

Print Management at “Mega-scale”: A Regional Perspective on Print Book Collections in North America

Brian Lavoie
Research Scientist

Constance Malpas
Program Officer

JD Shipengrover
Senior Web & User Interface Designer

OCLC Research



A publication of OCLC Research

Print Management at “Mega-scale”: A Regional Perspective on Print Book Collections in North America

Brian Lavoie, Constance Malpas, and JD Shipengrover, for OCLC Research

© 2012 OCLC Online Computer Library Center, Inc.

Reuse of this document is permitted as long as it is consistent with the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 (USA) license (CC-BY-NC-SA):

<http://creativecommons.org/licenses/by-nc-sa/3.0/>.

July 2012

OCLC Research

Dublin, Ohio 43017 USA

www.oclc.org

ISBN: 1-55653-450-7 (13-digit) 978-1-55653-450-8

OCLC (WorldCat): 799083301

Please direct correspondence to:

Brian Lavoie

Research Scientist

lavoie@oclc.org

Suggested citation:

Lavoie, Brian, Constance Malpas and JD Shipengrover. 2012. *Print Management at “Mega-scale”: A Regional Perspective on Print Book Collections in North America*. Dublin, Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2012/2012-05.pdf>.

Contents

Acknowledgements	6
Introduction	7
Context	8
A Framework for Models of Print Consolidation.....	10
Mega-regions: A Framework for Consolidation	14
Some Definitions.....	18
The North American and Mega-regional Print Book Collections	19
Stylized Facts	24
Key Implications	46
Conclusions	56
References	59

Tables

Table 1. North American print book collection in WorldCat	19
Table 2. Holdings to publications ratio, by regional collection.....	22
Table 3. Regional coverage of the North American print book collection	23
Table 4. Regional overlap of top 250 most frequently occurring topical subject headings with North American print book collection	34
Table 5. Cumulative coverage of the North American print book collection	40
Table 6. HathiTrust coverage of regional print book collections.....	41

Figures

Figure 1. A framework for print collection consolidation	11
Figure 2. Mega-regions of North America	16
Figure 3. Two distinct publications of the same work by Stephen Foster.....	18
Figure 4. Sizes of the North American mega-regional print book collections. (Circles are scaled to reflect the number of print book publications in each regional collection.)	21
Figure 5. Print books as percent of total holdings, by mega-region	26
Figure 6. Share of regional print book holdings, by institution type.....	27
Figure 7. Share of ARLs in academic print book holdings, by region.....	28
Figure 8. “Rareness” at the intra-region and inter-region levels	31
Figure 9. Global diversity in regional collections.....	32
Figure 10: Uniqueness and global diversity as percentages of regional collections	36
Figure 11. Bi-lateral overlap with the BOS-WASH collection, by region.....	38
Figure 12. PHOENIX, DENVER, and SO-FLO overlap with other regional collections.....	39
Figure 13. Top five concentrations of print book holdings outside the mega-regions, US and Canada	45

Acknowledgements

We wish to thank Michelle Alexopoulos, Ivy Anderson, James Bunnelle, Lorcan Dempsey, David Lewis, Rick Lugg, Lars Meyer, Roger Schonfeld, Emily Stambaugh, and Thomas Teper for their thoughtful comments on a draft version of this report; their feedback was immensely helpful in improving the final version. We also thank Michelle Alexopoulos for her aid in obtaining the ZIP/postal code data used to construct the mega-regional collections analyzed in the report. We owe debts of gratitude to several OCLC colleagues: Bruce Washburn, for his assistance in producing the HathiTrust overlap findings; and Lorcan Dempsey, to whom the credit belongs for perceiving the mega-regions framework as a valuable context for exploring library data, and who encouraged us to find application for the framework in our work.

Introduction

The future of print book collections has received much attention, as libraries consider strategies to manage down print while transitioning to digital alternatives. The opportunity for collaboration is a recurring theme in these discussions. The OCLC Research report *Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment* (Malpas 2011) considers the prospects for shifting the locus of print book management models from local collections to regionally-consolidated shared collections, and concludes that while the necessary policy and technical infrastructures have yet to be developed, a “system-wide reorganization of collections and services that maximize the business value of print as a cooperative resource is both feasible and capable of producing great benefit to the academic library community” (p. 64).

As the *Cloud-sourcing* report acknowledges, much work remains to be done before a system of consolidated regional print collections becomes a reality. Nevertheless, it is interesting to speculate on an imagined future where such a system has materialized. A key question is the nature of the consolidated regional collections themselves—what would they look like? How similar or dissimilar would they be? Taken together, would the regional collections constitute a system of similar print book aggregations duplicated in different geographical regions, or would each collection represent a relatively unique component of the broader, system-wide print book corpus? These and other questions are relevant to a variety of broader issues, including mass digitization, resource sharing, and preservation.

The answers depend on how the collections are consolidated, or in other words, how the regions are defined. Several regional models for shared print book storage facilities are in evidence today. For example, the Five College Library Depository is shared by Amherst, Hampshire, Mount Holyoke, and Smith Colleges, and the University of Massachusetts Amherst. All of these institutions are clustered in the Connecticut River Valley in western Massachusetts. On a larger scale, the Northern and Southern Regional Library Facilities provide book storage capacity for the northern and southern campuses, respectively, of the University of California system. And on an even larger scale, the Western Regional Storage Trust (WEST) project proposes a distributed print repository service serving research libraries in the western United States.

Investigating the characteristics of a system of regionally-consolidated shared print book collections requires two elements: a model of regional consolidation, and data to support analysis of collections within that framework. This paper employs the mega-regions framework for the first and the WorldCat bibliographic database for the second. Mega-regions are geographical regions defined on the basis of economic integration and other forms of interdependence. The mega-regions framework has the benefit of basing consolidation on a substantive underpinning of shared traditions, mutual interests, and the needs of an overlapping constituency.

This report explores a counterfactual scenario where local US and Canadian print book collections are consolidated into regional shared collections based on the mega-regions framework. We begin by briefly reviewing the conclusions from the *Cloud-sourcing* report, and then present a simple framework that organizes the landscape of print book collection consolidation models and distinguishes the basic assumptions underpinning the *Cloud-sourcing* report and the present report. We then introduce the mega-regions framework, and use WorldCat data to construct twelve mega-regional consolidated print book collections. Analysis of the regional collections is synthesized into a set of stylized facts describing their salient characteristics, as well as key cross-regional relationships among the collections. The stylized facts motivate a number of key implications regarding access, management, preservation, and other topics considered in the context of a network of regionally consolidated print book collections.

Context

The analysis in this paper builds upon findings from the *Cloud-sourcing* report, which was motivated by a growing concern within the academic library community about the perceived decline in use (measured by circulation) of print collections, as well as the anticipated shift toward use of, not to say preference for, digital surrogates produced through mass-digitization programs. The report addressed these issues by investigating the overlap across print book collections in US academic libraries and the growing corpus of digitized books. Given that few (if any) library directors would withdraw a local print book collection in favor of digital surrogates without a guarantee of continued access to print originals, and in view of the cost-efficiencies of shared library storage, the report also measured the level of duplication between digitized books and physical inventory in existing shared repositories.

Several key findings emerged from this investigation. First, a significant share of the print book collections in Association of Research Libraries (ARL) institutions is duplicated in the HathiTrust Digital Library digitized book corpus; moreover, the rate of duplication showed

a steady growth over a twelve-month period. The median level of duplication¹ was about 19 percent in June 2009, and exceeded 30 percent a year later. Estimates projected the median overlap with HathiTrust to reach 36 percent by June 2011.² While this analysis does not take into account issues concerning the substitutability of digital surrogates for print originals, it does demonstrate that the content in HathiTrust substantially duplicates—by as much as a third or more—the print content managed at much greater expense in local ARL print collections.

Another finding was that the locally-held print content duplicated in the HathiTrust library is typically held by many libraries. In other words, much of this content is neither obviously “at risk” from a preservation point of view, nor in short supply from a fulfillment perspective. Consequently, the operational concerns associated with shifting print management and access operations to a trusted partner are relatively modest. Once an acceptable digital access and use platform emerges, many academic institutions will likely seek to externalize or “outsource” their traditional print repository functions to other providers. A risk inherent in a large-scale transformation of the system-wide print book collection is that a disorderly transition from local to group management may exacerbate disparities in access and even jeopardize the preservation of distinctive print resources. A prime motivation for the present study was a concern that a reconfiguration of print books held by a relatively small number of institutions could have a dramatic effect on the library system as a whole.

The *Cloud-sourcing* report found a high level of overlap (about 75 percent) between the holdings of HathiTrust and a sample of holdings from the aggregate inventory of several large-scale shared print storage repositories. However, the overlap between an individual ARL university library, the sample print storage inventory, and the HathiTrust collection was surprisingly low, suggesting that bi-lateral agreements between individual institutions and storage repositories were unlikely to generate the kind of space and cost savings that library directors (or university administrators) are likely to seek in an outsourcing arrangement. The report considered two potential solutions to this problem. First, a cooperative agreement among existing large-scale library storage facilities might prove to be more effective in terms of collective preservation and on-demand fulfillment. Alternatively, individual storage facilities might choose to adopt a collection development policy that would be optimized for a shared print service, by deliberately accessioning resources that would be of value to many institutions in the region.

1. Comparing discrete publications in HathiTrust against print book holdings in individual ARL libraries.

2. Subsequent analysis confirmed this projection. The slowed growth in overlap between 2010 and 2011 is partly explained by the evolving composition of the HathiTrust partnership and collection. The overlap will continue to fluctuate as a result of changing content contribution patterns (which affect the composition of the aggregated corpus), and changes in library acquisition trends (which alter the baseline against which overlap is calculated).

The solutions explored in the *Cloud-sourcing* report focus on print collections held in academic research libraries and assume physical consolidation of individual print collections into an above-the-institution aggregation. This paper takes an alternative approach, based on a broader view of library print collections—including those held in public libraries—and assumes that local print collections remain local, but are *virtually* consolidated at the regional level. The next section places this in the larger context of potential print consolidation models.

A Framework for Models of Print Consolidation

For the purposes of this report, *print consolidation* refers to any strategy undertaken by a group of institutions to achieve a mutual purpose by imposing some degree of integration across their local print collections. This definition is admittedly vague, because as will be seen, its two key components—“mutual purpose” and “degree of integration”—can be manifested in multiple ways. However, the definition is useful because it identifies the two fundamental dimensions along which any model of print consolidation can be characterized: *why and how* print collections are being consolidated.

Each dimension can be characterized in numerous ways, but to keep the discussion tractable, we will focus on two facets within each dimension. In terms of the first dimension (why print collections are consolidated), we identify two general goals or objectives. First, consolidation of print collections could be motivated by the desire to create a *shared back-up collection of print originals*, with end-users relying primarily or even exclusively on digitized surrogates for access.³ Alternatively, the consolidated collection could serve as a *shared resource for use*, with the aggregated print book holdings of multiple institutions leveraged over a wider base of potential users.

In terms of the second dimension (how print collections are consolidated), we consider two general strategies for achieving consolidation. First, local collections can be physically combined into a single shared collection and housed at a centralized repository (or limited network of shared repositories). Alternatively, consolidation can be achieved *virtually*, where local print collections remain in the custody of their respective institutions, but are

3. This strategy was examined at length for the journal literature in an analysis conducted by Ithaka S+R (Schonfeld 2011).

linked through a layer of services, such as a shared discovery environment and fulfillment system.⁴

Combining these two dimensions yields a simple framework (see figure 1) that serves the dual purpose of providing a high-level mapping of the print consolidation landscape, and orienting the analysis in this report within the spectrum of potential print consolidation models.

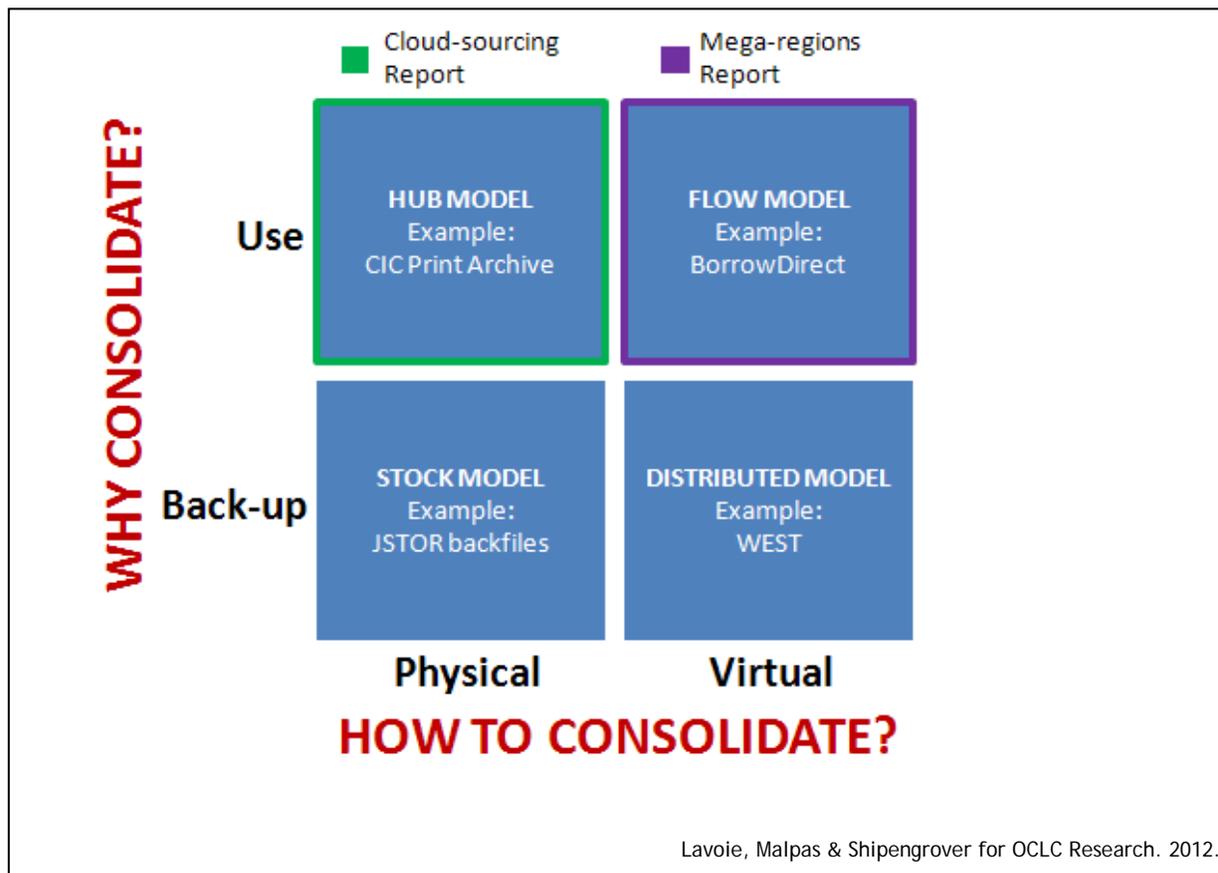


Figure 1. A framework for print collection consolidation

The framework identifies four basic models of print consolidation:

- *Hub model*: shared use of print materials is achieved through some form of physical consolidation of local collections.
- *Flow model*: shared use of print materials is achieved through some form of virtual integration across local collections.

4. The present study does not address the relative preservation benefits of physical or virtual consolidation of print collections (Maniatis et al. 2005). More recently, Paul Conway and colleagues have examined a variety of utility-based metrics for assessing the quality of digital surrogates as a replacement for print materials (Conway 2011).

- *Stock model*: shared back-up of print originals is achieved through a centralized consolidation of print materials into a shared repository.
- *Distributed model*: shared back-up of print originals is achieved through a virtual collection distributed across, and maintained within, local print collections.

Limiting the characterization of print consolidation models to these two dimensions omits other important aspects of consolidation. For example, these dimensions do not indicate whether local print collections are retained intact after consolidation, or some form of weeding/de-duplication is implemented across the participating institutions' combined holdings; nor do they address whether future collecting activity by participating institutions is subject to cross-institutional coordination. The purpose of the framework is to identify and distinguish a set of basic models; issues such as weeding or coordination are questions that can be asked in the context of any of the models.

