LIBRARY OF CONGRESS

Library Services

Collection Development Office

Collecting Digital Content at the Library of Congress

February 2017

In January 2017, the Library of Congress adopted a set of strategic steps related to its future acquisition of digital content. The purpose of this document is to provide background information and a high-level description of the strategy.

The Library has been steadily increasing its digital collecting capacity and capability over the past two decades. This has come as the product of numerous independent efforts pointed to the same goal – acquire as much selected digital content as technically possible and make that content as broadly accessible to users as possible. In the past few years, much progress has been made, and an impressive amount of content has been acquired through several acquisitions methods. Further expansion of the Library's digital collecting program is seen as an essential part of the institution's strategic goal to: *Acquire, preserve, and provide access to a universal collection of knowledge and the record of America's creativity.*

The scope of the newly-adopted strategy is limited to actions directly involved with acquisitions and collecting. It does not cover other related actions that are essential to a successful digital collections program. These primarily include the following.

- Further development of the Library's technical infrastructure
- Development of various access policies and procedures appropriate to different categories of digital content
- Preservation of acquired digital content
- Training and development of staff
- Eventual realignment of resources to match an environment where a greater portion of the Library's collection building program focuses on digital materials

The strategy also does not cover digitization, which is the process by which the Library's physical collections materials (printed text, images, sound on tangible formats, etc.) are converted into digital formats that can be stored and accessed via a computer.

Vision

The adopted strategy is based on a vision in which the Library's universal collection will continue to be built by selectively acquiring materials in a wide range of formats – both tangible and digital. Policies, workflows and an agile technical infrastructure will allow for the routine and efficient acquisition of desired digital materials. This type of collection building will be partially accomplished via collaborative relationships with other entities. The total collection will allow the Library to support the Congress in fulfilling its duties and to further the progress of knowledge and creativity for the benefit of the American people.

Current Status of Digital Collecting at the Library

In the past few years, much progress has been made in the Library's digital collecting effort, and an impressive amount of content has been acquired. As the eDeposit pilot began the complex process of obtaining digital content via the Copyright Office, other efforts were making great strides toward the goal of acquiring and making accessible digital content. Digital collecting has also been integrated into a range of special collections acquisitions.

Information about some of the most impactful efforts is provided below.

> ***Web Archiving***: Begun in the year 2000, this program has harvested from the Internet over one petabyte of content, including United States Election archives that contain web sites related to all presidential and congressional elections starting with the contested presidential election of 2000. A variety of other subject-based collections have also been created.

> ***Electronic Resources***: The Library offers its users an array of purchased or leased electronic resources comprised of nearly 800 databases that contain e-books, journals (full-text and abstracting and indexing services), datasets, digitized primary source material (e.g., manuscripts, newspapers, etc.) and more. More than one million digital works are made accessible though these resources.

> ***Copyright eDeposit***: This is a program to acquire digital content via the Copyright Office. Thus far, the only items in scope are e-serials without print counterparts. The program is now being expanded to e-books without print counterparts and digital sound recordings that are available online only.

> ***Cataloging in Publication e-Books***: In exchange for e-book metadata the Library creates and sends to the publisher prior to a book's publication, one copy of the electronic book is sent to the Library. Established in 2012, the program has grown to over 700 participating publishers.

> ***Geospatial Data***: A geospatial data program is coming to fruition, with the development of a Geospatial Hosting Environment that will allow for full utilization of this collected GIS content.

> ***Collaborations:*** Working with external institutional partners, the Library has participated in efforts to acquire, preserve and make available digital collections. Examples include the National Digital Newspaper Program, StoryCorps and the American Archive of Public Broadcasting.

> ***Gifts, Purchases and Contributions:*** Born-digital photographs, videos and manuscripts are examples of the materials being received in the special collections as a result of gifts and purchases. The Veterans History Project receives digital versions of documents and photographs from contributors.

Audio and Video: The Library has begun development of a Live Capture system to record unpublished audiovisual transmissions and other programming. Limited testing of video transmissions from the U.S. House of Representatives and U.S. Senate chambers has also begun.

Assumptions and Principles

The Library's digital collecting strategy is based on a number of assumptions, including the following.

- The Library will continue to build its universal collection.

