



2019

## Shared Collection Development, Digitization, and Owned Digital Collections

Michelle M. Wu

*Georgetown University Law Center*, [mmw84@law.georgetown.edu](mailto:mmw84@law.georgetown.edu)

This paper can be downloaded free of charge from:  
<https://scholarship.law.georgetown.edu/facpub/2133>  
<https://ssrn.com/abstract=3328624>

---

Collection Management, April 2, 2019.

This open-access article is brought to you by the Georgetown Law Library. Posted with permission of the author.  
Follow this and additional works at: <https://scholarship.law.georgetown.edu/facpub>



Part of the [Collection Development and Management Commons](#)

Michelle M. Wu

Georgetown University Law Library

## Shared Collection Development, Digitization, and Owned Digital Collections

Abstract: While library models already exist for sharing physical materials and joint licensing, this paper envisions an aspect of future collections involving a national digital collection owned, not licensed, by libraries. Collaborative collection development, digitization, and digital object management of owned collections can benefit societies in multiple ways, from expanding access to users otherwise unable to reach these materials, to preserving content even when disaster strikes, to reducing duplication of effort and expense in collection or digitization. This article will explore both the benefits of and the challenges to this type of collaboration.

*Keywords: collaborative collections, digitization, copyright, preservation*

Libraries exist in part to preserve culture and history for all generations, but the sheer amount of information available in the world today is overwhelming (Marr 2018), and even when limited only to published materials (Bowker), exceeds the resources available to any given library to acquire or preserve. Another persistent information challenge rivals the first, in the form of multiple formats. These formats are not viewed equally by users, and there are as many format preferences as there are formats. These will continue to change over the years, even more rapidly than they have in the past, and the ongoing dilemma for libraries will lie not only in meeting user preferences but in how to make all of that information equally available and discoverable.

Over centuries, libraries have adopted many practices to address these types of access and discovery issues, including the creation of circulating libraries, the use of interlibrary loan (ILL), the addition of discovery layers, and the move towards consortium licensing and shared collections. One of those practices has been to replace or supplement collections through the licensing of data stores, which has benefits in short-term access but also many long-term risks.

Materials subject to licenses can be unilaterally retracted (Stone 2009), can expire after a publisher-determined number of uses (Bosman 2011), can be unavailable to libraries (Crook 2011), can be subject to differential pricing (Coldewey 2012), and can restrict what libraries can do with the works (e.g., interlibrary loan, preservation copies). None of these actions are possible with physical items. Licensing terms, in many ways, subvert the traditional balance of copyright between a copyright owner's interest and the public's (Katz 2016) and limit a library's ability to access and preserve works. While LOCKSS and other efforts have addressed some of the preservation issues, a great many licensed works are still subject to too many restrictions for full library use.

While this paper does not suggest abandoning or replacing all licensing, it does recognize that there may be a better tool for libraries to gain electronic access without losing control of their materials. It also recognizes that many of a library's holdings may never be made available in electronic form by publishers. If libraries do not develop a plan to address transitioning their physical holdings, their communities will be at a disadvantage as the nation becomes increasingly populated with digital natives. Many physical materials, while technically available, will become practically undiscovered in that type of world.

A collaborative network to digitize physical materials and distribute their electronic equivalents through digital means could help to address all of the issues above. First, by

spreading the work of collection across all libraries, together we can build a richer collection than any of us can afford individually. Knowledge is power, as Francis Bacon reportedly said. As library budgets decline and costs rise, narrowing collections will have not only local but potentially national consequences. Communities using them may be unable to find information they need, or collections overall will reflect (and preserve) only majority viewpoints. If each library is charged with collecting one less common topic or viewpoint and these collections are made available to all, then we expand access and ensure preservation of more voices.

Second, technology allows for faster, more-automated transfer and retrieval than mail-based ILL, so sharing materials in digital form becomes less costly and shared collections become more useful. Third, by digitizing works and making them full-text searchable, we can level the searching playing field. The average user has difficulty understanding how to search a bibliographic record with defined authority fields but is familiar with searching full-text. While this may not be the most effective way to search in all instances, given the rise in search engines, it is the most likely user approach. Converting materials where only a bibliographic record is available to full-text documents will present the materials to users in the way they expect. Last, by digitizing the titles themselves, libraries retain ownership and control of the content. They can continue to use them in new forms in the same manner that they used the original, without fears that they will be limited in efforts to preserve or share content that they have acquired,

In its ideal form, this collaborative network would be a global, universal library, collecting materials in all languages and from all cultures, open to all comers and preserving all traditionally published information as well as selected self-published, archival, or unpublished materials (e.g., letters). Not only would materials be available online but every library would have technology available so that users would be able to access them. Due to differing copyright

laws across the globe and the practical challenges of providing technology to every corner of the world, though, this paper focuses on a narrower goal. I will begin by discussing briefly the backdrop for digital libraries, then propose a new model, and finally explore both the benefits and challenges to adopting the model for future collections.

### BACKDROP FOR A UNIVERSAL DIGITAL LIBRARY

The move towards a broad digital library model is styled as an evolution instead of a transformation because so many aspects of the model already exist today. Others have had similar ideas --- Google Books, HathiTrust, DPLA, and Internet Archive are some of the most notable players --- though with different conceptualizations and paths.

Google decided to build its library with limited collaboration; it partnered with major library systems to gain access to their materials, but it did not involve the libraries in planning, organization or usability testing (Google). Its agreements with major university and public libraries instead focused on the rights to digitize and use the resultant images (Wyatt 2005; Band 2006), and its intent was fairly narrow – to provide a searchable database. Despite its shortcomings, Google Books succeeded in building the largest library database available at that time, even though it itself was not a library and many of the underlying materials could not be viewed in full. It provided access to snippets through its search engine and developed data mining tools like ngram (Google Books Ngram), both actions which helped users and courts understand the value that aggregated information could provide above and beyond what could be found in individual volumes.

