

# BUILDING POLICY & CAPACITY

## *Launching the U.S. Library of Congress Digital Collections Management Compendium*

**Trevor Owens**

USA

*trow@loc.gov*

ORCID: 0000-0001-8857-388X

**Jesse A. Johnston**

USA

*jajohnst@umich.edu*

ORCID: 0000-0003-2617-0166

**Pedro Gonzalez-Fernandez**

USA

*pgon@loc.gov*

**Camille Salas**

USA

*csala@loc.gov*

**Andrew Cassidy-Amstutz**

USA

*acas@loc.gov*

**Kristy Darby**

USA

*kdar@loc.gov*

**Abstract** – In 2019, the Library of Congress launched the Digital Collections Management Compendium (DCMC). This publicly-available resource establishes and communicates institutional digital collections management policy. A core function of the DCMC is to articulate the connections between broader digital collections management policy and that of planning resources produced and/or consulted by Library staff. This paper relates the development of the DCMC, outlines its key features, and supports broader public use of the DCMC as an open resource. We intend this paper to be of use to others working on organization-wide digital collections management initiatives. More broadly, the paper contributes to the discussion on how work around digital collections policy and planning connects with practice.

**Keywords** – policy, planning, storage systems,  
**Conference Topics** – Building the Capacity

### I. INTRODUCTION

Ensuring long term access to digital collections requires an integrated approach to planning, policy, and practice. As institutions mature their digital collections management programs, it is critical that they consolidate and integrate individual projects into institutionalized programs [1]. In this context, policy functions as connective tissue between the

practical realities of an organization's structure, technologies, and resources [2]. Functional policy must be anchored in a continual dialog with ongoing organizational practice [3]. As such, digital preservation policy development is vital to enabling the development of digital collections management programs.

Over the course of 2017 and 2018, the newly formed Digital Collections Management and Services Division of the Library of Congress was tasked with supporting the refinement of digital collections management policy for the organization. This resulted in the creation of an internal Digital Collections Management Compendium (DCMC), as well as a public version that was launched in 2019. In what follows, we present a case study of the DCMC and discuss its development and connections to other digital preservation resources produced and maintained by the Library of Congress.

### II. ORGANIZATIONAL CONTEXT

The current state of digital collections policy development at the Library of Congress is shaped by the history of digital collections at the organization. The Library of Congress has been producing and managing digital collection content for nearly five

17th International Conference on Digital Preservation  
iPRES 2021, Beijing, China.

Copyright held by the author(s). The text of this paper is published  
under a CC BY-SA license (<https://creativecommons.org/licenses/by/4.0/>).

DOI: 10.1145/nnnnnnn.nnnnnnn



decades. In 1982, the Library's Optical Disk Pilot Project began capturing text and images of collections [4.] In 1989, those initial explorations were developed into the American Memory pilot program. This work included initial digitization of selected collection materials, which were distributed on laser disc and CD-ROM. Ultimately, these became the basis of the institution's first online collections [5].

Over the last twenty years, those initiatives rapidly grew to encompass a wide range of digital collections programs and projects. Of particular note, the Web Archiving Program launched in 2000, [6] the copyright eDeposit for eSerials launched in 2009, [7] and the Cataloging-in-Publication program began to acquire eBooks in 2012 [8]. More recently, work on the digital collections has become central to organizational planning and strategic efforts.

In 2017, the Library of Congress published "Collecting Digital Content at the Library of Congress," a five-year plan that established targets associated with six strategic objectives for growing the digital collections [9]. In 2018, as a companion to its strategic plan, the Library of Congress published its first Digital Strategy which asserts the intention to "exponentially grow" its digital collections and "ensure enduring access" [10]. Now, after four decades of digitization and collecting born-digital materials, the organization holds a digital collection that is, like its physical collections, varied, heterogeneous, and expansive, containing hundreds of millions of inventoried digital assets with more than sixteen thousand unique file extensions [11].