The framework suggests starker choices than what prevail in reality, where print consolidation strategies can shade between the various categories. The categories within each dimension are not mutually exclusive: for example, a consolidated print collection could plausibly serve as both a shared back-up *and* a shared resource.⁵ Similarly, a consolidation strategy could involve some combination of a centralized repository of physically consolidated materials, supported by a network of locally managed collections.⁶ However, the services and infrastructure needed to support each model are different; additionally, certain attributes of the consolidated collections themselves may align more readily to one model or the other. Given these considerations, we will treat the four models in the framework as distinct options, acknowledging that this is a simplification but still a useful conceptual device for orienting the analysis to follow.

The *Cloud-sourcing* report focused on print consolidation models falling into the upper-left hand quadrant: i.e., the hub model, where the objective of shared use is achieved through physical consolidation. In this report, we are focusing on print consolidation models represented by the upper right hand quadrant: the flow model, characterized by shared use achieved through virtual consolidation. The reason for this is two-fold. First, a recurring theme throughout current discussions of cooperative print book collection management is that institutions continue to favor direct access to print book originals over a deliberate redirection of demand to digitized surrogates. The prevailing presumption is that print books

5. This is the model being explored by the Western Regional Storage Trust (WEST), which allows low- and moderate-risk titles in the archive to be shared under prevailing inter-lending rules.

6. For example, JSTOR has adopted a model of physical consolidation for its paper journal backfiles, utilizing the print repositories at California Digital Library and Harvard for this purpose. But a virtual model of consolidation is employed for JSTOR's rare or special collections, whereby the print originals are retained and managed by the organizations that own them (JSTOR 2012).

in library collections are intended to be accessed and used, rather than serve merely as back-ups. This is partly an accommodation to anticipated (and sometimes demonstrated) patron preference for print formats; it is also a pragmatic stance, given the overwhelming dominance of in-copyright titles in most library collections, as well as digitized book collections.⁷ Second, there is as yet no indication that institutions are willing to dispense entirely with their local print collections, although there is certainly strong interest in making management of print collections more efficient and less costly. Given these considerations—a focus on print books for use, and the likelihood that institutions will continue to manage print book collections locally for the foreseeable future—the flow model was chosen as the basis for the analysis in this report.

Given this, it is useful to say a little more about flow models. A flow model for the management of print collections focuses on virtually consolidating local collections into a shared resource for use, by linking them through a layer of shared services. In these circumstances, access is the primary service offering, with print materials “flowing” through the network of participating institutions to wherever needed. The chief benefit of a flow model approach to print management is the opportunity to leverage greater value from the legacy investment in print collections, by encouraging and facilitating greater use over a larger user base. This is achieved by combining a group of individual print collections into a larger and richer collective collection, which is then made available to users at all participating institutions. Attributes of the flow model are reflected in current resource sharing (ILL) networks, although such networks vary in the degree of integration across collections and access services. A well-functioning flow model helps optimize supply and demand in the collective collection by facilitating the movement of print materials from various points of supply (local collections) to the point of need (users anywhere within the network).

Distinctiveness is a desirable feature of local collections in the context of a flow model. A key benefit of a flow model approach is to expand the scope and depth of the print book offering to all users across participating institutions. If a significant portion of each participating institution’s print book collection is distinctive—that is, comprised of publications not widely available at other institutions—then combining print book holdings into a collective collection yields a print book resource that is, from the perspective of the user, far more extensive than

7. A 2009 study by Lavoie and Dempsey estimated that 14 percent of US-published print book titles in WorldCat were published prior to 1923 and therefore clearly in the public domain (2009). As of February 2012, OCLC Research analyses of the HathiTrust Digital Library collection indicate that about 1.15 million of the more than 5.16 million unique titles in the digitized collection—or about 22 percent—are in the public domain. These estimates exclude the large number of publications that might be classed as “orphan works,” for which some copyright exceptions can be exercised. By some accounts, orphan works may account for as much as 50 percent of the digitized volumes in the HathiTrust collection (Wilkin 2011).

what is on hand locally. In contrast, the more similar collections are, the smaller the “gains from trade,” in that access to the collective collection would offer little beyond what is available locally. Of course, substantial operational efficiencies and cost avoidance might still be achieved through some rationalization of duplicative holdings.

Since by definition flow models involve a virtual consolidation of print inventory, good data about local print book collections is essential. Consolidation occurs not at the level of the physical collections themselves, but instead within a layer of services that extends over all collections in the region and permits them to be managed and accessed as a cohesive whole. The service layer will be data-driven, and therefore its ability to present distributed print book holdings as a “regional collection” and offer functionalities operating on that collection—such as support for cooperative collection management decision-making, or region-wide discovery and fulfillment services—will depend on the accuracy and completeness of the underlying data.

The flow model is illustrated by the Borrow Direct partnership between Brown University, Columbia University, the Center for Research Libraries, Cornell University, Dartmouth College, Harvard University, MIT, University of Pennsylvania, Princeton University, and Yale University. Borrow Direct permits faculty and students at each of the partner institutions to easily discover, request, and receive delivery of print books and other materials located at any of the other institutions. Although there are some limitations on cross-institutional borrowing privileges (e.g., one physical volume per request, loan renewal not permitted), users of Borrow Direct benefit from the larger scope and depth of the partners’ collective collection, and the speed with which requested materials can be delivered to the user’s location (Nitecki 2009). Each Borrow Direct institution maintains its own print collection but a layer of services link them together into a virtual collective collection. Greater value is extracted from the collective print investment by making more materials available to more users.

Mega-regions: A Framework for Consolidation

Given a model of print consolidation, a choice must be made as to the level of aggregation underpinning the consolidation. In other words, how many (and which) institutions will be involved, and where are they located? For the analysis in this report, we chose to examine consolidation at the regional level. Regions tend to be bound together by ties that can both motivate and facilitate interaction between organizations within the region, such as geographical proximity, shared infrastructure, and economic interdependencies. These ties are well-suited to support a print consolidation model based on virtual consolidation and

flows of materials around the system. The logistics of supporting a flow model of print consolidation would likely be simpler and more efficient within a region, in comparison to a grouping of geographically dispersed and disconnected institutions. Moreover, regions seem to be a natural scale of aggregation for print consolidation. Regional clusters of cooperative activity seem to be where current print management initiatives are gravitating: many discussions regarding cooperative print management are organized at the regional level, sometimes involving established regional consortia. For example, a recent *Chronicle of Higher Education* article notes that the WEST project aims to build a “large-scale regional trust for print journal archives,” while “talks are under way about setting up similar regional repositories in the Northeast and Southeast” (Howard 2011).

“Region” is a nebulous term, and can be defined at a variety of scales. We operationalize the concept of a region by adopting the mega-regions framework described by Richard Florida, Tim Gulden, and Charlotta Mellander in the 2008 paper, *The Rise of the Mega-region* (see also Florida 2008). A mega-region is a geographical concentration of population and economic activity, generally subsuming multiple metropolitan areas and their surrounding hinterlands, and linked together through a complex connective tissue of economic interdependency, shared infrastructure, a common cultural history, and other mutual interests. Florida et al. observe that “[t]he mega-regions of today perform functions similar to those of the great cities of the past—massing together talent, productive capability, innovation and markets. But they do this on a far larger scale” (Florida, Gulden, and Mellander 2008, p. 460). In contrast to Thomas Friedman’s idea that the global economy is “flattening,” there are, the authors argue, “a strong set of counter-forces that lead to geographic clustering and the pushing together, so to speak, of economic activity. The mega-region ... is a consequence of this clustering force” (p. 460).

Florida and his colleagues used satellite imagery capturing night-time clusters of lights around the globe to identify twelve mega-regions in the US and Canada (see figure 2). “... [T]he mega-region,” the researchers note, “has emerged as the new ‘natural’ economic unit. The mega-region is not an artifact of artificial political boundaries, like the nation state or even its provinces, but the product of concentrations of centres of innovation, production, and consumer markets” (p. 461).

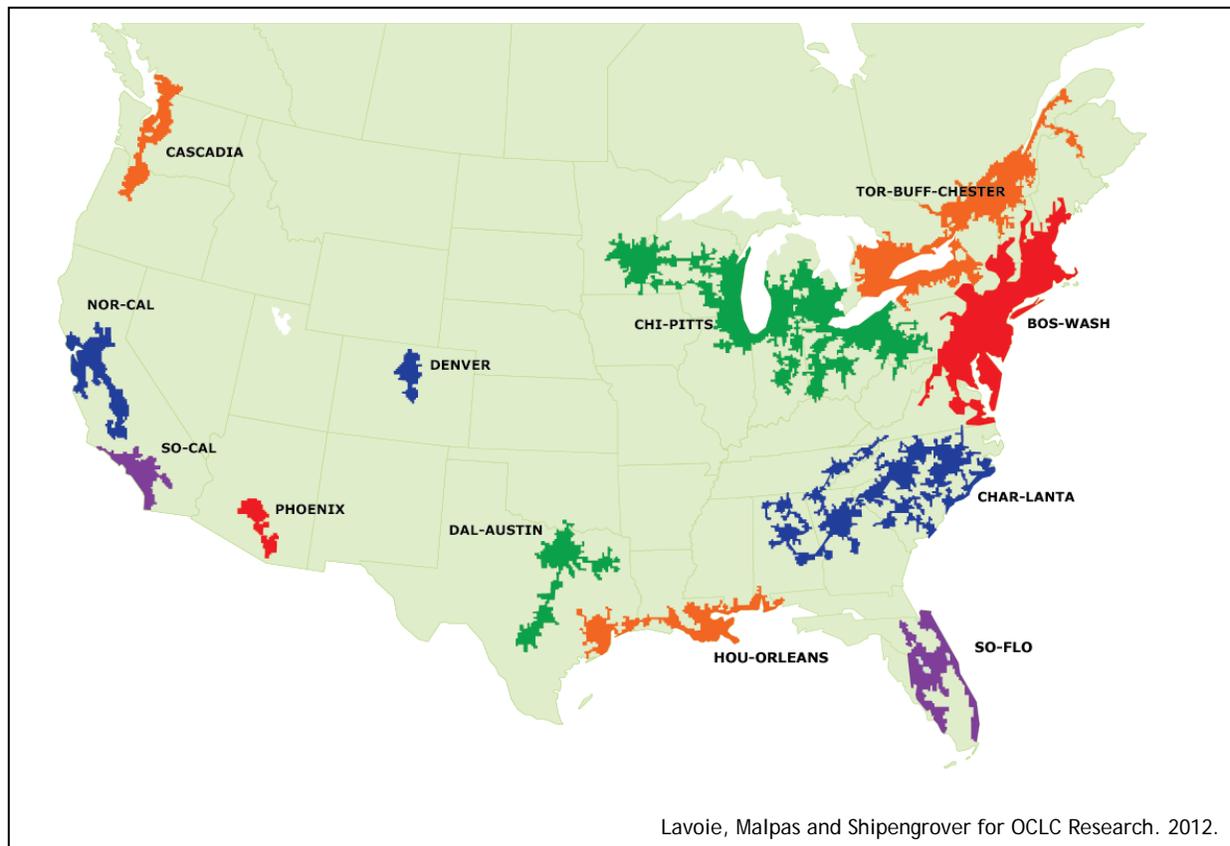


Figure 2. Mega-regions of North America⁸

As figure 2 illustrates, three of the twelve North American mega-regions extend over international boundaries: CASCADIA, CHI-PITTS, and TOR-BUFF-CHESTER. The extent of a mega-region is not limited by political boundaries, but rather by economic and cultural interdependency and mutual interests, which can occur in population centers that straddle an international border—Detroit and Windsor, for example.

Florida and his colleagues identify one mega-region in Mexico, centered around the Mexico City area. While Mexico is also part of North America, we exclude the Mexican mega-region from our analysis, and focus our attention on the remaining twelve US-Canadian mega-regions. The reason is that coverage of Mexican institutions in WorldCat is less extensive than for American and Canadian institutions, and therefore it is not clear that the Mexican presence in WorldCat would be sufficiently representative of the actual Mexican print book collection. For the remainder of the report, references to “North America” should be interpreted to mean the US and Canada only.

8. This visualization of the North American mega-regions, used here and in other graphics in this report, is based on figure 5 in Florida, et al. (2008, 470).

Mega-regions offer a compelling framework within which to think about a regional consolidation of print book collections organized as a flow model—that is, a virtual consolidation of local collections aimed at encouraging a flow of materials around the region. Mega-regions encompass existing networks—both physical and virtual—of integration and mutual interest that could potentially absorb and support a new network of cooperative print management and shared use. As we will show below, the vast majority of the overall North American print book collection is clustered within the twelve mega-regions. In this sense, mega-regions might be a “natural unit of analysis” for cooperative print management, as well as other cooperative library activities. Finally, mega-regions represent clusters of activity—research, innovation, learning, arts, and commerce—that library collections support. Therefore, it is useful to align clusters of library resources with clusters of activities that make use of these resources.

In a sense, the North American mega-regions illustrated in figure 2 are a snapshot, in that mega-regions are not static entities but instead grow and change over time. The boundaries of the twelve mega-regions in figure 2 will likely evolve in ways that absorb parts of the hinterlands surrounding the regions. Moreover, new mega-regions may form in areas where growing economic integration and other factors serve to bind people, institutions, and activities more closely than before. These dynamics will be at work not only in mega-regions, but almost any regional framework. From the standpoint of cooperative print management, the key implication is that regional boundaries will be in flux, likely resulting in the periodic appearance of new partners and an attendant need to adjust regional cooperative arrangements.

While the mega-regions framework is a useful and convenient tool for illustrating and analyzing regional consolidation of print collections, we are not necessarily advocating mega-regions as the appropriate scale for achieving consolidation and cooperative management in practice. Assuming that regions are in fact the natural unit of consolidation, the scale at which regions are defined will depend on a host of factors, including but not limited to the location of logistical networks, existing cooperative structures and agreements, and political jurisdictions (e.g., state or provincial boundaries). Mega-regions are one of many possible forms in which regional print consolidation can be manifested; careful analysis of the alternatives will help planners arrive at the most suitable choice for their circumstances.

Finally, as figure 2 makes clear, there is considerable space *between* the mega-regions. We do not imply that this space is “empty” or unimportant. In fact, the space between the regions—and more specifically, the aggregation of print books located there—has interesting characteristics in its own right, with important implications for cooperative print management and shared use. We discuss the areas outside the mega-regions in detail later in the report.

Some Definitions

The following terminology is used throughout this report:

- *Print book*: a book⁹ manifested in printed form. We exclude materials explicitly cataloged as theses, dissertations, or government documents from the analysis, as well as books in non-print formats such as e-books.

Publication: a distinct edition or imprint of a work. For example, *Walking Ollie, or, Winning the Love of a Difficult Dog* is a work—a distinct intellectual creation—by the author Stephen Foster. This work has appeared as several different publications, two of which are shown below (These would be counted as two distinct print book publications in our analysis).



Figure 3. Two distinct publications of the same work by Stephen Foster.

- *Holding*: an indicator that a particular institution (a library or some other organization) holds at least one copy of a particular publication in its collection. Note that a holding says nothing about the number of physical copies owned by the institution, other than at least one copy is available. For example, according to their catalog, the Dallas Public Library owns three copies of the Perigee Books publication

9. More specifically, we equate a “book” with a language-based monograph.

of *Walking Ollie*. All three copies would be represented in WorldCat by a single holding associated with the Dallas Public Library.¹⁰

- *Collective collection*: the combined holdings of a group of institutions, with duplicate holdings (i.e., those pertaining to the same publication) removed. This yields the collection of distinct publications that are held across the collections of the institutions in the group.

The North American and Mega-regional Print Book Collections

The WorldCat bibliographic database is the closest approximation available of the global collective collection—that is, the combined holdings of libraries and other institutions worldwide. While WorldCat data has certain limitations regarding coverage and interpretation of holdings information, it is nevertheless the best data source available for analysis of aggregate information resources such as regional print book collections. In January 2011, WorldCat contained 214.6 million bibliographic records representing information resources of all descriptions; these information resources accounted for nearly 1.7 billion holdings distributed across institutions all over the world.¹¹

Table 1 deconstructs WorldCat into the North American print book collection.