- The amount of available digital content will continue to grow at a rapid rate.

- Materials will continue to be published in physical formats for the foreseeable future.

- There will be much duplication in the marketplace, with the same content being available both in tangible and digital formats.

- The Library will be selective regarding the digital content it acquires.

- In selecting digital content, the Library will consider whether the content is already held and made available by one or more other trusted heritage institutions.

- There is a growing expectation of users to have wide access to digital content.

Likewise, there are a number of principles that support the strategy, including the following.

- The Library is developing one interdependent collection that contains both its traditional physical holdings and materials in digital formats.

- The Library will ensure that the rights of those holding intellectual property will be respected.

- Appropriate methods will be put in place to ensure that rights-restricted digital content remains secure.

- In acquiring new content, every effort will be made to adhere to the Library's Recommended Formats Statement since acquiring in those formats maximizes the content's sustainability and accessibility into the future.

- When desired content is available in both digital and tangible formats, the Library will apply various considerations (overall costs, preservation requirements, access restrictions, etc.) to determine which format – or formats – is preferable at time of acquisition.

- In acquiring new digital content, the widest patron access is the preference; however limited access, embargo periods and dark archive storage are appropriate in many cases.

- For electronic resource acquisitions that require payment, it is preferable to purchase and own the content permanently rather than to license access via a subscription.

Plan for Digital Collecting

Over the next five years, the Library intends to follow the framework of six objectives and their subordinate goals as described below.

Strategic Objective 1

Maximize Receipt and Addition to the Library's Collections of Selected Digital Content Submitted for Copyright Purposes

- Maximize the amount of e-serial content received via the eDeposit program.

- Expand the eDeposit program to include e-books and digital sound recordings.

- After implementation of the e-books and digital sound recordings expansion of the eDeposit program, systematically implement the acquisition of remaining e-formats (for example, digital photographs).

- Continue to pursue with publishers grants of special relief, obtaining digital files in lieu of physical copies, preferably for both serials and books, and resulting in publisher-provided online access to premium digital content for Library staff and patrons.

Strategic Objective 2

Expand Digital Collecting via Routine Modes of Acquisitions

- Maximize receipts of new publications via the Cataloging in Publication e-Books program.

- Continue to develop policies and implement workflows to enable routine acquisition of digital content via purchase.

- Develop policies and implement workflows to enable routine acquisition of digital content via exchange.

- Continue to develop policies and implement workflows to enable routine acquisition of digital content via gift.

- Develop policies and implement workflows to enable routine acquisition of digital content via transfer.

Strategic Objective 3

Focus on Purchased and Leased Electronic Resources

- Conduct a high-level assessment of the Library's entire electronic resources collection (subscriptions, purchased content and content acquired under special relief agreements).

- Fully identify all electronic resources as to their acquisition source – purchased ownership, subscription, gift, exchange, etc.

- Develop and implement a standard electronic resources license agreement.

Strategic Objective 4

Expand Use of Web Archiving to Acquire Digital Content

- Expand the use of web archiving as a standard acquisitions tool for digital content, such as openly available materials in various digital formats.

- Coordinate with NARA, GPO and other Federal agencies to ensure that the government's web sites are harvested, preserved and made available to researchers, with the Library taking on a special responsibility for harvesting Legislative Branch web sites.

Strategic Objective 5

Develop and Implement an Acquisitions Program for Openly Available Content

- Determine at a general level what types of openly available materials the Library wishes to acquire. Pilot technical methods for obtaining this content and develop appropriate guidelines for a permanent program.

Strategic Objective 6

Expand Collecting of Appropriate Datasets and Other Large Units of Content

- Selectively acquire large sets of content available in electronic format only.

- Provide datasets collecting guidance through a new Collections Policy Statement (CPS), Supplementary Guidelines document and/or through the incorporation of such information in other CPS documents as appropriate.

- Determine the manner in which research by our patrons using datasets will be supported, such as with special software and licensed or purchased content.

- On a pilot basis, acquire sample large sets of content for the purpose of providing bulk download and text mining capabilities to the Library's users.

Questions and Comments

Any questions or comments about this strategy or any aspect of the Library's collection building program may be directed to the Collection Development Office (cdo@loc.gov).