Alongside the development of these varied digital collections programs, the Library of Congress has developed resources that support planning for enduring access to these materials. Created and coordinated by a range of stakeholders in the organization, these include the development and maintenance of the Sustainability of Digital Formats technical resource [12], the PREMIS Data Dictionary for Preservation Metadata [13], and the Library of Congress Recommended Formats Statement [14]. The organization also convenes stakeholders annually for the Designing Storage Architectures for Digital Collections summit. In addition, the Library was a founding convener of the National Digital Stewardship Alliance Levels of Digital Preservation. All of these resources and initiatives have become critical components of international community approaches to digital preservation. Each of these resources supports work in specific areas of digital

collections management. The Library's recent DCMC initiative was undertaken, in part, to clarify how these resources work together to guide digital collections management practices in the organization.

### III. DEFINING THE NEED

In 2016, the Library's senior management identified the need for developing a compendium of practices and guidance for digital collections. This need, which became the DCMC, was envisioned by management to broadly synthesize, harmonize, integrate, and further clarify and define digital collections management policy for the organization. This work built directly on the efforts of the Library of Congress Digital Preservation Working Group, which developed and shared an initial set of policy recommendations and resources in 2014 [9].

That effort coincided in 2016 with the establishment of a new organizational unit focused explicitly on digital content management [10]. The newly-established unit, the Digital Content Management Section (DCM) was tasked with developing a process to establish the production and maintenance of an internal policy resource; however, the scale and complexity of Library operations posed significant challenges to the implementation of broad digital collections management policy.

### IV. ESTABLISHING THE DIGITAL COLLECTIONS MANAGEMENT COMPENDIUM PROCESS

The Library of Congress has many distinct organizational functions and roles. This required a process that would allow experts from across the Library to participate in refining and developing policy. A representative group of nearly forty staff from across the organization was established as a review panel facilitated by DCM. One DCM senior staff member took primary responsibility of facilitating review panel communications, in addition to serving as the main editor for the Digital Collections Management Compendium. Two other DCM staff worked extensively to research, draft, and revise policy items.

To facilitate a transparent and open process for managing the development of guidance items, an internal wiki was created using the Confluence platform. DCM developed a template for guidance elements, as well as a process by which to monitor the status of guidance elements (i.e., In Draft, In Review, In Effect).

In the spirit of openness and by design, policy items for the DCMC could be proposed or drafted by anyone in the organization. In practice, the variety of activities and collections management priorities limited the time that staff from the review committee could devote to policy creation alongside their existing duties; to date, therefore, staff in DCM were the primary proposers and drafters of policy items. Once drafted and ready for review, each individual policy item is announced to the broader review group for comments during a two-week period. Reviewers commented directly on the draft items. In many cases, these discussion threads on individual items were extensive and involved staff from a wide range of roles and functions in the organization. Based on the extent and nature of the comments received in the two-week comment period, the guidance items were either revised and moved to “In Effect” status or placed back in “Draft” status for further revision.

From 2017 through 2019, staff in DCM supported the development and review of twenty-one distinct policy items. During this period, the core DCM editorial group developed items to ensure coverage of key policy requirements articulated in resources like the NDSA Levels of Digital Preservation and ISO 16363. DCMC reviewers demonstrated extensive engagement during this process, providing 430 total comments across the twenty-one policy items reviewed.

After establishing the DCMC in 2017 and developing a substantive body of policy material in 2018, further work went into implementing guidance in the DCMC and publishing a publicly available version by the end of 2019. With the completion of a full initial set of guidance and publication in 2019, objectives for 2020 and 2021 are primarily focused on sharing out about the resource and working to support full implementation of the policy and guidance across the Library. Producing a public-facing version of the DCMC offered an opportunity for holistic review of the twenty-one guidance items in order to develop a framework and structure for public presentation. In what follows we discuss that structure and the reasoning behind this approach.

## V. DIGITAL COLLECTIONS MANAGEMENT COMPENDIUM STRUCTURE

Although guidance elements affect actions throughout the digital content lifecycle, we organized

the elements of the DCMC into three main areas: Digital Formats, Inventory and Custody, and Access.

### A. *Digital Formats*

The “Digital Formats” area includes a summary of how the Library’s Recommended Formats Statement can be applied broadly to digital collections management processes, as well as overarching statements regarding how the Library approaches management of various digital preservation issues for digital formats, which includes preserving copies of content “as received,” creating digital surrogates, and maintaining an inventory of format types in the digital collection.

This area underscores the key role of the Sustainability of Formats resource in supporting digital preservation planning. Given the varied nature of the Library’s holdings, its digital infrastructure prioritizes the long term preservation of content in both its original bitstream and format. This approach allows the Library to provide users with a basic level of access to acquired digital content.

