WORLD LIBRARY
AND INFORMATION
CONGRESS:
78TH IFLA GENERAL
CONFERENCE
AND ASSEMBLY

IFLA

2012 Helsinki

LIBRARIES NOW!
INSPIRING
SURPRISING
EMPOWERING

Inclusionary Tactics for Print and Digital: Managing and Remembering; Considerations to improving access, sharing resources and providing preservation of print collections

Jeanne Drewes
Library of Congress,
Washington DC, USA

Session:

139 — *Treating print in an increasing digital collection: issues, dilemmas and directions* **— Acquisition and Collection Development**

Abstract:

The Library of Congress (http://www.loc.gov/index.html) as a member of HathiTrust (http://www.hathitrust.org/) provides access to digital copies of many titles also held in print. Learn how the Library is working to improve access through providing digital and print choices from the online catalog (ILS); how they are managing physical condition by treating and binding damaged print copies when they are selected for digitization; how the Library is improving space for needed physical copies by storing physical copies that duplicate digitized titles differently after the digitization process is complete. Efficiency and cost effectiveness is a high priority when refining existing work flow process or when developing any new work flow process and the Library of Congress has spearheaded several projects that integrate the use of small computer applications used in conjunction with the ILS to improve accuracy, and speed of processing titles. The information required for effective decision making for processing damaged materials will be documented. How to routinely process materials for digitization, reformatting to paper copy, provide treatment or house appropriately will be delineated as currently applied to general collections within the Library of Congress. More efficient print collection management and speedy access through the on online catalog to the most appropriate format is the goal of this process and the focus of this presentation.

Learn the applications developed to reduce time for decision making as titles are selected for digitization. Workflow and thought process will be a major take away for this presentation.

BACKGROUND

The Library of Congress is the nation's oldest federal cultural institution and serves as the research arm of Congress. It is also the largest library in the world, with millions of books, recordings, photographs, maps and manuscripts in its collections with more than 151.8 million items on approximately 838 miles of bookshelves. The collections include more than 34.5 million books and other print materials, 3.3 million recordings, 13.4 million photographs, 5.4 million maps, 6.5 million pieces of sheet music and 66.6 million manuscripts. The Library receives some 22,000 items each working day and adds approximately 10,000 items to the collections daily. The majority of the collections are received through the Copyright registration process, as the Library is home to the U.S. Copyright Office. Materials are also acquired through gift, purchase, other government agencies (state, local and federal), Cataloging in Publication (a pre-publication arrangement with publishers) and exchange with libraries in the United States and abroad. Items not selected for the collections or other internal purposes are used in the Library's national and international exchange programs. Through these exchanges the Library acquires material that would not be available otherwise. The remaining items are made available to other federal agencies and are then available for donation to educational institutions, public bodies and nonprofit tax-exempt organizations in the United States. (http://www.loc.gov/index.html accessed May 19, 2012).

Since 1962, the Library of Congress has maintained offices abroad to acquire, catalog and preserve library and research materials from countries where such materials are essentially unavailable through conventional acquisitions methods. Overseas offices in New Delhi (India), Cairo (Egypt), Rio de Janeiro (Brazil), Jakarta (Indonesia), Nairobi (Kenya) and Islamabad (Pakistan) collectively acquire materials from more than 60 countries and acquire materials on behalf of United States libraries participating in the Cooperative Acquisitions Program. The Library is also collaborating with institutions around the globe to develop a World Digital Library.

The Library of Congress (http://www.loc.gov/index.html) as a member of HathiTrust (http://www.hathitrust.org/) provides access to digital copies of many titles also held in print. The Library of Congress' membership in the HathiTrust provides rich content that is available to researchers. The HathiTrust Digital Library contains digital surrogates for 5.7 million titles linked to more than 10 million volumes. A large portion of this material, 2.9 million volumes, is available in the public domain. The Library has contributed nearly one hundred thousand public domain volumes to the Trust's digital collections. All of these materials can be accessed at: http://www.hathitrust.org/.