Another effort, HathiTrust, began with the libraries who had partnered with Google and who had received digitized copies of their works in exchange for the partnership (Young 2008). Due to restrictions in the agreements that the libraries had signed with Google (University of

Michigan), they were constrained in how they could use these materials and their approach was consequently more conservative. Materials in the public domain were freely available to all, digitized works were available to visually impaired users under certain conditions (HathiTrust), but the vast majority of users had much less access. They could search all digitized materials to identify the books that held their search terms, and HathiTrust's search engine would produce results verifying the presence of the word in given book(s), the frequency of appearance, and the page number(s) on which the word(s) appeared. HathiTrust did not display snippets in the same manner as Google.

Both HathiTrust and Google were sued for these projects by authors, authors' groups, and/or publishers for copyright infringement, but in both cases, courts determined the uses to be fair use (Authors Guild 2014; Authors Guild 2015). The decision in Google was particularly notable, as Google had never owned the books it digitized and had not received permission from copyright owners for the reproductions. For a more detailed discussion of these cases, see Wu 2016a.

The other two projects to mention are closer in concept to the one described by this paper. The Digital Public Library of America (DPLA), at least during its earliest phases, contemplated building a national digital library which would be accessible through a single portal (Singer 2011). It would involve digitizing items (or facilitating the digitization of them), preserving them, and making them available to all users. As it evolved, the DPLA focused more on the portal aspect than it did on digitization and now features primarily free, full-text items provided by partners and free from copyright restrictions.

The Internet Archive (IA) is the final major U.S. player in testing out the concept of a universal digital library. Its vision is one that sees the widespread replacement of print items

with their digital equivalents in libraries across the United States, with shared effort in the digitization process (Kahle 2017). It began its effort in 2010 through circulating digitized items through Open Library in collaboration with libraries. It has since expanded its reach by partnering with some publishers and by providing a platform for controlled circulation to other libraries interested in lending digitized works, even if those works are not available through Open Library.

### PROPOSED MODEL

The model proposed today is not that far removed from DPLA's initial vision and IA's current vision, but the approach is somewhat different. The model proposes broader access to collections and a leveling of the searching field through a national, collaborative library (NCL), in which the nation's libraries (1) share collection development, (2) divide up responsibility for digitizing works that they own or acquire and (3) share these digitized materials through a centralized NCL system.

As materials are digitized by individual libraries, they could be shared with libraries with the same holdings. In other words, where one library funds the digitization of a work, another library need not digitize their own copy to use a digital replacement. Instead, they could use the copy already digitized replace a physical copy that they own.

This concept is probably best understood through illustration. Take Library A, which owns a single copy of a print title. It digitizes this title, takes the print copy out of circulation, and instead circulates the digitized copy through a central digital-rights-management (DRM) platform. Such a platform would be configured so that only one user can borrow the item at a time, mimicking how print circulation works. If the library owns five copies, it would similarly have the option to circulate five digital copies simultaneously, as long as it takes the print copies



off its shelves. Digitizing and lending in this manner is a form of controlled digital lending (CDL) (Bailey et al. 2017), ensuring that a library never lends more copies than it owns.

Now take Libraries B, C, and D. Let us say that they all own this same title. Since Library A has already digitized the title, the centralized platform would enable Libraries B, C, and D to use the digital copy that Library A produced to replace their print copy if desired. They would need to comply with general CDL principles (e.g., removing the print copy from their shelves) in using the digital copy. Each library continues to use only the number of copies that it owns, but they can all benefit from the efforts of digitization undertaken by their peers. They can circulate the digital item through their own system, only to limited users, or could share the materials with the nation through NCL.

The digital format is used over any physical format for many of the reasons detailed above: access, searchability, and ease of transfer. It has other advantages as well. It takes up less physical space, and standard procedures for electronic backup and redundancy provide greater guarantees of survival in case of natural disasters. It also will be easier to migrate materials from a single form to other forms – which is inevitable with today’s rate of technology evolution– than trying to migrate all original items to a new technology each time there is a format change.

There are losses from this format selection as well, the most obvious of which is that an e-book cannot convey the texture, smell, and features of a physical artifact. However, in assessing overall risk and reward, the potential benefits outweigh the costs (see Benefits, below).

Online circulation itself also carries benefits, from reducing the number of copies lost, to preventing patrons from keeping titles that are overdue, to reducing ILL shipping and personnel costs. Tasks that require human intervention with physical circulation, such as reclaiming an

overdue book, can be automated. As access to e-books automatically expire, users will also be freed of overdue or replacement fees.

An NCL would require some centralized funding, through membership fees or other means (see Proposed Model – Funding below), for acquiring and digitizing core materials that would be useful for all libraries (e.g., reference materials), setting up scanning centers, and/or developing and maintaining a centralized system. That system would provide a single interface (Wu 2016b), even where member libraries use different integrated library systems (ILS), and the unified interface would negotiate checkout between each library’s local system and the universal one.

#### PROPOSED MODEL: COORDINATION

A logical approach would be to rely on an existing organization (e.g., Library of Congress, Digital Public Library of America) to direct the effort, but this may increase complexity and litigation risks for the central entity unnecessarily (see Challenges below). Instead, a central organization could provide the system framework, set basic use rules (e.g., what a library has to certify to use a digitized copy through the system), and create tools to facilitate information sharing. Taking a coordinator’s role as opposed to a controlling role has the added benefit of avoiding the natural conflicts that can come into play when member entities have missions that might conflict with a controlling entity’s.

A coordinating entity would have four primary responsibilities: administrative oversight of costs and fees; system development and maintenance; collection analysis; and scanning in materials not acquired by other participating libraries. If centralized scanning centers become necessary, management of those would be an additional responsibility.

A largely decentralized system will have success in meeting the broader collection and preservation goals of the whole only if the coordinating entity provides core research to facilitate member libraries' efforts. For example, the coordinating entity should do a full collection analysis of member libraries' holdings to (1) determine where digitized copies have already been made, (2) assess the availability of those copies for use by the NCL (including talking to Google and the original partner libraries to see if there is any possibility of renegotiating the original agreement on Google images for broader use), and (3) identify the titles that have not been digitized or are unavailable for use so that any new digitization efforts can focus only on those. With so few resources, every effort should be made to avoid duplication where possible.

The coordinating entity would also take the lead in developing the centralized system --- hopefully partnering with a recognized technology expert -- and establishing any expert library panels necessary to provide input on the system's effectiveness. I am not suggesting a voting board, as that would slow down decisionmaking and likely is unnecessary considering the system's anticipated flexibility (see Proposed Model – Systems section).