Table 1. North American print book collection in WorldCat (January 2011)

Collection	Publications (millions)	Holdings (millions)
WorldCat	214.6	1,679.1
Print books	128.1	1,238.1
Print books in North America	45.7	889.5
Print books—US	40.9	840.0
Print books—Canada	14.2	49.4

10. Readers familiar with the FRBR entity relationship model will recognize that a publication is equivalent to a FRBR manifestation, and a physical copy to a FRBR item.

11. Quarterly snapshots of WorldCat are maintained and programmatically enriched by OCLC Research to support a range of projects and prototypes. External researchers interested in making use of this data in their own studies are encouraged to contact OCLC Research, which can make provisions for access.

An important caveat to note in regard to table 1, as well as other results presented in this report, is that they reflect institutional collections as they are cataloged and represented in WorldCat. The accuracy of holdings data in WorldCat may be lessened by the presence of duplicate records, cataloging errors, incomplete registration of collections, and other sources of inconsistency.

Of the 128.1 million distinct print book publications represented in WorldCat, 45.7 million are held by at least one institution located in either the US or Canada. This constitutes the *North American print book collection*, or the collective collection of print book publications held by North American institutions. Coverage of the North American collection varies considerably between the US and Canada: US institutions alone can muster 90 percent of the publications in the North American collection, while Canadian coverage is 31 percent. Similarly, 94 percent of the holdings comprising the North American print book collection are associated with US institutions, while the remaining 6 percent are of Canadian origin.

Richard Florida and his colleagues generously provided lists of the US ZIP codes and Canadian postal codes associated with each of the twelve mega-regions defined in their 2008 paper.¹² These ZIP and postal codes were then compared to location information associated with each of the nearly 1.7 billion holdings in WorldCat. In this way, all WorldCat holdings associated with each of the twelve North American mega-regions were identified, along with all holdings located in either the US or Canada that fell outside the mega-regions. Once the holdings for a particular mega-region were identified, the subset corresponding to print book publications were extracted, and this in turn established the regional collective collection of print books. The sizes of the twelve mega-regional print book collections, measured in terms of publications and holdings, are shown in figure 4.

12. The authors thank Michelle Alexopoulos of the University of Toronto for arranging the provision of the mega-region ZIP/postal code data for our work.

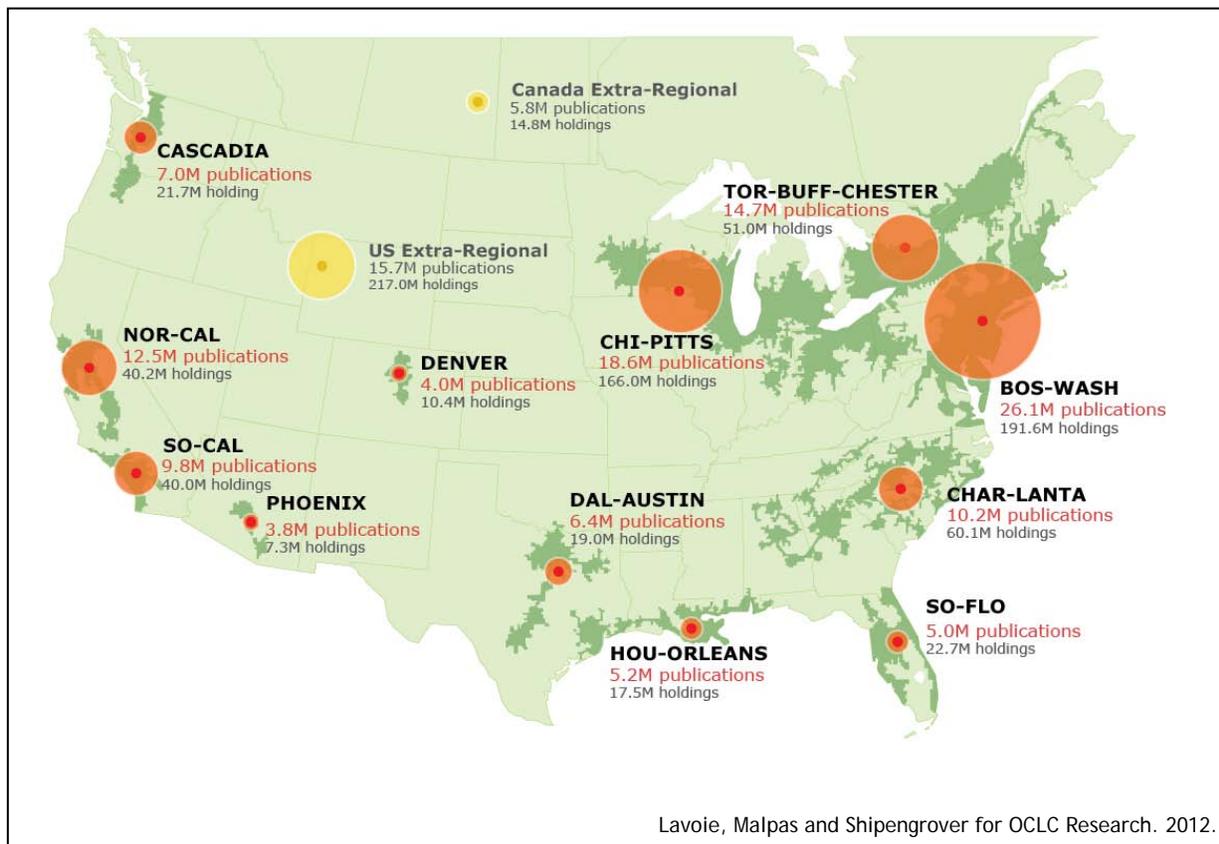


Figure 4. Sizes of the North American mega-regional print book collections. (Circles are scaled to reflect the number of print book publications in each regional collection.)

BOS-WASH is the largest regional print book collection, in terms of both distinct publications and total holdings. PHOENIX is the smallest, with only 15 percent as many publications, and 4 percent as many holdings, as BOS-WASH. The median regional collection size is 8.4 million distinct publications, and 31.3 million total holdings.

The ratio of holdings to publications provides a metric illustrating the degree to which a region’s collection of distinct print book publications is “amplified” into total print book holdings around the region. Higher ratios suggest higher levels of duplication—or from an access perspective, greater levels of availability—within a region, while lower ratios suggest the opposite. Table 2 reports the holdings to publications ratio for each of the twelve regional collections.

Table 2. Holdings to publications ratio, by regional collection

Region	Holdings (millions)	Publications (millions)	Holdings/ Publication
BOS-WASH	191.6	26.1	7.34
CASCADIA	21.7	7.0	3.11
CHAR-LANTA	60.1	10.2	5.92
CHI-PITTS	146.0	18.6	8.94
DAL-AUSTIN	19.0	6.4	2.98
DENVER	10.4	4.0	2.58
HOU-ORLEANS	17.5	5.2	3.39
NOR-CAL	40.2	12.5	3.22
PHOENIX	7.3	3.8	1.91
SO-CAL	40.0	9.8	4.09
SO-FLO	22.2	5.0	4.53
TOR-BUFF-CHESTER	51.0	14.7	3.47

The holdings to publications ratio varied widely across the regions, with CHI-PITTS exhibiting the highest value (8.94), and PHOENIX the smallest (1.91). Five regions exhibit a ratio of 4.0 or higher: that is, an average of four holdings across the region per print book publication. With the exception of PHOENIX, the remaining regions all exhibit holdings to publications ratios of 2.5 or higher. These results suggest that duplication (or availability) of print book publications is, on average, relatively low within the regions: even the highest ratio, associated with CHI-PITTS, suggests that on average only about nine institutions hold a given print book publication in their collections, despite the geographical extent of the region and the many institutions it contains. We re-visit this topic in more detail in the next section.

Table 3 reports coverage of the overall North American print book collection for each of the twelve regional collections.

Table 3. Regional coverage of the North American print book collection

Region	Coverage (%)
BOS-WASH	57
CASCADIA	15
CHAR-LANTA	22
CHI-PITTS	41
DAL-AUSTIN	14
DENVER	9
HOU-ORLEANS	11
NOR-CAL	27
PHOENIX	8
SO-CAL	21
SO-FLO	11
TOR-BUFF-CHESTER	32

The BOS-WASH region alone can account for nearly 60 percent of the entire North American print book collection. Other large regions, such as CHI-PITTS and TOR-BUFF-CHESTER, also exhibit significant coverage. Most regions, however, can only account for less than a quarter of the North American collection; for each of these regions, the vast majority of the print book publications available in North America are to be found elsewhere outside the region.

Before turning to a more detailed description of the twelve regional print book collections, it is useful to say a word about the areas *between* the regions. This report focuses on the regional collections, but this is not to diminish the importance of the print book holdings located outside the mega-regions. Indeed, these “extra-regional” print book holdings are significant in scale, accounting for more than 217 million holdings on 15.7 million print book publications in the US, and 14.8 million holdings on 5.8 million publications in Canada. Some of the local print book collections scattered through the extra-regional space are quite distant from even the closest mega-region; others are perched right on a mega-region’s boundary, or in its nearby hinterland. Clearly, US and Canadian print book holdings located outside the mega-regions constitute an important resource, but consolidating them into collective collections, like the regional collections, can be problematic. Unlike the mega-regions, there is no obvious collaborative structure or patterns of mutual interest binding these collections together. We will say more about the US and Canadian extra-regional collections in the next section.

Stylized Facts

Mega-regions provide a framework for organizing local print book collections into regional collections. But what would these regional collections look like? To answer this question, a detailed analysis of each of the twelve mega-region print book collections was undertaken using WorldCat bibliographic and holdings data. The result was a wealth of statistics characterizing the regional collections from numerous perspectives. Rather than attempting to present all of these statistics to the reader, we instead chose to synthesize the analysis into a set of *stylized facts*—in other words, a set of broad observations based on empirical findings. Taken together, the stylized facts constitute a general description of the North American mega-region print book collections, from which a number of implications regarding access, management, and preservation can be derived. We discuss several of these implications at the end of the report.

Library operations—and reputation—are still bound up with books

The OCLC (2011) report *Perceptions of Libraries, 2010: Context and Community*, reminds us that print books continue to be synonymous with libraries and library use, noting that “[t]he library brand is ‘books’ ... In 2005, most Americans (69%) said ‘books’ is the first thing that comes to mind when thinking about the library. In 2010, even more, 75%, believe that the library brand is books” (p. 38). The same report found that borrowing print books is still the top activity among library users (p. 35). Despite the attention (and funding) lavished on electronic and digital content in recent years, libraries of all types continue to devote significant resources to the management of print book collections.

While acceptance of e-books is increasing in academic and public libraries, the still-limited range of content, competing and incompatible platforms, and restrictive licensing regimes remain impediments to wide-scale adoption.¹³ This has important consequences for the organization of library service provision, as well as operating expenses. As shown in a 2010 study by Paul Courant and Buzzy Nielson, the long-term costs of storing print books are significant (estimated at \$4.26 per volume per year in open stacks) and relatively inelastic.¹⁴ In contrast to the journal literature, much of which has migrated into electronic formats and

13. In 2008, Mark Nelson predicted that many of the impediments to e-book adoption in academic libraries would be resolved within 5 years. In 2010, a survey of public library leaders found a high level of interest in e-book adoption but also pervasive concerns about restrictive licensing and platform interoperability (COSLA 2010). A recent report by the Pew Internet and American Life project finds that “the increasing availability of e-content is prompting some to read more than in the past and to prefer buying books to borrowing them” (Rainie et al., 2012). In the global consumer market, e-book adoption rates are already high and predicted to increase substantially (Bowker 2012).

14. Courant and Nielson’s study examines print book storage costs under a variety of different circumstances, and concludes that space is the single greatest cost driver. The sheer physicality of print books limits options for cost-effective management (2010).

aggregations managed by third-party agents, print books continue to occupy a significant share of local library space.

The long legacy of library investments in print books is reflected in the WorldCat database, where 60 percent of the bibliographic records describe print books and 75 percent of holdings are linked to print book titles. The outsized presence of print books in WorldCat records and holdings stems in part from cataloging practice. For example, title-level holdings for serials effectively mask the volume count of institutional journal holdings, which may significantly outnumber books on a per-volume basis. Likewise, format integration (single-record cataloging of titles produced in multiple formats) means that burgeoning e-book collections are not adequately accounted for in holdings counts, since electronic holdings may be intermingled with print holdings. Yet the millions of books acquired by North American libraries over many years of operation, the shared bibliographic infrastructure created to manage them as a collective resource, and the still powerful association between the codex and the library “brand” (or stereotype) serve to highlight the importance of print books to libraries and their users.

The impact of centuries of library investment in print books can be seen at the regional level. As figure 4 illustrates, print books account for anywhere from two-thirds to three-quarters of total holdings in each of the twelve mega-regions. The same characteristic is seen across different library types. Print books account for 68 percent of ARL library collections, while non-ARL academic libraries in North America are slightly higher at 69 percent. Eighty percent of North American public library collections are print books, while North American school (K-12) library collections are even higher at 87 percent. Again, while these results must be considered in light of cataloging practice and patterns of use of WorldCat as a bibliographic utility, they are nevertheless broadly indicative, and not only illustrate the ongoing predominance of print books in library collections, but also the importance and scale of the print collection management problem. Libraries retain responsibility for managing massive amounts of print book inventory, while at the same time they are transitioning their focus—and substantial portions of their budgets—to electronic and digital collections. Moreover, libraries face economic pressures to cut costs and justify value. A new system of print book collection management is needed to accommodate these conditions.

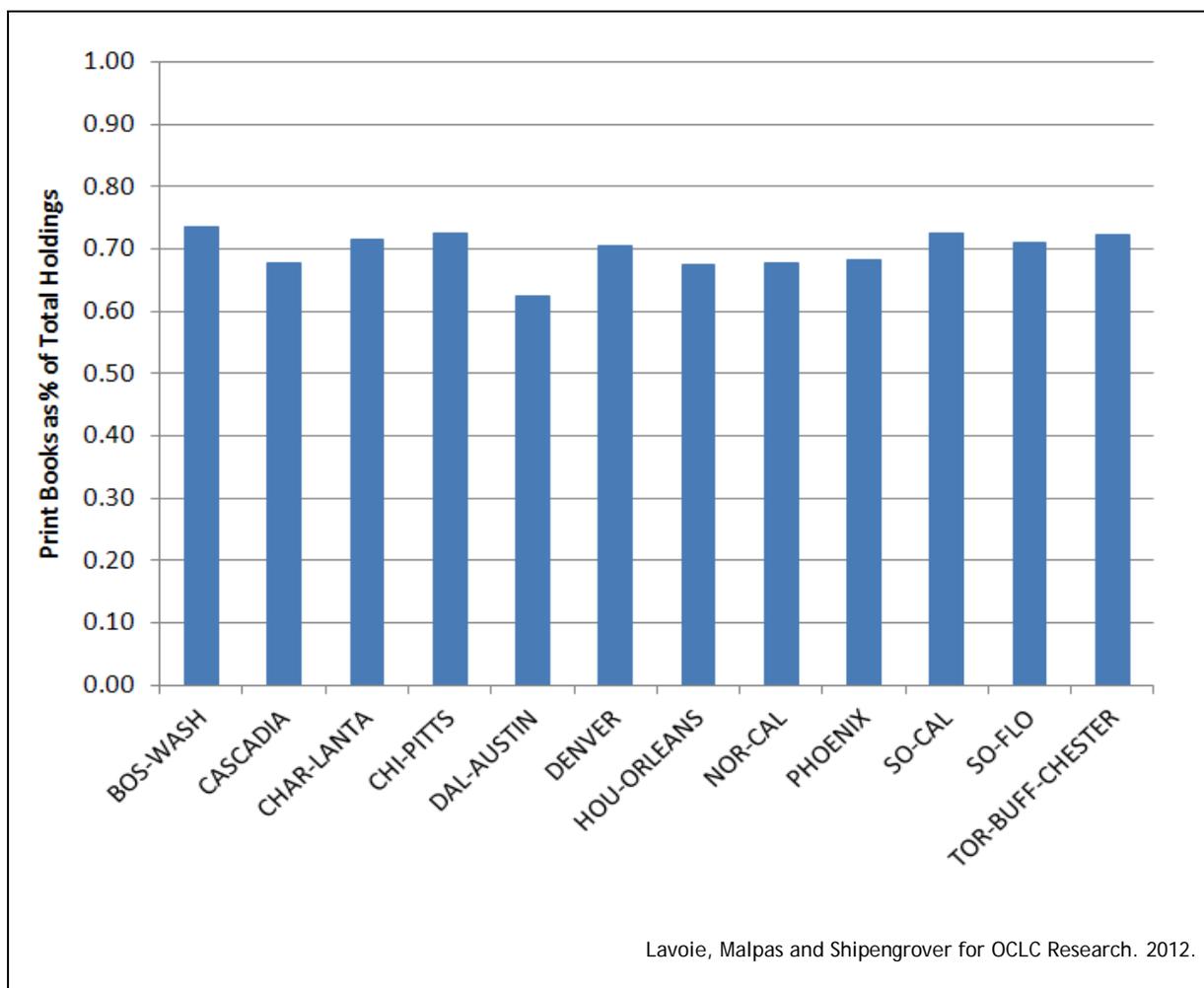


Figure 5. Print books as percent of total holdings, by mega-region

Academic institutions are the custodians of the majority of system-wide print book inventory

The success of a regionally-based cooperative model of print collection management depends on engaging institutions that control significant portions of the region-wide print book inventory. As the results in figure 6 show, the majority of the print book inventory in every region is in the custody of academic institutions.