The public domain material can be viewed by anyone accessing the HathiTrust Digital Library. Copyright protected materials have restricted access, but are searchable. As a member of the HathiTrust, the Library of Congress has

enhanced access to public domain materials held in the collection. The Library has loaded links to the Library's integrated library system (ILS) to all our physical titles that are also available through this membership. This provides seamless enhanced access for any Library user searching the Library's collection where all copies are shown, both physical for on site users and digital for both onsite and remote users thus providing multiple formats for the same title viewable from a Library computer or the website access.

The Library of Congress has made digitized versions of collection materials available online since 1994, concentrating on its most rare collections and those unavailable anywhere else. The following services are your gateway to a growing treasury of digitized photographs, manuscripts, maps, sound recordings, motion pictures, and books, as well as "born digital" materials such as Web sites. In addition, the Library maintains and promotes the use of digital library standards and provides online research and reference services.
The Library provides one of the largest bodies of noncommercial high-quality content on the Internet. By providing these materials online, those who may never come to Washington can gain access to the treasures of the nation's library. Such online access also helps preserve rare materials that may be too fragile to handle.

This process of adding digital content continues through a variety of initiatives, but for the HathiTrust there is a contractual agreement to create digital surrogates for public domain materials owned by the Library and scanned in-house by Internet Archive. Because of the number of volumes scanned and the value of materials, the Library has provided space on-site for the Internet Archive as contractors. This provides security control for materials and provides added preservation by reducing handling for these older materials to be scanned for access.

 After these electronic files are created they are added to the Trust's digital links and the physical titles are prepared for off site storage. Off site storage provides good storage conditions which include appropriate temperature and humidity put also provide appropriate support for each volume through shelving by size.   This provides needed space for new collections coming into the Library and in copyright, while not diminishing access to older volumes. Indeed access to these older volumes is greater than newer volumes that can only be used onsite in one of the many reading rooms. Should the older stored physical volume be needed it can be requested and returned for onsite patron use. The Preservation Directorate at the Library of Congress works closely with all aspects of this digital program as well as others within the Library.

Remembering
"Libraries and archives standardized microfilming practices during the latter third of the twentieth century. The Library of Congress published two guidelines in the 1970s, followed by the publication of a number of standards in the decades that

followed. (Slightly out of date references to these standards will be found on Web pages provided by OCLC and the Library of Congress.) Many of these standards were first developed under the auspices of ANSI and AIIM, and today they are being extended and integrated under ISO auspices. ANSI/AIIM standard MS23 was first published in 1979 and has seen several revisions, the most recent in 2004. For many years, it has been the central document for the field. ISO standards 6199 and 6200 incorporate many of MS23's specifications. ISO's Web site describes standard 6199 as specifying (1) "procedures that enable a camera operator to produce microfilm of appropriate quality of presentation and legibility, capable of yielding scanned images of acceptable quality" and (2) "methods for microfilming documents on 16 mm and 35 mm silver-gelatin microfilm, including orientation of images on microfilm, use of non-image areas and information required to facilitate identification of the microfilm."
(http://www.digitizationguidelines.gov/term.php?term=microfilmstandards accessed May 19, 2012)
The other very important piece that was an outgrowth of the microfilming was the National Register of Microform Masters (NRMM) Retrospective Conversion Project. "The Association of Research Libraries with funding from the national endowment for the Humanities created Master File creation. The NRMM Master File, the largest single file of records for microform masters in the United States, consisted of reports for microform masters sent to the Library of Congress (LC) by more than 200 different libraries, historical societies, and publishers between 1965 and 1983. NRMM was published annually by LC until 1984 with one cumulation covering the years 1965 to 1975. For all libraries, the automation of the NRMM files improves access to preservation microfilm masters, ensures against duplication of microfilming and reduces bibliographic searching time and costs. It is no longer necessary to search in the printed, multi-volume NRMM. The online NRMM makes possible a one- step search, which dramatically reduces the amount of staff time needed for pre-filming searches. The danger of missing a record is also minimized since the machine-readable records are accessible by several access points."
(http://www.arl.org/preserv/presresources/Microform_masters.shtml, accessed May 19, 2012). Remembering the importance of providing a means to avoid duplication when a large number of institutions were supporting the need to microfilm damaged materials to provide access, provides a roadmap for digital projects.