The downside to a decentralized approach is that it provides fewer guarantees in meeting the broad collection and preservation interests of the nation. A centralized controlling entity can mandate the broadest possible acquisition and digitization of materials, building a complete collection for current and future users. Decentralized collaboration will mean that individual libraries are likely to concentrate primarily on convenience for their respective user bases, and it is more likely that pockets of information will continue to be missed. However, if the coordinating entity continues to do collection analyses as libraries contribute materials, they could devote some of the centralized funding to acquiring and digitizing materials otherwise not captured by contributing libraries.

## PROPOSED MODEL: SYSTEMS

The centralized system would provide three core functions: coordinated acquisitions, circulation (including digital asset management, rights management and DRM), and a discovery layer. The effectiveness of these functions would require dynamic information exchange between the centralized system and libraries' ILSes if they want to continue to maintain their own.

In acquiring materials, each participating library would create an order record within their ILS and upon creation, the library would be informed by NCL's system (1) if another library owns the title already, (2) if another library has the title on order, and (3) if any of the owned/ordered materials are tagged as being available to the NCL. The ordering library would then be able to decide if it still wants to purchase the item or if demand will be such that they could reasonably rely on the existing copies through the NCL or ILL. The library is likely to continue to acquire popular materials for local needs, but through this mechanism, it would have the opportunity to leverage funds reserved for lesser-used materials and acquire a different, unique title for its patrons and the NCL instead of duplicating existing holdings. The centralized system could even interact with publishers' catalogs to highlight other books in the same subject that are not yet owned or ordered by any library in the NCL.

As a library places an order, it could simultaneously designate the title as being available for NCL users. Alternatively, it could set an overall default rule for its materials (e.g., each title is entered into the NCL one year after receipt). It would also have an option to add items to the NCL after the item is received (e.g., for very popular items where they are unsure when demand will lessen), but noting the intent at the time of order would be the most useful to other NCL libraries in assessing their own orders. Collection development otherwise need not be centralized, overcoming the usual issues with group decision-making.

If libraries have set umbrella rules for their new acquisitions (e.g., all added to the NCL one year after digitization), those rules are viewable by the general system, which will automatically “ingest” the item at the proper time and notify the owning library to withdraw the print item (if still in circulation) from general access.

Circulation could be initiated either from the central system’s interface or through an individual library’s ILS, and there would be a dynamic feed between the central system and all participating libraries’ ILSes, so that holdings on both would reflect current availability information. The coordinating entity would establish default rules, determined through identification of common elements across member libraries, but allow individual libraries to deviate from them. For example, if all member libraries have loan periods two weeks or longer, then the default period is two weeks. If a library notes in the central system that they will support longer loans (e.g., four weeks), then when users check out materials owned by that library, they will get the longer loan period. The system would alert users who check out materials with shorter loan periods (e.g., reserve) before they check out the item, so that they would not be surprised by the shorter-than-usual turnaround. No-loan materials (e.g., reference) would be viewed only online, with simultaneous use limited to the number of copies owned by or shared with the NCL. Default rules would apply unless a library actively overrides them.

The model described here is one that clearly prioritizes increased access over all else, including consistency for users, recognizing that this choice may cause inconvenience and make the system less attractive to users than it might otherwise be. The reason for this priority determination is because the overarching interests are in access and preservation; if libraries are discouraged from contributing works to the NCL because they do not want to adopt the defaults,

then the NCL's ability to preserve the materials long-term or to provide access to communities that otherwise do not have access is reduced.

As digitization costs fall and DRM platforms develop, this model should result in net benefits outweighing the costs for every member. Each library would have access to more materials than it owns, but each library could also preserve benefits from its investment in the materials it has acquired for their own community (e.g., delay sharing with the NCL until a title is less popular). While there is some inconvenience for members of libraries with fewer resources, they should still have access to more materials than they would have otherwise.

Patrons using the centralized system as a discovery layer, as opposed to their own ILS, can login to see titles that are available not only through the NCL but also at their local library. They would have access to a feature to "build" their own virtual libraries, organizing them in whatever manner makes the most sense to them. They could keep these libraries private, make them "public" for all to view, or share them with select individuals, as desired. In this manner, researchers working in similar areas could build shared libraries together to make it easier for scholars studying in the same field.

Their virtual shelves would essentially hold shortcuts to the titles they have selected, and when a user clicks on a title, the system would display the first available copy at the time with an option to check it out. If no copies are available, the system would notify the patron of this, allow them to place a hold on the first available copy across the entire NCL system or their local library, and would provide links to non-library sources (e.g., Amazon or Google books print-on-demand), where available, for those too impatient to wait for a library copy.

The accompanying DRM platform would have a rights management aspect similar to the one used by HathiTrust, permitting tagging of different types of materials (e.g., public domain)

and setting rules to go with each type (e.g., open access). It would also isolate publication dates so that titles could be automatically released into the public domain as copyright protection expires. It would also provide the standard elements of multiple-copy preservation, security, and periodic stability reviews.

Materials when lent would be controlled through an Overdrive-like DRM tool, able to limit uses to the number held by or shared through the NCL. The CDL platform would support the general functions available in common e-book readers today --- highlighting, bookmarking, searching --- and would also allow limited cutting and pasting (e.g., one paragraph) for citation purposes. Any local notes (e.g., highlighting) would only be visible to the individual user and would be stored on the local device, but would reappear if the title were ever rechecked out by the same user on the same device. Since the platform would be controlled by the NCL and not a for-profit entity, it could preserve reader privacy to a greater extent than current platforms allow.

Where a title has already been digitized and shared with the NCL, another participating library buying another copy can share their copy by changing the number of shared copies to reflect the additional copy upon receipt. It would do this by connecting the order number from its ILS to the NCL title (for accounting purposes) and noting how many copies were purchased and should be shared. All materials digitized would circulate primarily in digital form; the print materials would be stored in closed stacks to ensure that no more copies are used than are owned. If needed (e.g., a defective image), the physical item could be used, but the digital copy would have to be suppressed from patron view during use, to ensure that the library continues only to use the number of copies it owns.