### B. *Inventory and Custody*

The “Inventory and Custody” area contains guidance related to how the Library inventories and tracks digital content, monitors data integrity, aligns with the PREMIS standard, packages and stores content according to the BagIt standard, and maintains collections security.

This area also defines the roles and associated responsibilities for digital content managers at the Library. Importantly, this guidance delineates responsibilities between digital content managers, and IT systems and staff.

### C. *Access*

Finally, the “Access” area describes the Library’s principles for providing access to digital content, as well as the metadata that is created to manage and preserve the collections. The Principles of Access discuss the institution’s commitment to provide users with the broadest possible access to digital content within limits set to ensure that the rights of content creators are fully respected. In support of these principles, metadata for digital collections will clarify the effort required to enable discovery and use of digital collection content.

Collectively, the items in the DCMC present the first full digital collections management policy and guidance resource produced by the Library of

Congress. This resource is envisioned as a living document to be refined and revised over time in response to organizational needs. Staff across the Library are working on initiatives that will further implement policy requirements of the DCMC.

## VI. ALIGNED DIGITAL CONTENT MANAGEMENT INITIATIVES

Establishing policy is one thing, but supporting the enactment of policy requires substantive work as well. Policy, resources, and technology collectively constitute infrastructure necessary to support programs [12]. In the next section, we briefly describe a series of projects and initiatives that have developed as a result of this effort, or have concurrently developed at the Library.

### A. *Initial Efforts to Review File Formats In Use*

The DCMC established the need to track and manage information about digital file formats in use in the collection. As a result, DCM staff consulted an inventory of purported file types (according to file extensions) in order to conduct initial analysis of file types in the digital repository [12].

The results of these assessments continue to play a role in the prioritization of work related to collection development. Significantly, the results demonstrate substantive IT business needs. Future digital repository infrastructure developments are required in order to manage robust data about a widely varied body of digital file formats.

### B. *Phased Approach to Inventory Verification and Reconciliation*

The DCMC policy framework centers management and maintenance of inventory records for digital collections content. In response, DCM staff initiated the institution's first full attempt to verify inventory records for digital collections content stored in the Content Transfer Service (CTS), the Library's primary digital inventory system.

DCM staff noted that certain digital collections material was not recognized by CTS, and so staff reconciled the existing records with data retrieved from underlying storage systems capable of directly reviewing digital content that CTS was unable to "see" (i.e. uninventoried content). This review enabled further DCM review of inventory issues, which was then reported to institutional stakeholders to initiate DCM-supported inventory reconciliation projects.

This work has also identified a series of key IT business needs to improve workflows and processes so that inventory problems identified can be rectified through future system design.

### C. *External Media Survey of Collections*

The DCMC established that, due to the high risk of loss, digital content on external media (CDs, DVDs, hard drives, flash drives, etc.) generally needs to be transferred into the Library's managed storage systems.

As a result of this policy, routing workflows for external media have been revised to ensure that, as content is received on external media, the digital content is routinely extracted during processing. A significant backlog of digital content received on external media remains to be fully processed into the collections.

An effort in 2013 identified over 300 TB of digital content on external media in the collections awaiting this kind of further processing [13]. In collaboration with staff from the Preservation Services Division, DCM staff are now updating the results of that survey, ultimately using this data to develop plans for extracting and inventorying digital content from external media sources.

### D. *Digital Library Practitioners Group*

In parallel with the development of the DCMC, the Library of Congress established a Digital Library Practitioners Group. The objective of this group is to support ongoing dialogue, learning, and professional development for Library staff working on supporting all aspects of the digital content lifecycle.

Through 2021 and 2022, DCM staff will coordinate the group's quarterly meetings, which provide staff from across the organization with opportunities to learn from (and with) their colleagues regarding projects that demonstrate the impact of DCMC guidance.

This kind of practitioner-focused program is critical for supporting the refinement of ongoing work to meet policy requirements.

### E. *Policy-Driven Approach to IT Business Needs*

Each area of ongoing development and refinement to digital content management practice and policy connects directly back to identifying and prioritizing IT business needs for further development.

With the DCMC established, all work around investments and prioritization for future work for digital collections infrastructure can now be linked and traced back to these policies.