Managing
For any digital project there are overarching issues to be addressed. At the Library of Congress for the Internet Archive, HathiTrust program digitizing materials out of copyright and held in the general collections there are issues of duplication of effort, stabilization of damaged items before scanning, and consideration of relocation of physical volume after scanning.
　　　Duplication of effort
While the Library of Congress has a multi faceted approach to digitizing collections there is also the desire to avoid duplication of effort with other efforts

world-wide. While in the being stages of digital conversion projects in the early 1990's, duplication was not an issue, in today's world where many cultural institutions provide access to parts of their collections through digital surrogates, there is a need for discovery of titles in digital format before moving titles into a reformatting digital workflow. This is most needed for general published works out of copyright where the title may be held by a larger number of institutions.

With the membership to HathiTrust and the additional links from the Integrated Library Catalog (ILS) to those links it suddenly became possible to check with a minimum of effort to discover other copies of titles in digital format. At first the idea seemed not worthy of the time, given the scope of the Library's collection, so a pilot was conducted to discover the amount of overlap. Initially a staff person with excellent searching experience was tasked with searching a number of web based programs including Google Books, Hathi, Internet Archive and a general web search by title. The results of this pilot showed that over 50% of titles searched were already available in digital format. While the pilot was not large, (50 randomly selected titles in a reformatting workflow queue) it was clear that the searching had merit. The problem was the amount of time it took to search, check the link to assure the exact match, and that the digital copy was truly available beyond just the title page. Some projects limit access to the full text so a number of clicks into the document were required for the search reliability. A programming proposal was put forward to create an application that would work with the ILS to discover all holdings records for an individual title where that title's bar code identifier could be scanned. The programming was approved, the programmer was assigned and he quickly created an application because this task was similar to other applications that were already in production. A test was set up after programming was completed to assure accuracy. The time that elapsed between the first pilot and the first testing of the "checking" program was only about 6 months. The new application works in an elegant way. It searches across the ILS for holding records, which show multiple copies and copies in all formats where the copy links to the bibliographic record. For catalogers familiar with MARC records this is the 856 field.

This program is now used to review all general collections titles selected for digital based on copyright year, (pre 1923 for U.S. imprints) The average time to check a truck of 200 plus titles is less than an hour and the average match is anywhere from 30 to 60%. Those numbers warrant the continuing review before moving titles into the digital processing. The appropriate step to do this review works well because when a title is discovered to be in digital format that title is then moved to a nearby workflow for post digital processing, which prepares materials for off site storage. While this process may seem like a waste of time to do after the general selection has been made because digitized titles are further processed by then moving to off site storage this actually assists in the discovery of digital titles that can be moved to a less primary location. The strategy is to relocate off site those titles available in digital format since many patrons now prefer the digital version over the physical volume.

The condition of any particular physical volume is also an issue for digitization. Will the volume be damaged by the process of digitizing? Will the current binding allow proper access to provide a good scan image? Either of these conditions needs action before scanning can begin.

Before Internet Archive scanning work can begin there is a review of the physical volume. Should that review reveal the need, a Preservation directorate technician provides stabilization for physical volumes that otherwise might be damaged from the scanning process. This includes such action as paper mends, replacement pages for missing pages, reattachment of covers, and loosening of tight binding to assure good image capture while retaining binding structure. While stabilization is not as time consuming or as detailed as a full treatment it does provide what is needed to preserve the volume both for the scanning action and for possible future use. However it is the general policy to limit resources appropriately, meaning that the predominant resource goes to the digitalization not to the full treatment of a volume that will predominantly be used in digital format.

After scanning action all items are reviewed to assess the need for additional housing, in particular for fragile origin bindings, or minor stabilization to make materials fit for limited future use and for travel to the storage location. Since these items might be requested for future onsite use, additional information is added to the ILS record as part of the processing for off site location.   Items with original bindings are noted in the record so that information is available should the item be requested for onsite use. Information about original binding is not generally part of the Bibliographic record or item record. Original binding structure is considered to be the most likely need to see the physical volume when a digital version is also available. Information about deacidification, another preservation action that has been a standard process within the Library for fifteen years is also noted if it is not already in the record. This review of the record before relocation provides an added process to perfect the record and catch any discrepancies that have occurred over the long history of cataloging or these materials.