#### PROPOSED MODEL: OBLIGATIONS OF MEMBER LIBRARIES

Member libraries would have four core responsibilities for new materials: acquiring materials, identifying materials to be shared with the NCL, digitizing materials (or acquiring a digital copy), and setting local rules for their materials within the NCL.

The logistics of the first two responsibilities were discussed in the section immediately preceding this one. Digitizing materials would only be necessary for materials where no other NCL library has already digitized a copy. The first acquiring library of a title is the one that would be charged with digitization; if certain libraries are always the first acquirers, the digitization could be subsidized through central funding and/or outsourced, as the costs would be for the benefit of all NCL libraries. If outsourced, the library could have the materials shipped directly to a designated scanning facility where materials would be prioritized upon need (e.g., an item that has someone on a waitlist for it would go before a item with no holds) and where no needs are pending by demand for each genre (e.,g fiction more popular than non-fiction so would be first). Materials could be destroyed or stored (non-retrievable), depending on the library's preference, though libraries opting for storage would be charged storage costs. Alternatively, the system could alternate the responsibility, notifying a library when it is responsible for digitization, to ensure equal participation. Because some libraries will always be more poorly funded than others, the success of an equal distribution is unlikely. Centralized subsidization or outsourcing is more realistic.

Member libraries' obligations to digitize and make available older materials would likely be limited to funding. The coordinating body would need to determine what titles have already been digitized, whether those could be made available to the NCL, and then create a plan for digitizing all remaining titles. The digitization could be done through scanning centers,



government efforts (as has happened in Norway), or by assignment to libraries holding the materials and willing to contribute this to the national effort.

### PROPOSED MODEL: FUNDING

Since this type of collaboration is intended to bring benefits not just to individual community members but also to the nation as a whole – in terms of preservation and access -- this proposal suggests a combination of several funding components: (1) a state contribution for public libraries within each state, (2) a federal contribution for all federal libraries, (3) a member fee for individual libraries who are not covered by federal or state contributions (or where federal or state governments choose not to participate), and (4) credits for libraries who digitize and contribute works. The goal here continues to be non-profit, in that all that is desired is the amount of funding necessary to ensure the continuity of information access and preservation. Permutations of funding would be welcome. For example, if the federal government funds the entire preservation and scanning effort, payments by states and private libraries would become unnecessary. Given Congress' reluctance to expand funding other digital information efforts (e.g., GPO, Library of Congress), though, this seems unlikely, and a more modest request of them would have a greater chance at success.

The amount set for each state would be based on the state's population, so that costs would be proportional to potential use. While use might vary widely depending on populations, determining costs based on use would result in incredibly unpredictable costs for every state and therefore was not considered. This is a form of funding inspired by Carnegie's initial investment in libraries, where he required cities to commit to continued funding and maintenance to gain a library from him (Schindler 2007). Though most libraries are funded by local governments, I have selected the state as the paying entity for simplicity's sake; billing every local government

would quickly become unmanageable, and state legislatures have already shown willingness to negotiate for all public libraries contained within their boundaries (e.g., Florida and database licenses). Users with accounts at any of the public libraries within the state could then access the resources shared through the NCL.

Federal fees would be determined by the user base of its libraries (e.g., federal court libraries). Though many of these are open to the public, the user base would be determined only by the library's primary users (i.e., those with user accounts with the libraries), and again, users would need to have accounts with the providing entity to access the resources of the NCL, as some form of verification would be required for access.

State or federal governments may choose not to join in this effort or to pay the necessary fees, which would then mean that those libraries would be unable to access the resources of the NCL. However, there are reasons to believe that both might be supportive of such a system. First, the NCL provides more permanence than the database licenses that some states already fund, and the preservation and recovery options that it makes possible could reduce state and federal costs elsewhere (see Benefits, Preservation below).

Private libraries will not be covered by these payments and would instead be charged a flat annual fee – flat more for convenience than anything else --- though users may be able to get around this should a private library decline to participate. Since individuals can be members of more than one library (e.g., university, local public, state), they could seek out another library that is already an NCL participant.

Libraries who are disproportionately contributing digitized items to the NCL could have their fees waived and could get credits, much as OCLC once provided for cataloging services, to subsidize their efforts.

## BENEFITS

The benefits to an NCL, if widely adopted, are many, including the ones that have already been mentioned in other sections. This section will focus on the ones that are likely to have the greatest impact.

### BENEFITS: BROADER AND MORE EQUAL ACCESS

Access to information will never be equal for many reasons, mostly relating to financial circumstances, but shared collections through digital means can bridge several known divides.

The three populations that will benefit most from this model are those in rural locations, people with disabilities, and those in poor neighborhoods.

Those living in remote areas may have access to public libraries, but often at inconvenient distances (Donnelly 2015), and aside from the cost and time associated with travel, may find their options limited based on the hours that their library opens. Digital collections can be accessed at any time, whether or not a library is staffed. Holidays, evenings, and early morning hours present no problems for access, and costs --- financial for gas or simply in the form of time --- associated with travel become irrelevant.

Some people with disabilities will also be better served. As discussed by the court in the HathiTrust decision, the limited number of titles available in accessible formats was a tremendous limitation for the print disabled (Authors Guild 2014), and the digital form provided by HathiTrust relieved some of the pent up need. Similarly, digital forms can be useful in other instances, in particular with those with physical disabilities, whether temporary or permanent. Transportation can be difficult (e.g. requiring special reservations) or time consuming for individuals in this category (Field 2007; Rosenbloom 2007). Or, they may be confined to a location for health reasons (e.g., quarantine, weakened immune system). While these individuals

may have access to printed library materials through special transportation options, mailing services, paid assistants, or friends, digital access can be instantaneous and may be accomplished without assistance for some users. Autonomy makes a meaningful difference in people's lives, especially at a time when other matters are outside of their control (Ells 2001).

Last, for those who live in areas where the libraries may not be well funded or whose needs greatly outstrip that of its library's resources, a collaborative library can serve as a sharing of wealth that does not "take" from the donors to provide to the users. In most cases, the objections to wealth redistribution is that an entity expands benefits for one population only through reducing them for another. The sharing of collections across the nation can be configured in such a way that only excess or rarely used materials by one community are contributed to the collaborative. Those 100 copies of a popular book that a library system acquired when the title first came out may no longer be needed by a library three years down the line; perhaps 90 could be donated to the NCL so that others in the nation, who cannot afford them and who do not have access to them through their libraries, can nonetheless have the opportunity to read them.

