- How does your organization approach digital collection management policy?
- How does your organization approach issues with digital file formats? Has that approach evolved over time?
- What are the biggest shifts or changes you are working on with your technical infrastructure at this point in time?

Attendees will be encouraged to ask other questions of the panelists during an open-ended Q&A session.

III. PANELISTS

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 Center for Archival Futures.

Jeffrey van der Hoeven, MSc. is head of the Digital Preservation department at the National Library of the Netherlands (KB). In this role he is responsible for defining the policies, strategies and organizational implementation of digital preservation at the library, with the goal to keep the digital collections accessible to current users and generations to come.

Andrea Goethals is the Programme Director of the Preservation Research & Consultancy team at the National Library of New Zealand. She manages a talented team of digital preservation specialists and software developers, who look after the National Library's digital collections.

MSc. Lisandro Pablo Olivares is a software engineer working for the coordination of innovation and digital strategy at the Biblioteca Nacional de México and is in charge of implementing and promoting digital preservation processes on the digital collections the library manages.

Michael Day is Digital Preservation Research Lead at the British Library.

Paul Wheatley is the Head of Research and Practice Digital Preservation Coalition.

References

[1] KBNL's content strategy, available at: https://www.kb.nl/sites/default/files/docs/content_strategy_eng.pdf

[2] Dutch National Library steps into a new future of digital archiving, available at: <https://www.kb.nl/en/news/2021/dutch-national-library-steps-into-a-new-future-of-digital-archiving>

[3] ISSN IC Keepers website, available at: <https://keepers.issn.org/>