## VII. GOING FORWARD

The launch of the DCMC is a significant development for the Library's digital collections management program. The integration of related projects into institutionalized programs governed by an appropriate policy framework represents a significant milestone in the organization's development [1].

By making the resource public and sharing a bit about its development and implementation, we hope to support other organizations in further refining their own programs. We hope that this resource will invite further suggestions from colleagues, and will continue to improve.

Ultimately, the ongoing maintenance and revision of the DCMC is a core part of efforts to improve our approach to digital collections management. The launch of the DCMC is part of the ongoing dialogue between the technologies, resources, and organizational roles that serve as enabling infrastructure to ensure long-term access to digital content.

## REFERENCES

- [1] McGovern, N., & Kenney, A. (2003). The Five Organizational Stages of Digital Preservation. In P. Hodges, M. Bonn, M. Sandler, & J. P. Wilkin (Eds.), *Digital Libraries: A Vision for the 21st Century: A Festschrift in Honor of Wendy Lougee on the Occasion of her Departure from the University of Michigan*.
- [2] McGovern, N. (2007). A digital decade: Where have we been and where are we going in digital preservation? RLG DigiNews, 11. <http://hdl.handle.net/2027.42/60441>
- [3] Owens, T. (2018). *The Theory and Craft of Digital Preservation*. Johns Hopkins University Press.
- [4] Flynn, M., & Zinkham, H. (1995). The MARC Format and Electronic Reference Images: Experiences from the Library of Congress Prints and Photographs Division. *Visual Resources*, 11(1), 47–70. <https://doi.org/10.1080/01973762.1995.9658318>
- [5] C. R. Arms, "Historical collections for the National Digital Library: Lessons and challenges at the Library of Congress," *D-Lib Magazine* April: Part, vol. 1, 1996.
- [6] A. Grotke, "Web Archiving at the Library of Congress," *Computers and Libraries*, vol. 31, no. 10, Dec. 2011.
- [7] E. Delfino and J. Mandelbaum, "eDeposit for eSerials: Current Work and Plans at the Library of Congress," in *Journal Article Tag Suite Conference (JATS-Con) Proceedings*, 2011.

[8] T. Westervelt, "To Lead to Learning, Not to Madness: E-Books and E-Serials at the Library of Congress," *The Serials Librarian*, vol. 72, no. 1–4, pp. 122–127, May 2017.

[9] J. Puccio, "Collecting Digital Content at the Library of Congress," *The Signal*, 21-Mar-2017 <https://blogs.loc.gov/thesignal/2017/03/collecting-digital-content-at-the-library-of-congress/>.

[10] K. Zwaard, "Digital Strategy at the Library of Congress," presented at the Federal Depository Library Conference, Washington DC, Oct-2018.

[11] C. R. Arms, Fleischhauer C (2005). Digital formats: Factors for sustainability, functionality, and quality. *Archiving Conference*.

[12] R. Guenther et al (2015). PREMIS Data Dictionary for Preservation Metadata, v. 3.0. <https://www.loc.gov/standards/premis/v3/>

[13] T. Westervelt. (2014). Library of Congress Recommended Format Specifications: Encouraging Preservation without Discouraging Creation <http://dx.doi.org/10.5703/1288284315583>

[14] Cooper, M., DesRochers, A., Johnston, J., Salas, C., & Owens, T. (2019). Extensive Extensions: Preliminary Explorations of File Extensions in Inventoried Digital Collections of the U.S. Library of Congress. In M. Ras, B. Sierman, & A. Puggioni (Eds.), *16th International Conference on Digital Preservation* (pp. 64–71).

[15] Engle, E., & Moryma, A. (2014, October 9). Digital Preservation at the Library of Congress. <https://www.loc.gov/item/webcast-6493/>

[16] Owens, T. (2018, April 26). Teaming Up! Digital Content Management Joins the Signal. *The Signal*. [blogs.loc.gov/thesignal/2018/04/teaming-up-digital-content-management-joins-the-signal/](https://blogs.loc.gov/thesignal/2018/04/teaming-up-digital-content-management-joins-the-signal/)

[17] Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, 43(3), 377–391.

[18] Aydelott, M. (2015). The Tangible Media Program at the Library of Congress. *Against the Grain*, 27(4). <https://doi.org/10.7771/2380-176X.7127>