Of course, there are other alternatives for wealthier libraries which may very well be more cost effective. For example, simply licensing the digital e-books where available for an abbreviated length of time, may be less costly, not only in fiscal terms but also in personnel time and space utilization. This would result in no excess copies for the owning library and would correspondingly mean that the type of wealth sharing described would not occur. The ability of an NCL in reducing inequality will rest in part upon the commitment of its wealthier members to make choices that carry downstream effects.

**BENEFITS: MORE COMPREHENSIVE COLLECTIONS ACQUIRED AND PRESERVED**

There are several reasons why an NCL can collect and preserve a greater number of unique materials for the nation than libraries individually can accomplish today. Though collaborative collection development, we harness the resources of all libraries to a common goal, to ensure that we preserve for the nation the most comprehensive and meaningful set of materials that we can. NCL can use the principles in place at many county or system libraries where member libraries maximize their collections budget by identifying the number of titles needed by the overall community when purchasing, instead of making independent decisions resulting in more copies than necessary and less unique content than the system could otherwise acquire.

One approach for the NCL is discussed above, where all participating libraries do their collection development independently, with the coordinating entity filling in gaps as needed. Another approach would be more beneficial to the whole, though. If NCL participating libraries (or even a subset of them) agree that broad collection and preservation is a top priority, they can assure the broadest collection by assigning subject areas in advance (e.g., library A acquires all Georgia geography materials). Each library, by following through on its pre-determined subject purchases, will acquire unique materials and will digitize them for broad access through the NCL, ensuring a broad collection development plan.

This broad acquisition function would also serve another mission of libraries when coupled with digitization: preservation. In an era where climate change is becoming an increasing risk to property (III), having a broad collection plan and a digital collection increases the chances that the content of library collections, if not the objects themselves, survive even where fire, flood, or another disaster renders the object unusable.

By digitizing materials acquired for NCL, and backing up the materials in a manner consistent with data preservation best practices (DPC), an NCL could save both the federal

government (FEMA) and each local community the funds necessary to reacquire materials lost to a natural disaster. Backup copies, or rather, access to the materials, can be provided to the owning library with little delay and (assuming that the NCL has a digitized copy of any unique collections) no content loss. The artifacts themselves would still be lost, which may be of concern to historians, but they would have been lost in any case, so while the NCL proposal would not meet all needs, it would at least ameliorate damage.

As materials are digitized, the coordinating entity could establish combined scanning and storage locations – whether local, regional, or national --- to serve the more traditional meaning of preservation through dark storage of the physical materials digitized. These could be dispersed across several geographic areas for redundancy and kept at temperatures ideal for print materials, if not humans, but the creation of these would also increase the costs to participating libraries, as they involve not only real estate and equipment but staffing.

#### CHALLENGES: USE ISSUES

Perhaps the most commonly raised challenge for collaboration is use. When discussing collaboration with libraries – both those who have current collaborative relationships as well as those without – common themes arise. What advantages does it bring to their users, will this disadvantage their users, will it encourage free riding, and who gets to make decisions (e.g. loan periods) for the collaborative?

These are weighty issues, and every library will have its own interests, so the goal with this proposal is not to override libraries' individual decisions but to find mechanisms to fit them into an overarching system that aims to balance both local and national needs. For example, if some libraries wish to recall materials and others do not, the unified system would advertise that

materials may be recalled, but libraries can individually apply whichever rule they prefer to the materials they loan.

Perhaps the trickier issue is where a library wishes to control the actions of another library – e.g., not only do they object to recalling their own materials, they do not think other libraries should be allowed to do so – and herein is where we reach a point where the project stalls or we establish a framework that allows each library sufficient control over its materials so that it is more likely to share than not. As shown in the description above, this proposal suggests adopting the broadest language, which permits a wide range of differences within participating libraries.

The funding model has been designed to limit freeriding, though the risk of this is not possible to eliminate entirely. As noted earlier, users (including librarians) can be members of multiple libraries and if NCL is widely adopted, it would not be surprising if some non-NCL libraries still found ways to use NCL resources. Whether or not this becomes a problem largely depends on how much funding is obtained and the level of freeriding. Library policies, though, generally do recognize and permit some freeriding. Libraries will often lend materials through ILL even if the requesting library does not lend their materials (or has few materials to lend). For that reason, I do not anticipate that this would be a significant objection by NCL participants.

This section will not address end-user concerns about e-books both because there is sufficient coverage on these topics (Ennis 2018; Rosenwald 2015; Pew Research Center 2015), and because the underlying technologies are changing so rapidly that it is virtually impossible to provide a current analysis.

## CHALLENGES: LEGAL ISSUES

If libraries want to use their materials with whatever current technologies exist, even as those technologies change with growing regularity, digitization is necessary. With digitization comes the legal challenge of copyright as it does involve the reproduction right, which is normally the exclusive right of the copyright owner. United States copyright as it applies to CDL has been covered extensively in other literature (Bailey et al. 2017; Wu 2016a; Wu 2017), so this article will give only an abbreviated justification for digitizing and lending materials.

In the U.S. Constitution and the resultant copyright laws, the federal government has made clear its interests in copyright. The primary interest is the public's access to information, and to encourage the creation of such information, a secondary interest is born: the copyright owner's right to control the work for a specified time so that she can make a living from it (United States Constitution). The theory was that authors needed incentives to create works, and without those incentives, there would be no information for the public to consume. Later research has cast doubt on that proposition (Ku, Sun & Fan 2009) but the laws remain consistent with the original premise.

While the laws recognize the author's right to control his work, it also recognizes limitations on the exercise of that right. One of those limitations is library lending, a core library function to serve the public interest. Libraries have been given the authority to lend their materials, whether to their own users or to others; this is based on the principle of first sale (United States Code, Title 17, Section 107).

Copyright protects a work, not an individual embodiment of a work (with limited exceptions in works of art). Here is an illustration: the work of *Catcher in the Rye* is the content (i.e., the words) of the book; the individual embodiments include any copy printed or recorded (e.g., audiobook). Laws are structured so as not to damage the work or its market but provide no



protection for the manifestations. This enables people (even outside of libraries) to destroy, sell, lend physical books that they have legitimately acquired and is the core concept behind the principle of first sale.