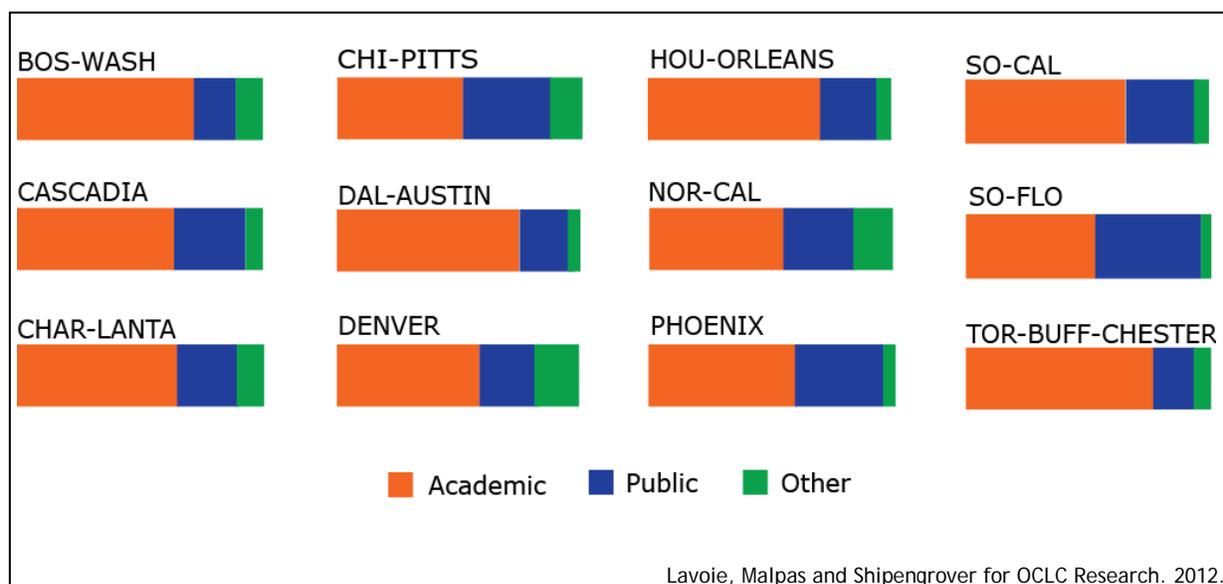


Figure 6. Share of regional print book holdings, by institution type

The extent to which academic institutions dominate print book holdings varies across regions, with the highest proportion in the TOR-BUFF-CHESTER region (76 percent), and the lowest in CHI-PITTS (51 percent). But the key point is that in every region, more than half of the regional print book inventory is in the hands of academic institutions—and in some regions, considerably more than half. We are aware that many public library holdings are not represented in WorldCat, and this will tend to amplify the relative presence of academic institutions in regional print book collections. But even taking this coverage gap into account would not serve, in our judgment, to overturn the conclusion that most print book inventory in the regional collections belongs to academic institutions, given the wide gap between the relative shares of each institution type exhibited in figure 6.

Print book holdings associated with academic institutions can be divided into those belonging to ARL institutions (the most research-intensive academic institutions), and those belonging to non-ARL academic institutions. BOS-WASH has the greatest number of print book holdings belonging to ARLs, at 65.3 million—more than twice the number of the region with the second-highest total, CHI-PITTS. However, it is in fact PHOENIX—the smallest regional collection—that has the highest percentage of its print book holdings associated with ARLs (52 percent); TOR-BUFF-CHESTER is next at 46 percent. In contrast, SO-FLO (12 percent), and CHI-PITTS and DAL-AUSTIN (both at 19 percent), are the regions with the smallest percentage of ARL holdings. Another way to assess the presence of ARLs in the regional collections is to compute the share of academic holdings in each region belonging to ARLs; figure 7 reports these results.

Considerable cross-region variation is apparent: in PHOENIX, nearly 90 percent of all academic print book holdings belong to ARLs, compared to less than a quarter in SO-FLO.

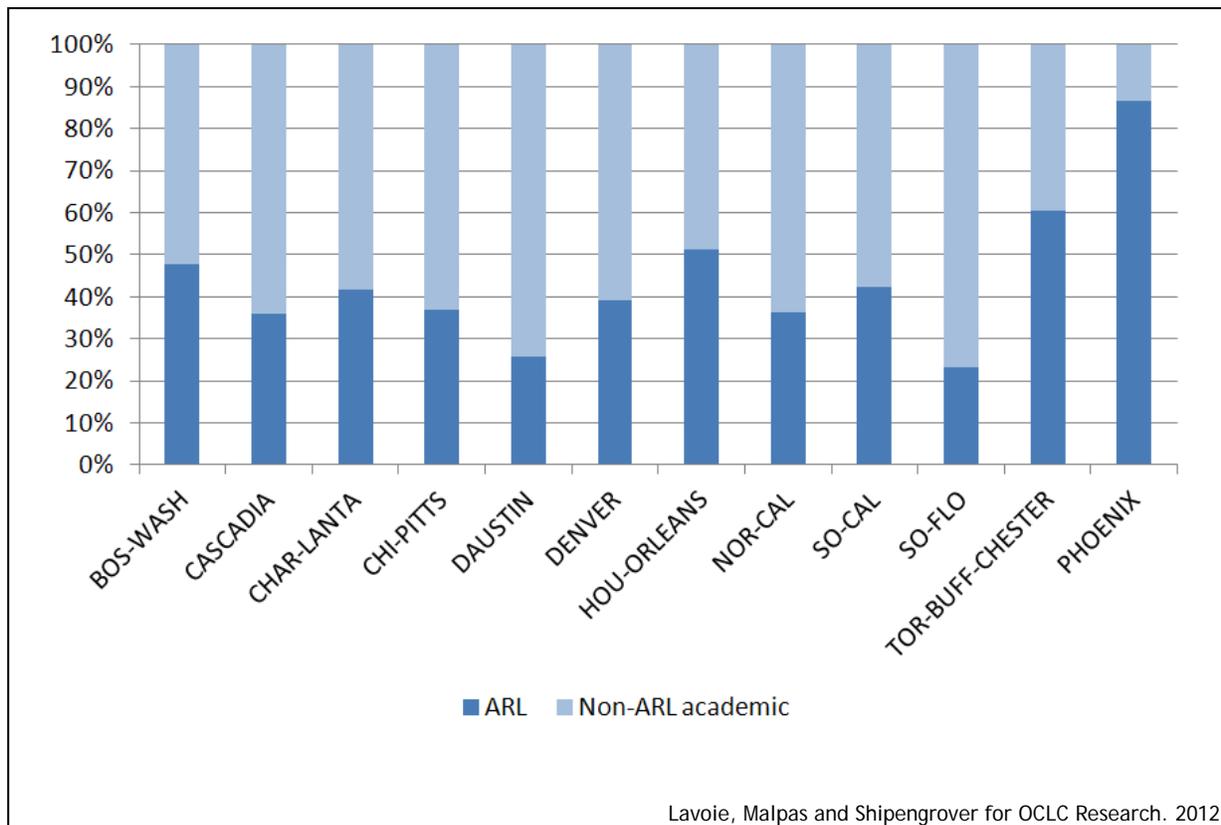


Figure 7. Share of ARLs in academic print book holdings, by region

The fact that most regional print book inventory is managed by academic institutions suggests that regional print book collections are, on average, geared toward the needs of faculty and students in higher education. This is further evidenced by the relatively low percentage of print book holdings belonging to public libraries in most regions (see figure 6): in half the regions, the share of public libraries is below a quarter, and in several regions the share is particularly low (BOS-WASH and TOR-BUFF-CHESTER, both at 17 percent). However, a few regions do exhibit relatively high percentages of public library print book holdings: SO-FLO (43 percent); CHI-PITTS (36 percent); and PHOENIX (35 percent). These regional collections would seem to be better positioned, vis-à-vis other regions, to serve the needs of general users.

Rareness is common within and across regional collections

WorldCat holdings data suggests that a significant share of print book inventory is relatively scarce both within regions and across regions. At least three quarters of the print book publications in each regional collection can be found at five or fewer institutions in the

region. Recall that a print book publication is a distinct imprint or edition of a printed book. Therefore, other publications pertaining to the same work may be available at other institutions. For example, while a particular publication of *A Tale of Two Cities* may be rare in the sense that it is held by only three institutions in the BOS-WASH region, many other publications of the same work may be available at other institutions in the region. Moreover, a print book holding indicates that an institution holds at least one copy of the publication in question; it may be that the institution holds many copies, which would alleviate to some degree the apparent scarcity observed at the publication level.

In some regions, the percentage of print book publications held by five or fewer institutions is particularly high: in DENVER, it reaches 89 percent, while in PHOENIX it is 95 percent. A partial explanation for exceptionally high percentages of “rare”¹⁵ publications (that is, held by 5 or fewer institutions) might be found in a correspondingly high fraction of print book holdings within the region associated with ARL institutions. Intuition would suggest that the largest research libraries are likely to possess relatively unique print book collections vis-à-vis other institutions. In fact, the percentage of rare publications in a region and the share of print book holdings belonging to ARL institutions do exhibit a moderate degree of positive correlation,¹⁶ indicating that regions with a relatively heavy ARL presence tend to have higher shares of rare materials.

The apparent “lack of abundance” of many print book publications within regional collections suggests both opportunity and challenges. Low levels of duplication correspond to high levels of uniqueness within a regional collection, which in turn suggests that a regionally-consolidated collection would represent a significantly richer information resource, in terms of scope and depth, than what is available at any single institution. However, the ability to capitalize on this uniqueness—and confer benefits on regional users—will depend on the geographic size of the region and the robustness of its inter-lending infrastructure. Potential benefits will also be scaled to the extent that aggregate regional demand for a particular print book publication exceeds local demand at the institution or institutions where the publication is held.

Rareness is also common *across* regional collections. Forty-nine percent of the publications in the North American print book collection are only available in one regional collection, or are only available in either the US or Canadian “extra-regional” collection.¹⁷ Eighty percent of

15. It should be noted that a print book publication’s “rareness”—i.e., the fact that it is held by only a few institutions—does not necessarily imply that it is an exceptionally valuable contribution to the regional or system-wide print book resource. For example, its scarcity may owe to the obsolescence or low quality of its content.

16. The Pearson correlation coefficient is 0.46 for the twelve regions.

17. The US and Canadian extra-regional collections are the collective print book collections of all institutions located outside of the mega-regions in the US and Canada, respectively. We will say more about these collections later in the report.

the publications are only available in five or fewer regions.¹⁸ Significant portions of several regional collections are unique to their regions: a third of the BOS-WASH collection, and a quarter of the TOR-BUFF-CHESTER collection, can be found in no other region. The majority of the regionally-unique materials are concentrated in regions located in the eastern half of the United States and Canada; more specifically, about 70 percent of the regionally-unique materials are located east of the Mississippi River.

Scarcity or uniqueness within a region does not seem to be a predictor of scarcity or uniqueness across regions. As figure 8 shows, a strong relationship between these characteristics is not apparent. In fact, if any relationship exists at all, it appears to be a negative one: regions with higher levels of intra-regional uniqueness tend to have relatively fewer materials unique to the region. This counter-intuitive relationship seems to be driven by regional size. Regions located to the upper left on the chart tend to be smaller: PHOENIX, DENVER, DAL-AUSTIN, CASCADIA; regions located toward the lower right tend to be larger: BOS-WASH, TOR-BUFF-CHESTER, CHI-PITTS. A possible explanation for the pattern in figure 8 is that smaller regions tend to have fewer institutions, which may act to reduce rates of duplication within the region. On the other hand, fewer institutions also means fewer materials in the regional collection, and therefore fewer opportunities to include rare or unique publications not available in other regions.

18. The US and Canadian extra-regional collections are counted as "regions" in this result. For example, if a particular publication was available in BOS-WASH, CASCADIA, and several US locations outside the mega-regions, this would be counted as three "regions."

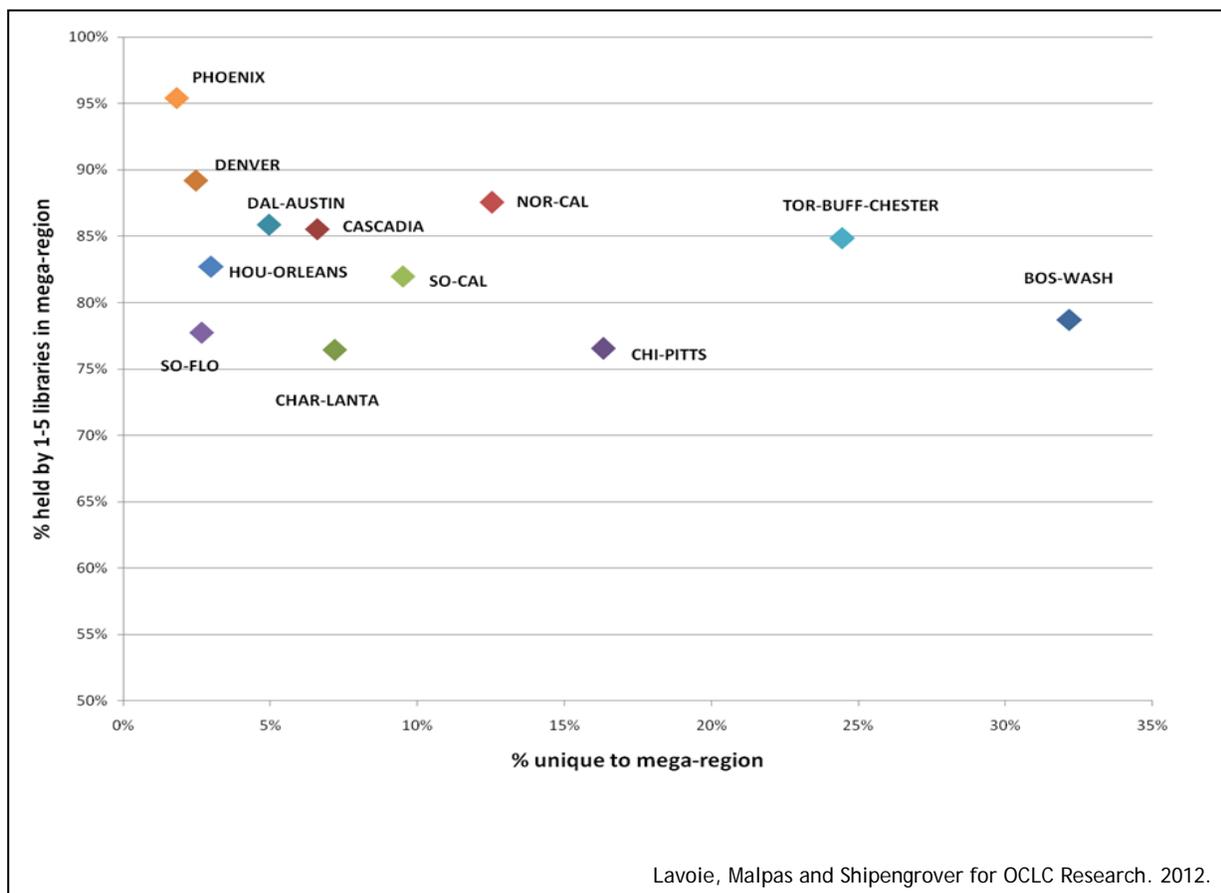


Figure 8. “Rareness” at the intra-region and inter-region levels

Analysis of overlap within and across regions indicates that considerable distinctiveness attaches to the regional collections at several levels. Consolidation at the regional level yields an aggregate print book resource that is richer in scope and depth than any single local collection. But distinctiveness also manifests at the inter-regional level, where a significant portion of the overall North American print book collection is available in only a few or even a single region. It is worth noting that no regional collection is completely subsumed within another regional collection, or can be entirely duplicated through the combined holdings of a group of regions. All regional collections have a store of print book publications that are unique to that region. Even the smallest regional collection—PHOENIX—contains a fraction of materials (2 percent, or nearly 70,000 distinct print book publications) that are only available in that region.