While the Preservation Directorate does not catalog materials they do work closely with the catalogers to discover and correct cataloging irregularities for materials in the workflow of the various sections. The actions taken by the preservation directorate are recorded in the record of the item treated, bound, or deacidified. While this information is not shown to the public it is visible to staff at the Library. The ILS provides a function for batch updating that is used to document the actions taken. In a similar manner the contract staff who prepare materials for off site storage are able to use time saving batch update or programmed keys that efficiently insert appropriate information in the record.

Off Site Storage
Materials that have a digital surrogate and link to that file in the ILS are prepared
for off site storage by updating the ILS record to the new location for the physical
item. The Library of Congress since 2005 has maintained a state of the art, high-
density storage facility thirty miles from the main buildings on Capitol Hill. Other
high density storage is currently used for the workflow of items coming from the
digitization project of the general collections. The Preservation Directorate works
with custodial units to prepare materials for transport and storage,
(http://www.loc.gov/preservation/about/conserv/storage/index.html accessed May
19, 2012).

Decision Making
The Library of Congress has working groups which cross divisions and
processing units to provide the appropriate decision making for cross divisional
workflow. At the beginning of the Internet Archive contract appropriate managers
were brought together in a working group to design the process of retrieval of
materials, staging, record updating, treatment of damaged items, and digitization
handling and review. All steps were developed as part of this overlap of
responsibility and to provide for more seamless workflow across units. One
division has responsibility for selecting materials from the shelving areas and
bringing those materials to the digital scanning work area. Another unit does the
now computerized review to assure the title has not been scanned and that it is
in the best condition for scanning. Another unit, in this case a contractor, does
the scanning, another unit provides the quality assurance of the scanned image
before the physical volume is reviewed for any additional needs before the last
group processes the item for off site storage. And most importantly on going
communication across these groups allows for improvement and refinement of
the system. The pilot which revealed the need for searching before scanning is
now part of the workflow, although it was not thought of at the beginning of the
Internet Archive scanning project.

It was in 2007 that the Sloan Foundation provided a grant to the Library of
Congress to digitize thousands of public-domain works, with a major focus on at-
risk "brittle books" and U.S. history volumes. The project, "Digitizing American
Imprints at the Library of Congress," included the scanning of brittle volumes
which past digitization projects have shied away from because of the condition of
the materials, but "Digitizing American Imprints" was a demonstration project of
best practices for the handling and scanning of such vulnerable works."
(http://www.loc.gov/today/pr/2007/07-020.html accessed May 19, 2012). This
grant became a routine workflow with internal funding that served to extend the
subject areas to be scanned and also served as a means to open needed space
through digital access and off site storage of the physical volume. Such actions
serve the patron through multiple access points both online and within the
reading rooms, preserve the collections with multiple formats and safe storage,
and document actions within the ILS for future working groups to make decisions
about new technological avenues. The knowledge of how microfilming was

accomplished across many institutions while reducing duplication helped to inform the digitization workflow that is currently in place at the Library of Congress.

Author:

Jeanne Drewes is Chief, Binding and Collections Care Division, Preservation Directorate at the Library of Congress. She worked to help develop the workflow to eliminate duplication of effort. The processes described were developed in collaboration with the Internet Archive project within Library Services in the Library of Congress and CALM: Collections Access and Library Materials Division. The author works at the Library of Congress, United States.

Contact: jdre@loc.gov

Author Bios

Jeanne Drewes has been Chief of the Binding and Collections Care Division since 2006 and Mass Deacidification since January, 2011 at the Library of Congress. She has taught numerous courses in collections care, treatment and binding for general collections. She was previously head of preservation programs at Michigan State University Libraries and Johns Hopkins University library.   She works directly with the general collections digitization projects and was a partner in the design and development of the systems currently in use. She has been a member of the IFLA Preservation and Conservation Section from 2006 until 2012 and served as the newsletter editor for the section. She has been an active member of the American Library Association's preservation section for over 20 years.