Damaging the market usually means creating unauthorized copies that take the place of authorized copies that people would otherwise buy. For example, if you buy one copy of the most recent Harry Potter, make 100 copies and give them away to friends, then those 100 copies have replaced copies on the market, theoretically reducing sales by 100. Putting aside objections that those 100 individuals may never have been willing to purchase the book, one can see how the making of these copies could be viewed by Congress as damaging to the book's market.

However, not all damage to the market is protected by copyright laws. Libraries are authorized to lend books, and arguably, every loan replaces a copy on the market. Those borrowing library books likely are not buying them. But because libraries are lending the same item every time --- and not additional, unauthorized copies of the work -- this type of market harm is permitted.

When a library digitizes a work, CDL advocates argue that so long as the library (1) takes the print title out of circulation while the digital copy circulates, (2) lends on a one-to-one own-to-loan model (i.e., if it owns one copy, it can only lend one copy), and (3) controls the use of the digital item through DRM, it is essentially performing the same function as it had with the print book, but just using a different medium.

For that reason, scholars have argued that this type of use is justified through a combination of first sale and fair use. Though it is possible that sales could decrease for the actual work where CDL is used, this is also true of circulating the print copy. The argument here is that the digitized copy does not take the place of an authorized copy on the market but actually

takes the place of the library's legitimately-acquired copy. Therefore, any "harm" that is done is no more than what the library could have done with the print copy, though it would have been more expensive and labor intensive to do it in print (e.g., shipping costs). Considering how rapidly technology is changing today, it does not make sense for libraries to expend funds repeatedly on the same content, just in different form. CDL contemplates full and fair use of the content that a library has purchased even as technology changes.

As these two legal principles are derived from American law, the analysis above would not be the same in other nations. Very few countries have an equivalent legal principle to fair use, though Canada, Australia and India do have variations on the practice. For this reason, this paper contemplates only a national effort and not an international one. This also means that materials available through the NCL would be confined to users in the United States whether through geofencing or other means.

Even should the reasoning above ultimately prevail, it may not lessen the general risk of litigation itself. Content providers are understandably motivated to protect their content and their income streams, and filing suit has been very effective in deterring people from using technology in creative ways. Since the NCL has multiple partners, its design needs to recognize the litigation risk in both planning and system implementation to avoid jeopardizing the whole effort if a single bad actor abuses the system.

In the planning stages, the libraries that participate should be required to document their understanding that they are each responsible for their own compliance with copyright laws, that they have independently consulted their counsel or conducted a legal analysis, and that they are authorized to act on behalf of their organization. This protects the coordinating entity and other participating libraries should an authorized user unilaterally deviate from the principles of CDL.

The system itself must also have a robust authorization and accounting system. Establishing authorization hierarchies ensure that only those authorized can share materials with the NCL. And the system needs to record not only every change made to a record but also who made them, so that any abuse of the system can be traced back to its originator.

#### CHALLENGES: PUBLISHERS

Unfortunately, for-profit content provider patterns are fairly predictable by now and several examples have been described above. They have objected to most changes in technology --- photocopiers, digital recording devices, VCRs, streaming media --- where the technology has made it easier for users to capture or duplicate their content. Even where they adopt the new technology, they tend to apply it in a restrictive manner. We have already seen this with e-books, where publishers have used licenses to restrict use of ILL, printing or copying. We have also seen them use devices like limiting the number of uses of an e-book, even though libraries can prove through actual use that a book's lifespan is much longer than publishers suppose.

Given past actions, it is likely that vendors will fight against both CDL and a model of e-book sale/ownership. The manner of objection could take any number of forms, from threatening litigation to refusing to sell to libraries to stopping the print form entirely. Litigation is so effective because it is costly, even when the claim itself may have weaknesses. An entity being sued may be able to recover attorneys' costs if it prevails in a case, but that still means that it needs to come up with the funds at the front end, and many libraries may not have those resources.

Ultimately, though, no matter what objections they raise, I believe that publishers would adapt, just as they have in all the technological changes over the last century. In most cases, new avenues for revenues open up (e.g., video library opportunities after after litigation over Sony's

Betamax) and there are obvious choices with e-books: licensing directly to individuals or licensing tools and interfaces instead of content. They could also develop innovative programs like swap programs, where libraries turn in their print copies for e-book equivalents and the publisher charges only costs for the technology not the content. This would only be worthwhile to libraries if the swap results in ownership of the e-book, not a license, as the swap would not be equivalent otherwise.

Libraries have always been willing to pay for value-added features, so it makes sense to focus on aspects of technology where both the publisher and the library recognize the value. The reasons why e-book practices are so frustrating today is that the providers have taken away common rights that used to come with a book and libraries are paying repeatedly for the same content.

Another possible challenge is on the part of ILS vendors, as building one interface that is interoperable with all of them necessarily involves some sharing of code or at least cooperation in building dynamic feeds. Not all ILS vendors have been open to cooperating at this level, as demonstrated with libraries' struggles with discovery layers, so it may be that success to an NCL will require replacement of ILSes with those made by vendors more inclined to support collaboration than not.

#### CHALLENGES:TECHNOLOGY

Even though the seeds of the systems we need do exist today --- in consortial library systems like Alma, in FRBR-ization efforts by organizations like OCLC, in DRM readers like Overdrive, search and lending platforms like Open Library, in various reading list sharing apps (e.g., Wise) --- none of them are yet sophisticated enough to accomplish the integrated and dynamic workflow that is described here.

Consortial systems share information but are mostly responsive, not actively querying information entered and pushing related information to library staff. FRBR, while improving with each organization's implementation, is still hit or miss, resulting in strange connections of unrelated materials and occasional failures to identify the same title. Common DRM tools are too constraining, prohibiting any copying or printing, and making the research and citation process much more difficult than it needs to be. Individual database providers, like Westlaw, have a better approach, allowing limited copying and providing an embedded reference with the copied text for easy citation. DRM readers need to incorporate similar capabilities if they ever expect to be widely adopted for all purposes and not just leisure reading.