Regional collections are globally diverse and exhibit similar collecting patterns across broad subject areas

Global diversity is a characteristic common across all regional collections, measured by the presence of non-English language materials and books published outside North America. Each of the twelve mega-regional print book collections included well over 200 countries of publication, with the highest total (247) found in BOS-WASH. Similarly, the publications in each regional collection reflected a wide range of languages, although the cross-regional variation in the number of languages was higher than for countries of publication. The region with the most languages represented in its collection was BOS-WASH with 473; the regional collection with the fewest number of languages, DENVER, had no less than 265 languages represented. Figure 9 shows the percentages of each regional print book collection published outside North America, and published in a language other than English. As the figure illustrates, non-North American and non-English print book publications account for significant portions of each regional collection.

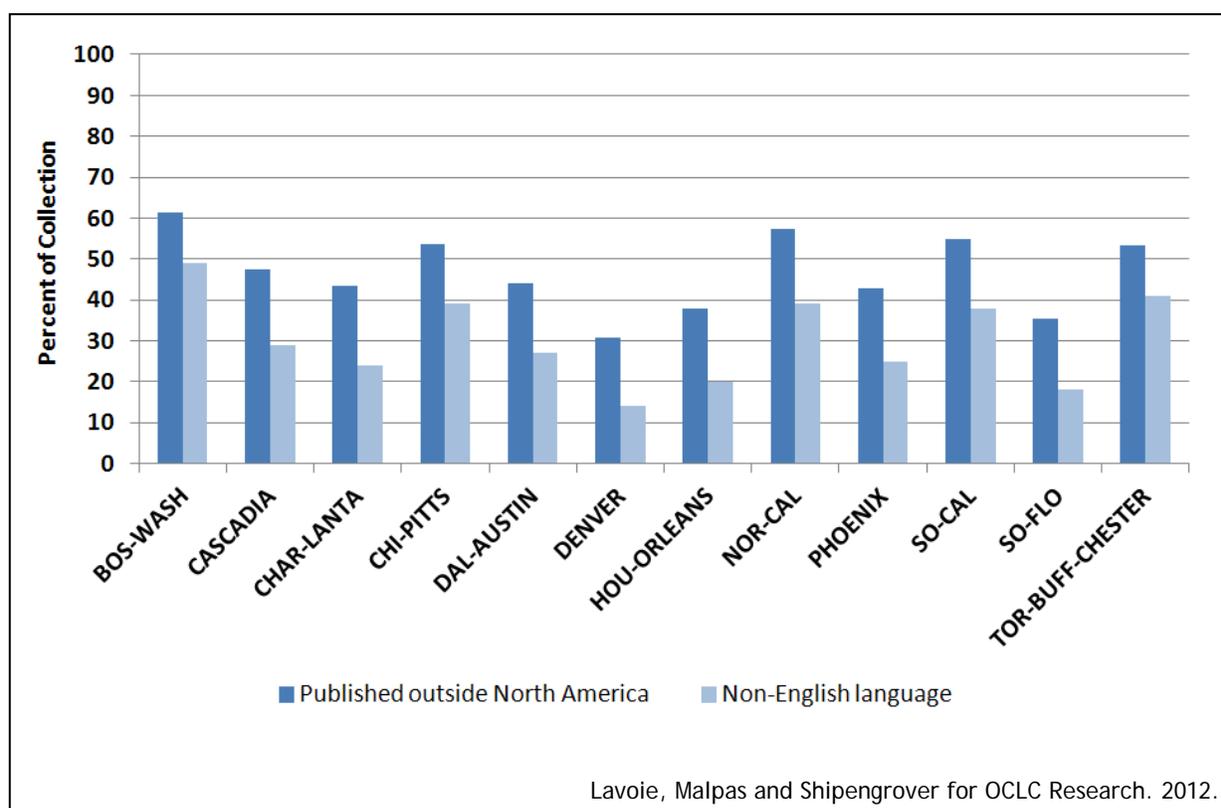


Figure 9. Global diversity in regional collections

The mega-regional print book collections display considerable similarity in regard to their subject make-up. A detailed subject analysis of the twelve collections could fill another

paper; here, we limit ourselves to a few rough but indicative measures, based on analyses using the OCLC Conspectus¹⁹ and FAST²⁰ topical subject headings in WorldCat records.

All of the regional collections had “Language, Linguistics, and Literature” as the most frequently assigned Conspectus division, and “History and Auxiliary Sciences” as the second most frequently occurring division. Ten of the twelve regions had “Philosophy and Religion” as the third most frequently occurring division. “Business and Economics” and either “Art and Architecture” or “Engineering and Technology” generally rounded out the top five Conspectus headings for the regional collections. The top five most frequently occurring divisions accounted for anywhere from 45 percent (BOS-WASH) to 60 percent (DAL-AUSTIN) of print books with Conspectus headings assigned. With the exception of BOS-WASH, the top five divisions accounted for more than half of the publications in the regional collections, suggesting a “long tail” shape to the distribution across Conspectus divisions in each region.

Similar results are obtained through an analysis of FAST topical headings, which characterize subjects at a more granular level. In each region, “political science” was the most frequently encountered topical heading, with little cross-regional variation in the top five topical headings. Expanding to the top 250 topical subject headings in each region provides further evidence of cross-regional similarity in subject representation. As table 4 shows, the degree of intersection between each region’s top 250 topical subject headings, and a benchmark top 250 list for the North American print book collection as a whole is substantial. No region overlaps less than about three-quarters with the benchmark collection, and for most regions, the overlap is much greater. In short, no region appears as an “outlier” in terms of an exceptionally distinctive subject composition for its regional print book collection.

19. The OCLC Conspectus is a subject hierarchy, ranging from broad to specific subject descriptions. The analysis focuses on Conspectus divisions, which are broad disciplines of knowledge. Our analysis is confined to print book publications in WorldCat that have an assigned Conspectus division, which includes the majority of the publications in each regional collection (for more information, see OCLC 2012, sec. 1.2).

20. FAST (Faceted Application of Subject Terminology) is a streamlined, simplified version of the Library of Congress Subject Headings schema (for more information, see OCLC Research 2011).

Table 4. Regional overlap of top 250 most frequently occurring topical subject headings with North American print book collection²¹

Region	Intersection	Overlap (%)
BOS-WASH	228	91
CASCADIA	203	81
CHAR-LANTA	206	82
CHI-PITTS	225	90
DAL-AUSTIN	199	80
DENVER	184	74
HOU-ORLEANS	192	77
NOR-CAL	211	84
PHOENIX	187	75
SO-CAL	204	82
SO-FLO	189	76
TOR-BUFF-CHESTER	212	85

One distinction that is discernable across regional collections in terms of subject composition is that regional collections tend to have a regional flavor. FAST subject facet distributions show that regional collections collect relatively heavily in subject areas related to the region itself, pertaining to geography, local history, local events, and so on. For example, in DAL-AUSTIN, books about Texas and the Mexican War (1846-1848) figure prominently relative to other regions; similarly, in SO-FLO, relatively heavy concentrations of books can be found about Florida and Cuba. So while all regions seem to collect materials in the same general topical areas (perhaps reflecting production trends or other meta-regional trends), each seems to also specialize in books about region-specific subjects.

Taken together, this stylized fact, which identifies the subject-based similarity between the regional collections, and the previous one that asserts that “rareness is common” both within and across regional collections seem at first glance contradictory. Yet these stylized facts co-exist quite easily. A useful metaphor to illustrate this is to consider the regional collections as a group of retail stores that, by and large, sell the same type of merchandise, but tend to carry different brands of that merchandise. So while all the regional collections have relatively similar subject distributions, they also display a significant degree of distinctiveness

21. Overlap computed using Whitehead (2012) comparison tool.

in terms of individual offerings within subject areas—that is, specific publications. Although we did not examine subject data at the level of an individual institution’s holdings, we would hypothesize that this phenomenon also holds within regions—i.e., similar subject distributions across institutions, but distinctive individual offerings within subject areas. This suggests that institutional collections consolidated at the regional level, or regional collections consolidated at a national or international level, result in a richer set of offerings within a shared pattern of collecting across broad subject areas.

Size is a driver for uniqueness, diversity, and age as characteristics of collections

The twelve regional collections vary considerably in size, with the largest, BOS-WASH, nearly seven times the size of the smallest, PHOENIX. Given that the larger regional collections by definition have more publications than smaller collections, it is easy to predict—and to confirm—that the larger collections will have greater numbers of publications that are unique to their region; greater numbers of publications originating from countries outside North America, or published in languages other than English; and greater numbers of older publications, than the smaller collections. Examination of the data shows, however, that in fact the difference between large and small regions on these points is more fundamental: differences can be detected not just in absolute terms, but in terms of proportions, which in turn suggests differences in collecting behavior.

The three largest regional collections—BOS-WASH, CHI-PITTS, and TOR-BUFF-CHESTER—unsurprisingly have the largest numbers of print book publications that are unique to their respective regions. By “unique” we mean that according to WorldCat holdings data, a particular print book publication is present in the collection of a single region. But these regions also exhibit the highest proportion of their collections corresponding to materials unique to the region: 32 percent for BOS-WASH; 16 percent for CHI-PITTS, and 24 percent for TOR-BUFF-CHESTER. On the other end of the spectrum, the three smallest regional collections—PHOENIX, DENVER, and SO-FLO—exhibit the three lowest proportions of regionally-unique materials: 2 percent, 2 percent, and 3 percent, respectively. This suggests that the presence of unique materials is proportionately less in smaller collections, or in other words, such materials are collected less intensively than in larger regions.

Similar results pertain to the presence of print books published outside North America, and those published in languages other than English. Again, the three largest regional collections exhibit high proportions of these materials, with BOS-WASH devoting 61 percent and 49 percent to non-North American and non-English materials, respectively; CHI-PITTS, 54 percent and 39 percent; and TOR-BUFF-CHESTER, 53 percent and 41 percent. The smallest collections, on the other hand, displayed relatively small proportions of these materials: PHOENIX, 43

percent and 25 percent; DENVER, 31 percent and 14 percent; and SO-FLO, 35 percent and 18 percent.

Figure 10 summarizes the percentages of each regional collection corresponding to publications that are regionally-unique; published outside North America; and published in a language other than English. The regional collections are ordered from largest to smallest. The figure demonstrates the strong correlation between size of the collection, and its propensity to be comparatively unique and globally diverse. There are some exceptions, of course—for example, the PHOENIX regional collection seems to be more globally diverse than its size would predict²²—but in general, larger collections have greater shares devoted to regionally-unique and globally diverse publications than smaller ones.

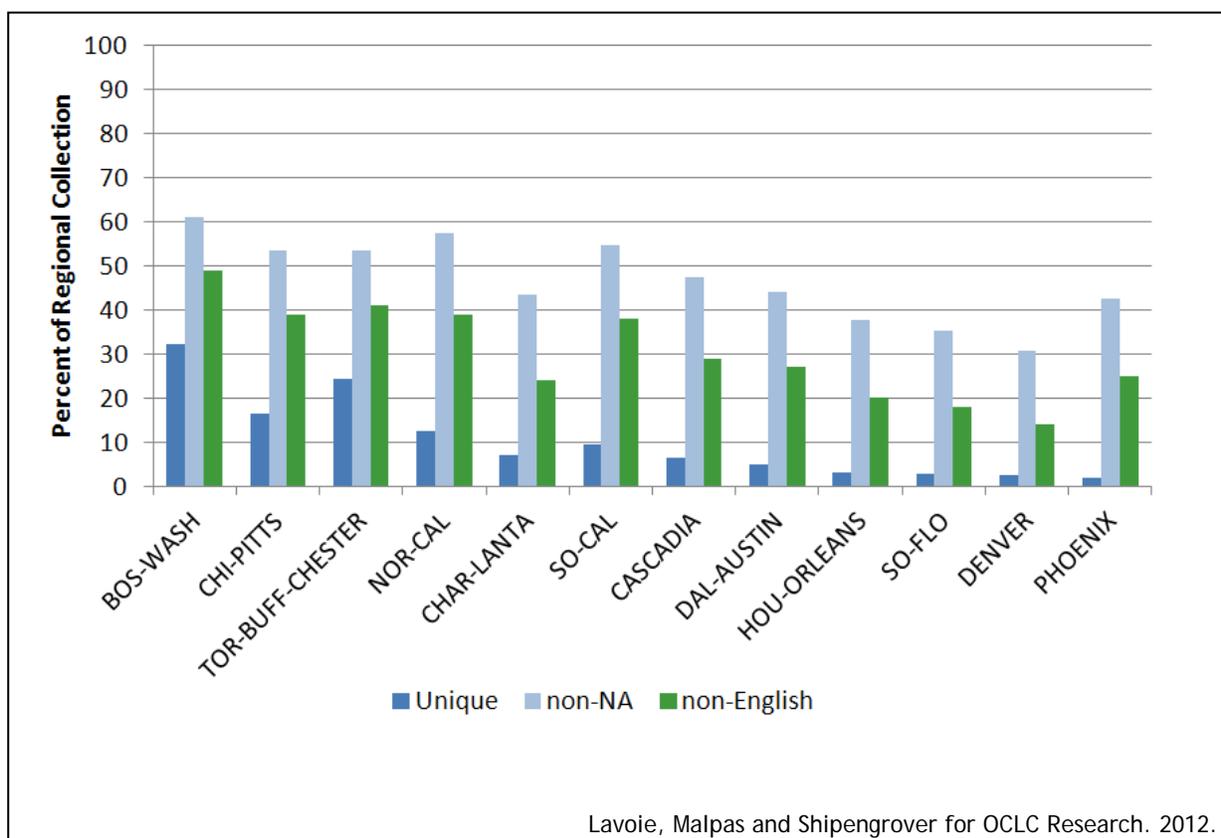


Figure 10: Uniqueness and global diversity as percentages of regional collections

22. The PHOENIX anomaly may be due to the relatively high proportion of ARL holdings associated with its collection. Fifty-two percent of print book holdings in the region belong to ARLs, the highest percentage of any region. One might expect that ARLs would tend to have higher proportions of unique and diverse materials in their collections, as well as higher proportions of older materials.