While ILS systems allow users to save preferences and lists of materials, they do not allow users to organize materials from different libraries into their own preferred order or to share those shelves with others. Apps and websites allow the sharing of lists but also lack the interactivity with multiple sources.

All of what has been described is theoretically possible but requires an entity with both the technical skills and the understanding of information usage, or strong relationships to both, to develop it. Further, the links to necessary pieces – other ILS systems, publishers' catalogs, and interfaces to new and used bookstores --- require more than technical expertise. It requires cooperation across industries and interests in a manner that is rare. Fortunately, there have been hints that such collaboration is possible in projects like LOCKSS, Google Books, and IA's work with university presses.

#### CHALLENGES: DIGITAL DIVIDE AND WEALTH GAP

Despite the rapid rate of technology advancement, there continues to be a significant gap between those who use technology and those who do not (Vick 2017; Karl 2017). As libraries

move online, they create more access opportunities but only if people have meaningful access to technology, in the form of connectivity as well as reading devices and assistance in using each. If they do not have access to broadband, computers, or e-readers, they may technically have the opportunity to access library content, but practically, they will be worse off than before.

Related to the digital divide is the wealth gap (World Inequality Lab 2018) and its impact on access. Even where technology is available, users may not have the funds to acquire it. Some libraries have already been tackling this through providing personal wifi access and partnering with schools for other devices but those efforts are limited by the libraries' finances and the technology infrastructure in any given region.

Education and assistance is also a necessary component for access, as seniors have adopted technology at a lower rate in part due to difficulties with understanding and using new technologies (Anderson & Perrin 2017).

In undertaking the building of an NCL, careful thought has to be given to access and ensuring that whatever approach is taken, it does not shrink access to already disadvantaged populations.

#### CHALLENGES: OTHER PRIORITIES

The largest challenge to this type of initiative is its low profile. The collection and preservation of information simply does not carry the same urgency for people as other issues. It is natural for those struggling with immediate needs or concerns to concentrate their own limited resources – whether time, money, or attention – on issues that directly respond to those worries. Social security, debt, job security, immigration, and discrimination occupy larger roles in societal consciousness and are more readily accepted as issues that require attention and funding.

Past instances involving information collection and preservation have shown the limited attention span the country has for these issues. In 2013, Harvard released a study showing how web documents cited by law journal articles and Supreme Court justices had disappeared (Zittrain, Albert, and Lessig 2014). There was a flurry of coverage by the press, but it rapidly disappeared. Similarly, over the last few years, Vint Cerf and Rick Whitt of Google have sounded the alarm about the loss of information recorded on various types of media like videotapes (Whitt 2017). Again, when they speak, they often receive press coverage but the interest is not sustained.

Politicians follow their constituencies and tend to be loudest on the issues on which voters are most vocal. Further, information collection and preservation is a more charged issue than it has been in the past, since this era has ushered in disagreement on what information is true and can be trusted. Though this debate should elevate the importance of what libraries do, it can serve to muddle the waters as well, if society believes that what is published is skewed and that is what libraries seek to preserve.

Without broad societal support for an effort that is expensive and time consuming, this type of library collection will be slow and heavily dependent on a few libraries that will choose to lead the effort.

## CONCLUSION

The individual elements raised in this paper are not new. Decentralized systems such as OCLC have existed for decades, and consortia, collection sharing agreements, and collaborative library efforts have been in place for much longer. CDL efforts, while newer, have now been in use for nearly a decade, and format shifting issues have been in and out of the courts for a little

longer than that. So while the individual concepts in this paper are not new, the combination of all of them together in an NCL to fill a noticeable gap in access and preservation is.

Communities without adequate access to information are at a disadvantage and any library alone is without the resources to represent every viewpoint. However, an NCL could harness the power of libraries together to provide all communities with access to more information sources than what they have available to them today. Even if that effort cannot represent every viewpoint, it almost certainly would represent more than what any library currently collects.



## REFERENCES

Anderson, Monica & Andrew Perrin. 2017. "Tech Adoption Climbs Among Older Adults." Pew Research Center, May 17. <http://www.pewinternet.org/2017/05/17/tech-adoption-climbs-among-older-adults/>

Authors Guild v. Google, Inc.. 2015. 804 F.2d 202.

Authors Guild, Inc. v. HathiTrust. 2014. 755 F.3d 87, 103.

Bailey, Lila, Kyle K. Courtney, David Hansen, Mary Minow, Jason Schultz, and Michelle M. Wu. 2017. "Controlled Digital Lending by Libraries." Accessed August 31, 2018. <https://controleddigitallending.org/>

Band, Jonathan. 2006. "The Google Library Project: Both Sides of the Story." *Plagiarism* (2006): 6-21. <https://quod.lib.umich.edu/cgi/p/pod/dod-idx/google-library-project-both-sides-of-the-story.pdf?c=plag;idno=5240451.0001.002;format=pdf>

Bosman, Julie. 2011. "Publisher Limits Shelf Life for Library E-Books." *New York Times*, March 14.

Bowker. "Print ISBN Counts, USA Pubdate 2002-2103." Accessed August 13, 2018. [http://media.bowker.com/documents/isbn\\_output\\_2002\\_2013.pdf](http://media.bowker.com/documents/isbn_output_2002_2013.pdf)

Coldewey, Devin. 2012. "Necessary Evil? Random House Triples Prices of Library E-Books." *TechCrunch*, March 2.

Crook, Jordan. 2011. "Penguin Shuts Down Libraries' Access to New E-Titles on Amazon's Kindle." *TechCrunch*, November 23.

Donnelly, F. 2015. "Regional Variations in Average Distance to Public Libraries in the United States" *Library & Information Science Research*, 37(4). DOI: <https://doi.org/10.1016/j.lisr.2015.11.008>

DPC (Digital Preservation Coalition). "Storage" in *Digital Preservation Handbook*. Accessed August 22, 2018. <https://www.dpconline.org/handbook/organisational-activities/storage>

Ells, Carolyn. 2001. "Lessons About Autonomy from the Experience of Disability." *Social Theory & Practice* 27(4)(October): 599-615.

Ennis, Matt. 2018. "College Students Prefer Print for Long-form Reading, e-Books for Research" *Library Journal*, March 27. <https://www.libraryjournal.com/?detailStory=college-students-prefer-print-long-form-reading-ebooks-research-lj-survey;>

Field, MJ and AM Jette, eds. 2007. *The Future of Disability in America*. Washington (DC): National Academies Press.

Google. "Google Books Ngram Viewer." Accessed July 26, 2018.  
<https://books.google.com/ngrams/info>

Google. "The Library Project – Books Help." Accessed July 26, 2018.  
[https://support.google.com/books/partner/faq/3396243?visit\\_id=1-636685549038642622-3207901398&hl=en&rd=1](https://support.google.com/books/partner/faq/3396243?visit_id=1-636685549038642622-3207901398&hl=en&rd=1)

HathiTrust. "Accessibility." Accessed August 21, 2018. <https://www.hathitrust.org/accessibility>

III (Insurance Information Institute). "Number of World Natural Catastrophes, 1980-2017." Accessed August 22, 2018. <https://www.iii.org/graph-archive/96424>

Kahle, Brewster. 2017. "Transforming Our Libraries from Analog to Digital: A 2020 Vision." Accessed on July 27, 2018. <http://er.educause.edu/articles/2017/3/transforming-our-libraries-from-analog-to-digital-a-2020-vision>

Katz, Ariel. 2016. "Copyright, Exhaustion, and the Role of Libraries in the Ecosystem of Knowledge." *I/S: A Journal of Law and Policy for the Information Society* 13(1):81-123.

Ku, Raymond Shih Ray, Jiayang Sun, and Yiyang Fan. 2009. "Does Copyright Promote Creativity? An Empirical Analysis of Copyright's Bounty." *Vanderbilt Law Review* 62(6): 1669-1746.

Marr, Bernard. 2018. "How Much Data Do We Create Each Day? The Mind-Blowing Stats Everyone Should Read." *Forbes*, May 21.  
<https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#74a03f7a60ba>.

Meyer, Robinson. 2015. "After 10 Years, Google Books is Legal." *The Atlantic*, October 20.  
<https://www.theatlantic.com/technology/archive/2015/10/fair-use-transformative-level-google-books/411058/>

Ota, Alan K. 1998. "Disney Locks in 20 More Years of Protection for Its Characters Under Copyright Extension Bill." *CQ Weekly*, 56 (41) (October 17): 2820.

Perrin, Andrew. 2017. "Digital Gap Between Rural and Nonrural America Persists." Pew Research Center, May 19. <http://www.pewresearch.org/fact-tank/2017/05/19/digital-gap-between-rural-and-nonrural-america-persists/>

Pew Research Center. 2015. "Report: Book Reading 2015."  
<http://www.pewinternet.org/2016/09/01/book-reading-2016/>

Rosenbloom, Sandra. 2007. "Transportation Patterns and Problems of People with Disabilities." In *The Future of Disability in America*, edited by MJ Field & AM Jette, 519-560. Washington DC: National Academies Press. <https://www.ncbi.nlm.nih.gov/books/NBK11420/>

- Rosenwald, Michael S. 2015. "Why Digital Natives Prefer Reading In Print. Yes, You Read That Right." *Washington Post*, February 22. [https://www.washingtonpost.com/local/why-digital-natives-prefer-reading-in-print-yes-you-read-that-right/2015/02/22/8596ca86-b871-11e4-9423-f3d0a1ec335c\\_story.html?utm\\_term=.0f67de1a4a19](https://www.washingtonpost.com/local/why-digital-natives-prefer-reading-in-print-yes-you-read-that-right/2015/02/22/8596ca86-b871-11e4-9423-f3d0a1ec335c_story.html?utm_term=.0f67de1a4a19)
- Schindler, Steven. 2007. "Case 5. Carnegie Public Libraries for America's Communities." In *Casebook for the Foundation: A Great American Secret*, edited by Joel L. Fleishman, J. Scott Kohler, and Steven Schindler. New York: PublicAffairs. [https://cspcs.sanford.duke.edu/sites/default/files/descriptive/carnegie\\_public\\_libraries.pdf](https://cspcs.sanford.duke.edu/sites/default/files/descriptive/carnegie_public_libraries.pdf)
- Singer, Natasha. 2011. "Playing Catch-Up in a Digital Library Race." *New York Times*, January 9.
- Stone, Brad. 2009. "Amazon Erases Orwell Books from Kindle." *New York Times*, July 17.
- United States Constitution. Article I. Section 8. Clause 8.
- University of Michigan and Google. "Cooperative Agreement." Accessed July 8, 2018. <https://www.lib.umich.edu/sites/default/files/services/mdp/um-google-cooperative-agreement.pdf>
- Vick, Karl. 2017. "The Digital Divide: A Quarter of the the Nation is Without Broadband." *Time*, March 30. <http://time.com/4718032/the-digital-divide/>
- Whitt, Rick. 2017. "'Through a Glass Darkly': Technical, Policy, and Financial Actions to Avert the Coming Digital Dark Age." *Santa Clara High Technology Law Journal* 33(2): 117-229.
- World Inequality Lab. 2018. "World Inequality Report 2018." <https://wir2018.wid.world/download.html>
- Wu, Michelle M. 2011. "Building a Collaborative Digital Collection, a Necessary Evolution in Libraries." *Law Library Journal* 103: 527-551.
- Wu, Michelle M. 2016a. "Collaborative Library Digital Collections Post-Cambridge University Press, HathiTrust, and Google Decisions on Fair Use." *Journal of Copyright Education & Librarianship* 1: 1-19.
- Wu, Michelle M. 2016b. "Food for Thought: Should Libraries Partner with Non-Library Search Engine Providers for their OPAC and Discovery Layers?" *Legal Information Review* 1: 51-74
- Wu, Michelle M. 2017. "Piece-by-Piece Review of Digitize-and-Lend Through the Lens of Copyright and Fair Use." *Legal Reference Services Quarterly* 36:51-73.
- Wyatt, Edward. 2005. "Google Library Database is Delayed." *New York Times*, August 13.

Young, Jeffrey R. 2008. "In Case Google Bails Out on Its Library Project, Universities Create a Backup." *Chronicle of Higher Education*, October 24.

Zittrain, Jonathan, Kendra Albert, and Lawrence Lessig. 2014. "Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations." *Harvard Law Review* 127: 176-199.