Median age of the print book publications in regional collections is also strongly correlated with size. BOS-WASH, the largest collection, also has the “oldest” collection with a median age (years since publication) of 34 years. CHI-PITTS, the second-largest collection, also has the second-oldest with a median age of 30 years. The “youngest” collections, on the other hand, are the smallest: PHOENIX at 23 years; SO-FLO at 24 years; and DENVER at 25 years.²³

In summary, our findings suggest that large regions not only collect larger numbers of unique and globally diverse materials, but in general they collect proportionally more of these materials than small regions. Moreover, the higher median age of the larger regional collections suggests they tend to either collect proportionally more older materials, or retain materials in their collections for longer periods—or both. We conclude, therefore, that regional collection size seems to be a driver for uniqueness, global diversity, and age.²⁴

The largest regional collections can serve as rough substitutes for smaller regional collections

Although each regional collection has a distinctive contribution to make to a North American network of consolidated regional collections, a pair-wise overlap comparison across regional collections reveals that the largest regional collections subsume most—although not all—of the print book publications available in the smaller regional collections. In this sense, the largest collections closely approximate, and therefore could serve as reasonable substitutes for, the smaller collections.

Bilateral overlap comparisons across the twelve regions reveal some interesting patterns. The BOS-WASH regional collection stood out as the collection subsuming the highest portions of the other regional collections. As figure 11 illustrates, BOS-WASH subsumed at least three quarters of nine of the other eleven regional collections. In addition, BOS-WASH accounts for 70 percent of the CHI-PITTS collection, and 65 percent of the TOR-BUFF-CHESTER collection. Some of the smaller regional collections are almost entirely subsumed within the BOS-WASH

23 CHAR-LANTA, a medium-sized collection, also is relatively young (25 years). An interesting outlier is NOR-CAL, which is the fourth largest collection but is tied with PHOENIX for the youngest collection with a median age of 23 years.

24. It should be noted that the largest regional collections are typically those where urban growth and associated educational and social infrastructure have built up over centuries of economic development. Thus, as one moves from East to West, the mega-region collections tend to decrease in size, scope and relative distinctiveness. To a significant extent, the size of regional collections is determined by historical factors that have shaped (and continue to alter) the identity of each mega-region. One might also argue that the diffuse nature of the “extra-regional” print book resource in North America is an artifact of social and cultural movements, including the institutionalization of land-grant universities and the library extension movement in the second half of the nineteenth century. Both of these contributed to the dispersion of library resources across geographic areas where urban development was limited. Additional study might show if the utilitarian orientation of these two movements has had a lasting effect on the complexion of the regional library collections. An overview of the Library Extension movement is provided in deGruyter (1980).

collection. For example, 95 percent of the PHOENIX collection, 93 percent of the DENVER collection, 92 percent of the HOU-ORLEANS collection, and 92 percent of the SO-FLO collection is duplicated within the BOS-WASH collection.

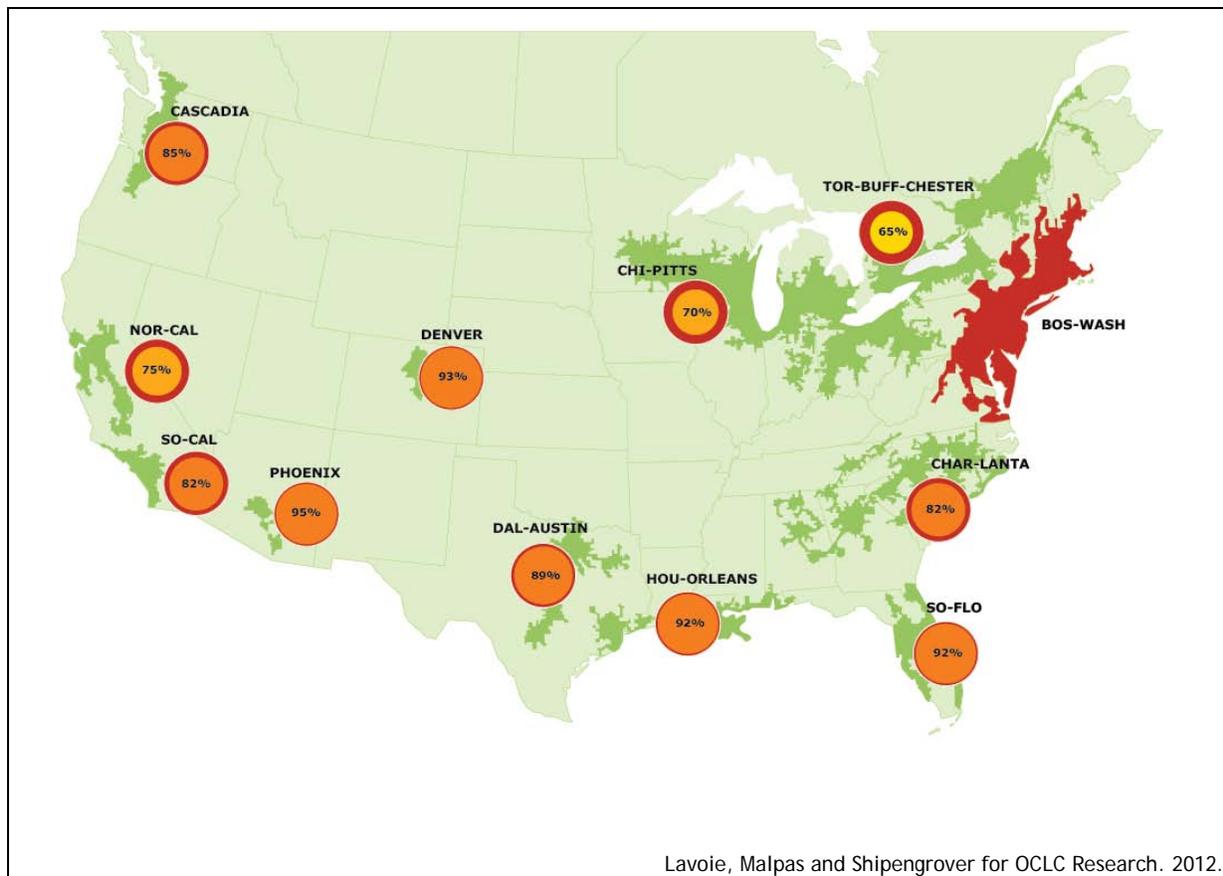


Figure 11. Bi-lateral overlap with the BOS-WASH collection, by region

More generally, we find that as a rule, the smallest regional collections are roughly duplicated by the largest collections. Figure 11 illustrates this in the context of the three smallest regional collections: PHOENIX, DENVER, and SO-FLO. Eighty percent or more of the PHOENIX collection is duplicated within six other regional collections; 80 percent or more of the DENVER collection is duplicated within five other regional collections (and 93 percent within BOS-WASH alone); 80 percent or more of the SO-FLO collection is duplicated within four other regional collections. Significantly, each of the three small regions is geographically near to at least one other region that overlaps with at least 80 percent of its collection. This would likely ease the logistical challenges involved with one of the small regions partnering with a larger neighbor to meet some part of its print book management and access needs.

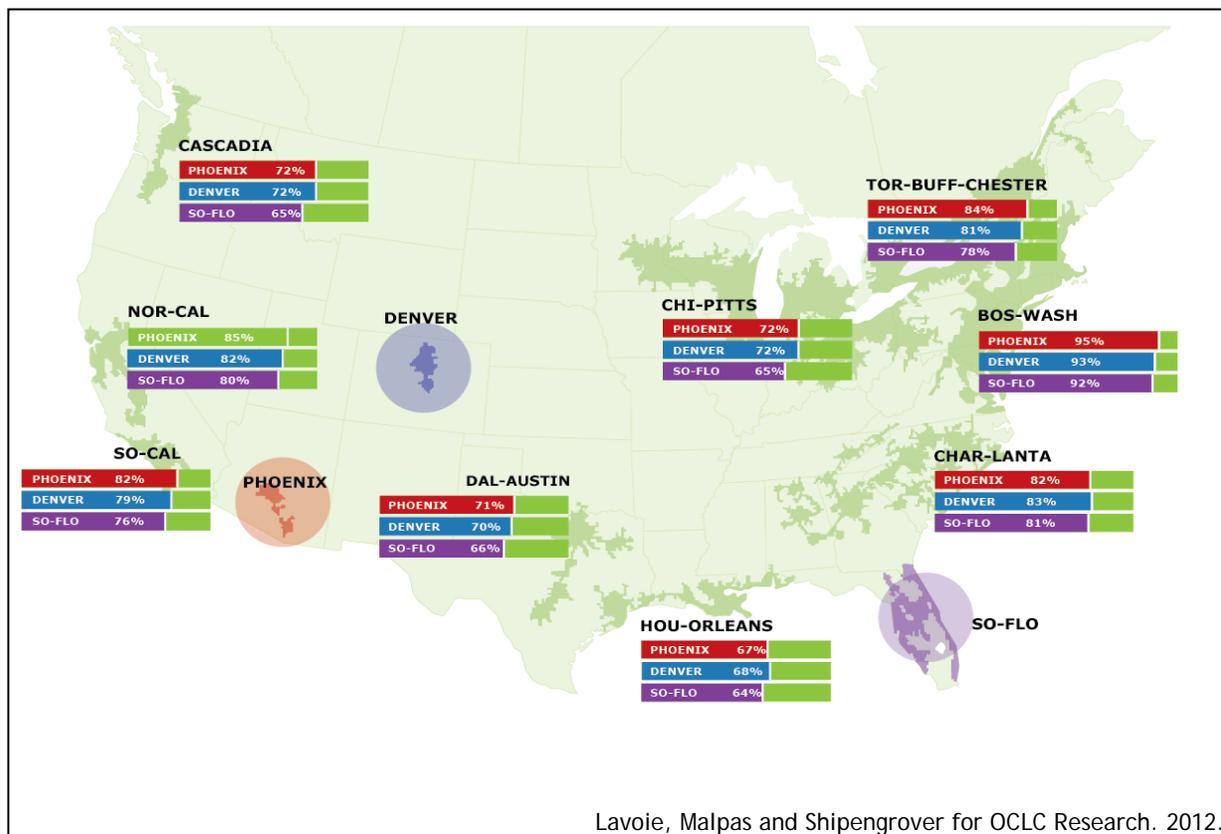


Figure 12. PHOENIX, DENVER, and SO-FLO overlap with other regional collections

An important take-away from this analysis is that all of the regional collections can be paired with another collection that overlaps significantly with the first collection’s holdings, and more specifically, the BOS-WASH collection is the highest overlapping collection for all the other regional collections (see figure 11)²⁵. But it is equally important to note that no regional collection is completely subsumed within a larger collection. Each collection includes a slice that is found nowhere else, and therefore represents a unique contribution to the overall North American print book collection.

Another way of considering cross-region overlap is to examine the degree to which the cumulative holdings of the regional collections progressively cover the extent of the overall North American print book collection of 45.7 million publications. Table 5 presents the cumulative coverage of the North American collection, by the twelve regional collections, ranked from largest to smallest.

25. CHI-PITTS is the highest overlapping collection for BOS-WASH at 50 percent.

Table 5. Cumulative coverage of the North American print book collection

Region	Print Book Publications	Percent of N. American	Percent increase
BOS-WASH	26,105,425	0.57	--
CHI-PITTS	31,699,504	0.69	0.12
TOR-BUFF-CHESTER	36,246,732	0.79	0.10
NOR-CAL	38,303,267	0.84	0.05
CHAR-LANTA	39,223,404	0.86	0.02
SO-CAL	40,310,654	0.88	0.02
CASCADIA	40,851,328	0.89	0.01
DAL-AUSTIN	41,245,684	0.90	0.01
HOU-ORLEANS	41,433,933	0.91	< 0.01
SO-FLO	41,598,163	0.91	< 0.01
DENVER	41,719,110	0.91	< 0.01
PHOENIX	41,800,348	0.92	< 0.01

The findings reported in table 5 suggest a highly skewed distribution of the North American print book resource across the twelve mega-regions, with the vast majority of the North American collection concentrated in a few of the largest regions, and these large regions concentrated in the eastern half of the US and Canada. The BOS-WASH collection, containing 26.1 million print book publications, alone comprises 57 percent of the overall North American print book collection. Adding the next largest collection, CHI-PITTS, increases coverage of the North American collection to 69 percent. The three largest regional collections together account for almost 80 percent of the North American collection, while the six largest collections cover nearly 90 percent. It is clear from table 5 that diminishing returns set in quickly as new collections are added to the cumulative total. Taken together, the twelve mega-regional collections account for 92 percent of the overall North American print book collection; the remaining 8 percent is located in the US and Canadian areas outside the mega-regions.

Digital surrogates exist for significant portions of the regional collections

Strategies chosen for managing legacy print collections will depend in part on the availability of digital surrogates for print book publications. Digitization of print books has been preceding apace for several years now, under the auspices of mass digitization

programs such as Google Books. As the volume of digitized materials has increased, services such as HathiTrust have emerged to manage and share this content. Reliable access to digital surrogates creates a strong incentive to reduce the scale of print book inventory managed locally.

HathiTrust is a digital archiving service that manages digitized materials on behalf of more than sixty partner institutions (primarily American universities). As of March 2012, HathiTrust reports its collection to include about 5.4 million digitized book titles.²⁶ Table 6 reports HathiTrust's coverage of each of the twelve regional print book collections.

Table 6. HathiTrust coverage of regional print book collections

Region	Publications	In HathiTrust	Share of Regional Collection
BOS-WASH	26,105,425	3,719,184	0.14
CASCADIA	6,987,064	1,977,901	0.28
CHAR-LANTA	10,156,810	2,240,706	0.22
CHI-PITTS	18,558,201	3,700,089	0.20
DAL-AUSTIN	6,383,756	1,790,966	0.28
DENVER	4,047,196	1,216,497	0.30
HOU-ORLEANS	5,162,621	1,470,582	0.28
NOR-CAL	12,481,999	2,927,296	0.23
SO-CAL	9,771,974	2,433,638	0.25
SO-FL	5,008,657	1,286,838	0.26
TOR-BUFF-CHESTER	14,699,921	2,827,021	0.19
PHOENIX	3,827,173	1,258,297	0.33

The results in table 6 indicate that the HathiTrust corpus of digitized books accounts for significant portions of the regional collections. The PHOENIX and DENVER regions stand out in this regard, with about a third of their respective collections represented in HathiTrust. Seven of the twelve regions have a quarter or more of their collections in HathiTrust, while all except two regions have at least 20 percent. The outlier is BOS-WASH at 14 percent, although this result is primarily due to the size of the BOS-WASH collection; in absolute

26. OCLC Research periodically compares digitized book titles in the HathiTrust collection to print book titles in WorldCat. An analysis of the HathiTrust collection as of late February 2012 identified about 4.9 million discrete book titles that could be mapped to print book titles (and holdings) in WorldCat.

terms, BOS-WASH has the largest number of publications with digital surrogates in HathiTrust.

Examination of the subject content of the portion of the HathiTrust collection overlapping with the North American print book collection suggests a weighting toward the humanities. The three Conspectus subject areas “Language, Linguistics and Literature,” “History and Auxiliary Sciences,” and “Philosophy and Religion” account for half of the publications in HathiTrust overlapping with a print publication in the North American collection; in contrast, these subject areas account for 32 percent of the North American collection itself. Several factors may account for this. One has to do with collecting patterns on the part of HathiTrust contributors. Most of the Hathi collection was contributed by large research universities, and there is some evidence that institutions of this kind tend to have these subject areas more prominently represented in their collections (Lavoie and Dempsey 2009). Second, there are indications that some of the contributors to the HathiTrust collection made a special effort to provide materials that were likely in the public domain. By definition, these would be older materials, many pre-dating 1923. It is possible that institutions are more likely to retain older humanities-related materials—e.g., literature, works of history, etc.—than older materials in the sciences, which have a faster rate of obsolescence.

Future strategies for managing print book collections will hinge on the library community’s capacity to provide access to alternative formats, including digital surrogates. Digitized texts offer a broad range of features and conveniences to readers in comparison to the print originals, and could support a transformation of library operating models, enabling broader access to the collective resource while also reducing costs associated with managing redundant physical inventory. The current state of HathiTrust coverage of the twelve mega-regional print book collections, and the North American print book collection as a whole, suggests that significant progress has been made in this regard.

Print book publications held outside the mega-regions in the US and Canada constitute significant collections in their own right, although diffused over more institutions and a larger geographical space.

The analysis in this report focuses on the consolidated print book collections of the twelve North American mega-regions. As noted above, 92 percent of the overall North American print book collection is represented within these twelve collections. The remaining eight percent is found exclusively in US and Canadian print book holdings scattered across the space between the regions. While 8 percent may seem a small proportion, in absolute terms it represents nearly 4 million print book publications that are not available in any of the twelve regional collections. Moreover, the general characteristics of the materials in the US and Canadian “extra-regional” collections resemble those of a large and small regional

collection, respectively. However, both of the extra-regional collections also have a few unique characteristics.

The US extra-regional collection consists of 15.7 million print book publications (distinct imprints or editions of books in printed form) which in comparison to the twelve mega-region collections, would make it the third largest collection after BOS-WASH and CHI-PITTS. The US extra-regional collection alone can account for about a third of the overall North American collection. The more than 217 million total holdings in the US extra-regional collection—the sum total of the number of print book publications held in each institutional collection in the US extra-regional space—exceeds the total holdings of every regional collection; BOS-WASH has the next highest total at 191.6 million. As a consequence, the ratio of holdings to print book publications in the US extra-regional collection (13.83) is more than half again as high as the regional collection with the highest ratio (CHI-PITTS at 8.94). This in turn suggests a relative abundance of print book inventory (and higher level of duplication) in the area between the regions in the US.

The US extra-regional collection resembles the larger regional collections in terms of several key characteristics. Fifty percent of the collection was published outside North America, and 33 percent consists of non-English language materials, which would place the collection among the upper half of the regional collections with respect to both characteristics. The median age of the collection is 31 years, which makes it the second oldest collection after BOS-WASH (34 years). Fourteen percent of the collection is unique to the US extra-regional area.

The Canadian extra-regional collection encompasses 5.8 million publications and 14.8 million holdings, which places it on a scale similar to the smaller mega-region collections. The Canadian collection alone would cover about 13 percent of the overall North American print book collection. The percentages of materials published outside North America (43 percent) and in languages other than English (25 percent) are more in line with the smaller mega-region collections than the larger ones. But the Canadian collection does depart from the smaller regional collections in terms of its median age, which is 30 years, placing it among the “older” regional collections.

Based on available evidence, print book inventory appears to be relatively scarce in the Canadian extra-regional collection, with 89 percent of the publications available at five or fewer libraries, the second highest percentage after the PHOENIX region. The Canadian region differs from smaller mega-region collections in that a relatively high percentage of its collection is unique to the area (15 percent). This percentage would place it fourth among the mega-region collections, after the largest regions BOS-WASH, TOR-BUFF-CHESTER, and CHI-PITTS, and slightly ahead of the much-larger US extra-regional collection.

The extra-regional collections in the US and Canada constitute important collections in their own right, representing rich collections of print book publications with significant uniqueness vis-à-vis other regional collections. However, these collections are dispersed over a considerable geographic area with no natural framework within which to consolidate them into a single virtual collection. Within the vast extra-regional space, a number of successful and sometimes overlapping inter-lending networks do exist. It remains to be seen if such networks could be effectively federated into a system that would leverage the highly diffused aggregate print book resource as a regional asset. In short, the extra-regional collections are a varied and abundant inventory that is not organized in a way that easily supports shared access or management. The total value of this resource is imperfectly reflected in current management and inter-lending systems; it is in effect an “unaddressable resource” in the network.

A number of significant concentrations of print books can be found outside the mega-regions. Figure 12 indicates the locations of the top five largest concentrations of print book holdings outside the mega-regions in both the US and Canada. In the US, the comparison was conducted on the basis of geographical areas associated with US Postal Service sectional center facilities (identified by the first three digits of a US ZIP code)(Wikipedia 2012a). In Canada, the comparison was based on a similar geographical unit, the forward sortation area, defined by the first three characters of a Canadian postal code (Wikipedia 2012). Although much smaller than mega-regions, these geographical units were the largest identifiable areas that are distinct from—i. e., do not overlap with—the mega-regions.

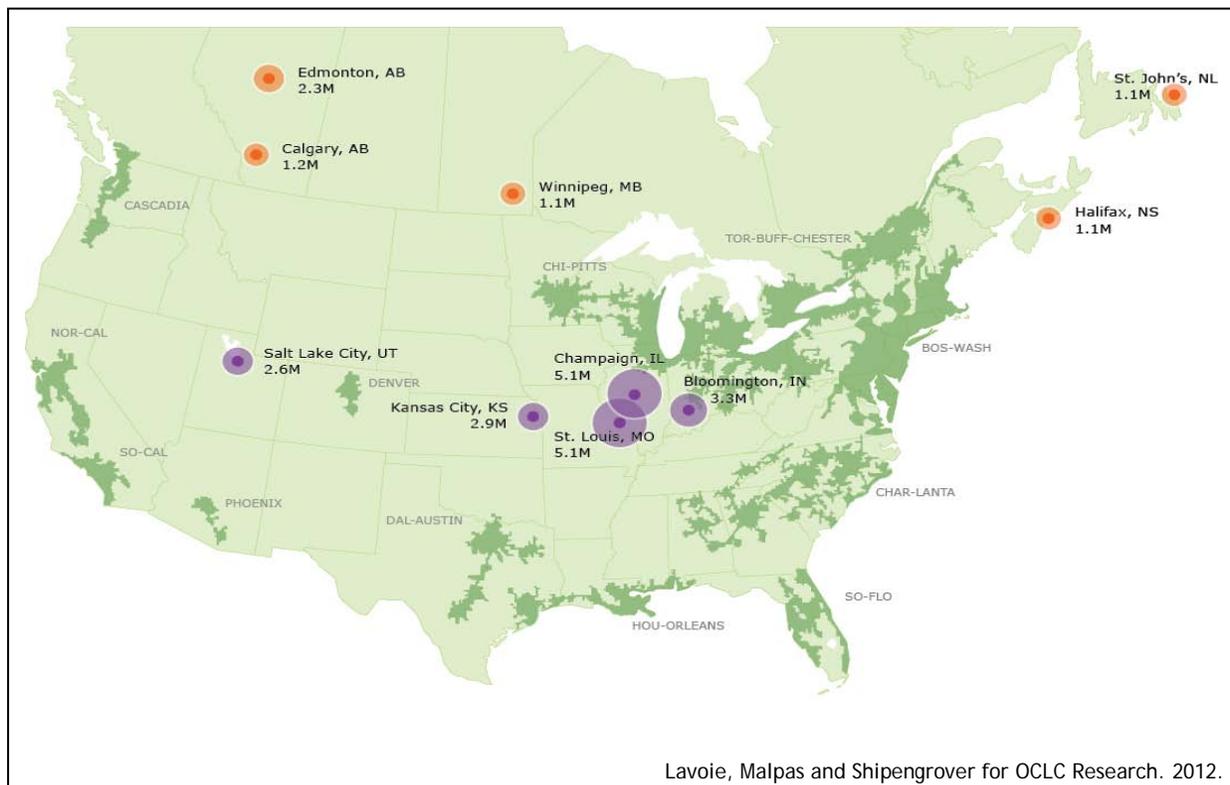


Figure 13. Top five concentrations of print book holdings outside the mega-regions, US and Canada

As figure 13 illustrates, the largest concentration of books outside the mega-regions is the St. Louis (Main) area, with more than 5 million print book holdings. Washington University and the St. Louis Public Library have significant holdings in this area, as well as St. Louis University and the University of Missouri, St. Louis. Champaign (North) in Illinois is not far behind: the University of Illinois, as well as three public libraries—Urbana Free Library, Champaign Public Library, and the Danville Public Library—all have significant print book holdings in this area. In Canada, the largest concentration of print books outside the mega-regions is found in Edmonton (North Capilano), with nearly 2.3 million print book holdings—the University of Alberta is located there. Calgary (Kensington/Westmont/Parkdale/University) is the home of the second largest concentration in Canada with nearly 1.2 million holdings; the University of Calgary is the key institution in the area.

It is interesting to note that in the US, the second- and third-largest extra-regional concentrations of print book holdings—Champaign, Illinois and Bloomington, Indiana—are situated very close to the edge of a mega-region (CHI-PITTS). It is easy to imagine that institutions in these locations would find opportunities to connect themselves to cooperative print management efforts within CHI-PITTS. Indeed, the geographical distribution of the

membership of the Committee on Institutional Cooperation (CIC), a collaborative body that includes the twelve Big Ten Conference institutions plus the University of Chicago, tracks very closely to the CHI-PITTS region. But the CIC also includes the University of Illinois and Indiana University, owners of the largest print collections in the Champaign and Bloomington areas, respectively. Moreover, CIC also includes more distant extra-regional institutions such as the University of Iowa and the University of Nebraska. Cooperative print management strategies within the CIC would therefore include institutions from outside CHI-PITTS, even though most members are located within the region or on its borders. In general, cooperative print management initiatives based on existing cooperative structures like CIC are likely to draw in partners from outside the mega-regions.

In all of the extra-regional concentrations of print book holdings shown in figure 13, an academic institution represents the entity in the area with the most print book holdings. This aligns with the stylized fact above that identifies academic institutions as the custodians of the majority of the print book inventory in the twelve regional collections. However, in all five of the US extra-regional areas shown in figure 13, a public library represents the second largest collection of print books in the area. In some of the areas, the share of holdings of this public library is significant: for example, 18 percent (Salt Lake City Public Library) and 15 percent (St. Louis Public Library). These results serve to remind us that public libraries are also key stakeholders in the future of print book collections. Regionally-based cooperative print management strategies should not forget the potential contributions of public libraries to print consolidation activities.

Key Implications

A number of implications for the future of print management emerge from the stylized facts described above.

Stewardship of the North American print book collection will require coordination on a supra-institutional scale.

Our study reveals that the geographic distribution of the aggregate North American print book collection is remarkably uneven, with the largest regional collections—all located in the eastern half of the US and Canada—accounting for the vast majority of print book publications and holdings in the North American collection. At the same time, we discovered that scarcity of holdings on a regional basis is not a reliable indicator of scarcity elsewhere in the larger library system. Taken together, these facts suggest that effective stewardship of the aggregate resource will require both a supra-institutional view of the

system-wide collection and cooperative agreements that transcend organizational and even geographic boundaries.

Individual institutions, and even regional cooperatives, are likely to misjudge the relative preservation risks to which their collections are exposed if they fail to appreciate the highly diffuse distribution of the collective print book collection. Just as supply chain management in the retail and manufacturing sectors is informed by a global view of demand, local and regional library inventory must be considered in the larger context of system-wide holdings and aggregate demand. Without this broader perspective, individual institutions run the risk of over-investing in local print preservation strategies at the expense of other, mission-critical services.

The appropriate scale of consolidation and cooperative management for regional print book collections will vary

There are many examples of collaborative print management schemes organized by consortia operating at the state and provincial level, from OhioLINK’s long-running effort to coordinate the purchase of books-not-bought-in-Ohio (Gammon and Zeoli 2003), to more recent state- and province-wide initiatives to identify and preserve “last copies” of books in Illinois, New Jersey and Ontario.²⁷ Libraries in Florida and Maine are currently working to develop state-wide strategies and shared infrastructure to manage local print and journal collections.²⁸ These initiatives will produce real benefits for libraries operating within their respective geographic areas; yet, as the mega-regions analysis suggests, the boundaries of the system-wide library print collection do not necessarily align with state borders. To the extent that mega-regions represent a natural unit of economic organization, undergirded by common social, cultural, and economic interests, one might expect cooperative print management solutions to operate at a similar scale.

Print books acquired by libraries in Northern California are, in the aggregate, different from print books acquired by libraries in Southern California; library collections in the western area of New York State will more closely resemble collections in Ontario than collections in Ithaca or Syracuse. Titles judged to be last copies on a state-wide scale may be relatively abundant if evaluated against regional or supra-regional holdings. Management schemes focused on rationalizing collections on a per-state level risk disrupting supply and demand patterns

27. Several state-based last copy policies, including those organized by CARLI in Illinois and by VALE in New Jersey, are identified in Malpas’s *Shared Print Policy Review Report* (2009). The Center for Research Libraries (CRL) maintains a registry of print archiving efforts that includes several state- and province-level initiatives, including the Ontario Council of University Libraries (OCUL) Thunder Bay Agreement (CRL 2012).

28. For details on planning for a shared storage facility for academic libraries in Florida, see University of Florida 2010. For information about the Maine Shared Collections Strategy project, see MSCS 2011.

operating on the higher, mega-region scale. A holistic approach to managing the collective resource will require cooperation across multiple states. This presents obvious challenges, as the social and technical infrastructure that is needed to support coordinated management of library resources at this scale is relatively limited. We consider this issue in further detail in the concluding section.

Paradoxically, the fact that “rareness is common” in the North American print book collection may constitute the single greatest imperative to collective action. While individual libraries are generally pleased to know that some part of their collection is distinctive, rare, or even unique, the capacity of regions to secure the long-term preservation of all such resources is limited. For regions with relatively few natural preservation partners—institutions with a mandate or other motivation to assume curatorial responsibility on behalf of the collectivity—stewardship may prove a burdensome challenge. The relatively scarcity of holdings (multiple copies of discrete titles) on a regional basis suggests that efforts to preserve distinctive resources will need to scale up to include multiple regions, if only to distribute the total cost of the curatorial enterprise.

As noted above, 75 percent or more of the print books in any of the North American regions are held by 5 or fewer libraries within that region. The sheer volume of material that would appear to warrant special preservation investment would overwhelm library budgets in any one region. Regional library cooperatives would arguably do better to concentrate their limited resources on the preservation of print book publications that are scarce or unique in the system as a whole, and redistribute the cost of preserving more ubiquitous material across a broader range of regional interests.

Significant bilateral duplication between mega-regional print book collections suggests opportunity for inter-regional cooperation in print management and fulfillment services.

While intra-regional supply of print books is generally limited, we discovered surprisingly high rates of pair-wise overlap between the largest mega-regional collections and those of smaller mega-regions. An implication of this is that smaller regions may find it advantageous to externalize some print management operations to larger regional partners, who can achieve greater economies of scale in preservation and access services. This has important implications for inter-regional shared print agreements and, more generally, for the limiting factors for large-scale print management regimes. If a robust, inter-regional library logistics infrastructure is available, it is conceivable that regions with relatively high rates of duplication vis-à-vis other regions may source preservation and fulfillment services elsewhere. While it is highly improbable that any region would outsource management of distinctive resources, it is logical to imagine that as fulfillment models continue to diversify—with an

increasing range of digital surrogate, print-on-demand, and inter-lending supply options available for a growing part of the retrospective print collection—strategies for sourcing commodity content will also change. Some efforts along these lines are evident with consortial and state- or province-wide licensing of digitized books in aggregations like Early English Books Online (EEBO). Content that was once difficult or even impossible to acquire in print, or regarded as a distinctive institutional asset, has become increasingly ubiquitous and as such lends itself to collective management.

A related implication is that mega-regions with a high rate of inter-regional duplication (PHOENIX and SO-FLO, for instance) may have greater options for sourcing cooperative print management solutions than regions with relatively low inter-region duplication. Larger mega-regions, which tend to have a lower inter-region duplication rate, are likely to be viewed as natural suppliers of print access and preservation, whether or not they actively seek to assume such a role. MINITEX, a resource sharing service funded by the Minnesota Office of Higher Education, offers a potential model for multi-state preservation and access services. Through state contracts, MINITEX provides libraries in North and South Dakota with inter-lending and document supply services that leverage the vast collections of the University of Minnesota-Twin Cities. Such multi-state resource sharing arrangements may have a downside for supplier institutions, however, as contractual obligations are likely to constrain their options for reconfiguring their own print book inventory.

A substantial part of each regional print collection has been digitized, creating opportunities for a virtual redistribution of system-wide assets; even so, major challenges for regional print management remain

Our investigation revealed a sizeable overlap between regional print collections and the mass-digitized corpus in the HathiTrust Digital Library, ranging from a low of 14 percent in BOS-WASH to a high of 33 percent in PHOENIX. As with the bi-lateral duplication analysis, where we observed that the largest regional collections subsumed most of the smaller collections, we found that smaller regional collections were more likely to be substantially duplicated in HathiTrust. This is a notable finding, as most of the content contributors to the HathiTrust corpus are located in the larger mega-regions and a majority of the contributed content was sourced from three zones: CHI-PITTS, NOR-CAL and SO-CAL.²⁹ It is remarkable, for example,

29. As of March 2012, there are 23 institutional content contributors to the HathiTrust Digital Library. One is located outside of North America. Of the remaining 22, about 25 percent are located in the BOS-WASH mega-region and an additional 40 percent are located in CHI-PITTS and CHAR-LANTA. If one looks at the relative contribution of content (digitized volumes) by mega-region, CHI-PITTS has a dominant presence, representing more than 50 percent of volumes in the aggregation. This is unsurprising, as the University of Michigan’s contributions alone account for almost 45 percent of the total. The strong representation of NOR-CAL and SO-CAL

that regions like PHOENIX and DENVER, which have (as yet) contributed no content to the aggregation, have a higher than average duplication rate with HathiTrust. An implication of this is that these smaller regions may stand to benefit more from the "replacement" value of HathiTrust than larger regions, assuming the digitized corpus is ultimately made available as a source of surrogate supply. This is a vexed issue at present, given uncertainties about the likely outcome of a legal dispute between rights-holders, Google, and HathiTrust.

The growing overlap between digitized books and retrospective print book collections may have another, equally transformative effect on the library system. As shown above, the supply profile of regional collections varies dramatically across North America, with a relatively small number of regions concentrated in the Northeast, Upper Midwest and (to a lesser extent) West coast holding a significant share of the aggregate resource. The increasing "dematerialization" of this content may ultimately diminish the striking asymmetries in supply that currently characterize the North American print book collection. This in turn may create opportunities for institutions in mega-regions with comparatively small regional print book collections, as well as those located outside of the mega-regions, as new fulfillment options emerge that level the playing field by reducing reliance on traditional print distribution networks.

As an increasing part of the source material that informs scientific research and inspires creative innovation moves online, the infrastructure that supports knowledge creation and information exchange will become less dependent on locally concentrated inventory and more reliant on systems that improve the flow of content, both in print and online. The degree to which libraries stand to gain or lose from this transition will depend in large measure on the success (or failure) of library fulfillment channels to compete with alternative supply chains. This is especially true for digitized print books, most of which are subject to copyright protections that prohibit electronic redistribution. Regional inter-lending networks might provide the backbone for a more robust library logistics network, enabling a greater number of libraries to benefit from a diminishing but still widely dispersed print book inventory. Looking forward, it is reasonable to imagine that regions that are successful in implementing "flow"-based strategies for print management that maximize the total value of the aggregate print book resource will have a competitive advantage over regions where management of print resources devolves to individual, differently equipped institutions.

content is explained by the very large contributions sourced from the University of California system. Estimates presented here are based on published figures as of February 2012. HathiTrust Statistics on contributions to the HathiTrust collection are published monthly (HathiTrust Digital Library 2012).

While the vast and still growing digitized corpus is likely to have a major impact on print management strategies on both an institutional and regional basis, one should not lose sight of the fact that a large part of the system-wide print book collection has not yet been digitized. As of March 2012, we calculate the total number of print book publications in the HathiTrust digital library to be approximately 4.9 million, representing only about 11 percent of the 45.7 million distinct print book publications in North America.³⁰ This estimate excludes titles digitized by Google and other agents that are not part of the HathiTrust collection, but nevertheless represents the core of the mass digitized resource that can be said to have a specific utility—long-term digital preservation—to institutions seeking to reduce local print preservation investments. Consequently, for the many millions of print book publications not replicated in digital archives like HathiTrust, the range of possible preservation strategies is relatively limited. Our application of the mega-regions framework suggests that institutions in the BOS-WASH, TOR-BUFF-CHESTER and CHI-PITTS regions—where the total print book collection is very large and the overlap with HathiTrust is lower than average—will have fewer preservation options available to them than institutions in PHOENIX, DENVER or HOU-ORLEANS. One can infer from this that large-scale, multi-region cooperative preservation strategies will be needed if the aggregate print resource is to be secured for the long term as a collective resource.

Changes in the organization of higher education will have a profound effect on academic library infrastructure and the disposition of print books in particular

While the majority of the North American print book inventory is held in academic libraries, the distribution by type of academic library differs significantly from one mega-region to the next. This has important implications for the redistribution of institutional investment in print collections, as colleges and universities with a primary mission of teaching and learning are increasingly focused on cost-effective provisioning of course materials, especially electronic content. Investment and attention that was once directed toward building and maintaining local print collections now has other aims. As a consequence, stewardship responsibility for the aggregate print collection is increasingly concentrated on a relatively small—and unevenly distributed—population of research-intensive institutions.

Academic research libraries, which typically view stewardship of the scholarly record as central to their academic mission, are challenged to uphold a preservation mandate that

30. As noted above, the figure of 4.9 million print book titles in HathiTrust is based on analysis of discrete book titles that could be identified in WorldCat as of March 2012. HathiTrust reported a total of 5.4 million book titles for the same period. The discrepancy in title counts reflect a difference in the way titles are disambiguated. Our analysis counts only those titles for which a discrete OCLC number corresponding to a print book title in WorldCat could be identified.

encompasses an increasingly diverse range of information resources. Print preservation is just one among many stewardship responsibilities, and at even the largest North American research universities it must compete with other institutional priorities for scarce attention and resources. Looking ahead, we can anticipate that changes in print management strategies in academic libraries outside the ARL sector will have a decisive impact on libraries within the ARL sector. As an increasing number of academic institutions begin to reduce retrospective inventory and purchase fewer print books, a small minority of libraries with a mission-driven commitment to preservation will feel compelled to step into the breach to ensure continuing access to and the long-term survival of the print published record.

In mega-regions where a majority share of the print book stock is managed by non-ARL academic libraries, it may prove difficult to identify an entity with a “manifest destiny” to assume responsibility for the regional print resource. An upside of this topsy-turvy world, in which the largest and best resourced libraries have less control over their destinies than nimble, smaller institutions that have divested from legacy print, may be that ARL institutions can capitalize on the increased reliance on their collections to negotiate new business and mutual aid agreements that may help redistribute the costs of long term preservation.

Resources located outside of the established mega-regions may prove difficult to mobilize as a collective asset

As noted in one of the stylized facts, a resource equal in size to the third largest mega-regional print book collection is located in the portions of the United States outside the twelve North American mega-regions. A significant print book resource also exists in the Canadian extra-regional area. While significant in terms of size (library holdings) and scope (number of publications), these extra-regional collections are purely notional, in the sense that they are not underpinned by any existing collaborative arrangements, shared infrastructure, or other substantive bonds of mutual interest. Nor would one expect that they would be, given that the extra-regional areas stretch across the lengths of the US and Canada, and encompass collecting institutions of all descriptions.

Mobilizing the regional print book collections as a collective resource will be difficult; mobilizing the print book holdings scattered across the areas outside the regions will be even more challenging. This suggests that the extra-regional print book resource is at greater risk than print books undergirded by the comparatively strong infrastructure—both physical and collaborative—found in the mega-regions that can serve as a starting point for building regionally-based cooperative print management strategies. Moreover, the print books distributed across the vast extra-regional area, especially those that are not duplicated in

other mega-regions, will be difficult to mobilize for use beyond their local settings in the absence of more robust library logistics.

In the US, a partial solution to this problem might lie in the fact that the largest extra-regional concentrations of print book holdings are often not far away from a mega-region. As figure 13 illustrates, three of the top five extra-regional print book concentrations are quite close to the CHI-PITTS region; the fourth (Kansas City) is situated mid-way between DENVER and CHI-PITTS, while the fifth is near DENVER. One could imagine clusters of print book holdings in the areas surrounding the borders of the mega-regions being incorporated into the collaborative print arrangements of the mega-region itself, although the “thickness” of this border area—i.e., the distance from the mega-region at which effective integration is still feasible—is an open question. In the case of CHI-PITTS, the border area encompasses several universities that participate in the Committee on Institutional Cooperation, which serves as the cooperative infrastructure for a large part of the regional research enterprise. Indeed, this regional infrastructure also serves a number of institutions located far outside the CHI-PITTS area, including the Universities of Iowa and Nebraska. Similarly, one can look to WEST, the regional print archiving effort, as evidence that institutions located in the extra-regional zone can leverage cooperative infrastructure that is anchored within an adjacent mega-region.

This solution works less well in Canada, where, as figure 13 illustrates, the largest clusters of print book holdings outside the mega-regions are more isolated. In these circumstances, another possibility might be the development of “mini-regions,” perhaps taking the form of a collaboration between institutions located in a metropolitan area and its immediate hinterland. For example, Edmonton is distant from any existing mega-region, but might consider a cooperative print strategy that includes academic and public libraries in Edmonton and the surrounding area, perhaps even extending down to include Calgary, another metropolitan area outside of a mega-region. In this way, locations outside the mega-regions might replicate to a degree the benefits inherent in regional cooperation, albeit at a smaller scale. One can point to the cooperative print strategy of the Ontario Council of University Libraries (OCUL), which leverages the strengths of institutions located in the metropolitan areas of Southern Ontario, as an exemplar of this kind of sub-regional strategy.

Canada is distinctive in having a strong regional cooperative library infrastructure that spans vast geographic areas, sometimes including universities and always encompassing populations located far from metropolitan centers, often with limited Internet connectivity. These supra-provincial agglomerations will be challenged to establish cooperative print management strategies that can balance the interests of libraries hoping to reduce their print footprint and communities that have a reasonable expectation of equitable access to information. For example, the Consortium of Prairie and Pacific University Libraries (COPPUL) supports 22

academic institutions in four Western provinces, covering a landmass of nearly 75 thousand miles—a significant part of the extra-regional Canadian print book collection. COPPUL members have initiated a plan for shared management of a print journal archive, focusing on titles that are available in electronic format (COPPUL 2011). Extending this model to the monographic literature may prove difficult, given the geographic scale of the region and the comparatively attenuated infrastructure available to support a “flow”-based model.

Whether through alliance with a geographically proximate mega-region, the cultivation of a “mini-region,” or some other strategy, the potential for considerable system-wide benefits is created by incorporating as large a share of the extra-regional print book resource as possible into some form of multi-institutional collaborative arrangement for print management and access. Library consortia serving institutions located outside the mega-regions will have the opportunity to model a variety of potential solutions and best practices.

Optimizing print collections on a mega-regional basis will have system-wide impact

The high concentration of print book inventory in BOS-WASH, CHI-PITTS and TOR-BUFF-CHESTER, and the relatively elevated levels of duplication in holdings across these regions, suggest that a large-scale cooperative effort to optimize and secure this resource against potential loss will deliver benefit to the library system as a whole. All libraries in North America benefit from the availability of these legacy collections, and their longevity will depend in part on the economic sustainability of the preservation and access that ensure their continued usefulness. Simply put, greater economies of scale can be achieved in regional approaches to print management that leverage existing infrastructure and inventory, than in approaches that require a massive redistribution of inventory.

Since there is a comparatively high level of duplication within and across these three mega-regions, and because most of the inventory is held by academic libraries that are, individually and collectively, revisiting their long-term investment in print collections, it will be necessary to guard against the uncoordinated withdrawal of materials upon which the North American library system as a whole depends. The consequences of a disorderly “draw down” of library holdings in BOS-WASH, CHI-PITTS and TOR-BUFF-CHESTER, which collectively hold 80 percent of the North American print book resource, would be felt across the library system as a whole. Thus, while significant system-wide benefit might be achieved through a deliberate rationalization of this supra-regional resource, significant harm may result if institutions continue to view print management (including the withdrawal of materials) as a purely local concern.

A recent survey of academic library directors in the US found that while about half of respondents felt that they lacked sufficient data to make informed judgments about the withdrawal of print journal back-files, more than 90 percent were planning or already actively engaged in projects to de-accession local serial holdings (Long and Schonfeld 2010). The same survey reported that, while academic library directors did not view a transition to reliance on e-books as imminent, a considerable majority (74 percent) “said that the withdrawal of print books would be an important strategy for their libraries in the future” if appropriate preservation and access mechanisms were in place (p. 36). The emergence of successful business ventures supporting the managed de-selection of print books in academic libraries suggests that this trend is already well underway, even where cooperative preservation infrastructure is still lacking.³¹

As libraries look to optimize the amount of local print inventory, it will be important to balance institutional imperatives to maximize library space recovery by reducing local physical holdings, with the core library mission of broadening access to information. Our investigation has revealed that each of the twelve mega-regions hold some distinctive print resources, assets that are not duplicated in any other regional collection and that add richness to the system-wide print book resource. Improving the “flow” or circulation of these resources—whether in print or digital form—will benefit the system as a whole by ensuring that the total value of library investment is effectively leveraged.

Looking at the aggregate print book collection from a supply-side perspective, it is clear that the relatively diffuse distribution of inventory is not optimized for fulfillment except perhaps at the local level. Inter-lending networks support a certain amount of load-leveling among potential suppliers, but they presently lack the system-intelligence or “shelf-awareness” that would enable a more dynamic alignment of supply and demand. As currently configured, library logistics are highly inefficient: the transaction costs of distributed print fulfillment are high (estimated at \$30 per inter-library loan) and relatively inelastic, in part because the sources of supply are not located where the demand originates.³² It is not possible to aggregate potential users around existing sources of supply, so instead solutions must be found to improve the flow of resources to users. A flow-based approach to consolidation that leverages aggregate demand will amplify the impact of library investment in collections

31. For a view of the growing market for services related to “managing down” print book collections in academic libraries, see Lugg and Fischer 2008, and Gilson and Strauch 2012. In January 2012, OCLC announced a strategic partnership with Sustainable Collection Services, LLC, an organization founded by Rick Lugg and Ruth Fischer to provide consulting and decision-support services to libraries reducing their print book holdings (OCLC 2012a).

32. The estimated cost of \$30 per transaction is based on total-cost calculations established in 1993 (Roche 1993). Nearly two decades on, the costs have scarcely changed: a recent article cites costs ranging from \$25 to \$40 per volume (Esposito 2012).

within and across mega-regions, while also enabling institutions located outside of these zones to benefit from greater economies of scale.

Conclusions

In this study we have applied the mega-regions framework to the distribution of the aggregate print book resource in North America, and considered some implications for cooperative approaches to print management. Among our key findings is that the variable distribution of print book publications (distinct imprints or editions of books in printed form) across mega-regions—with a few mega-regions accounting for a majority of the North American print book collective collection—along with variations in the characteristics of regional collections, is likely to result in divergent regional strategies for print management. At the same time, we find that the growing overlap between locally managed print book collections and collectively managed aggregations of digitized books may have a leveling effect, enabling regions with comparatively small collections to achieve greater efficiencies in print management than is feasible in the mega-regions with larger collections. We also find that the aggregate print book collection located outside of established mega-regions may prove difficult to leverage as a collective resource, except at a scale commensurate with existing social and technical infrastructure.

This report provides a supply-side picture of the North American print book collection, mapped against clusters of population and economic activity—i.e., the mega-regions. More work is needed to understand demand patterns within, across and outside of the North American mega-regions. It is hoped that regional consortia participating in intra-consortium borrowing programs—including Borrow Direct, OhioLINK p-circ, UC Request, and the Orbis-Cascade Alliance’s Navigator system, among others—will begin to share and analyze inter-lending data to build a common understanding of aggregate demand patterns within and across regions. Equally important is additional evidence and analysis to improve our understanding of how collections located in the vast extra-regional areas function in the larger picture of supply and demand. Does the higher rate of duplication in collections in this zone correspond to a greater aggregate (and more diffuse) demand than that found in the mega-regions? Or does the additional inventory dispersed across this less populous area compensate for weaker distribution networks? A more complete picture of the aggregate demand characteristics within and outside of the mega-regions will help to address the striking asymmetries in supply that the present study has revealed.

An important corollary to the findings discussed in this report is that existing cooperative infrastructure may not be equal to the task of managing print resources at a mega-regional scale. The absence of a cooperative infrastructure that is fit-to-purpose for achieving an

integrated regional print management strategy, or negotiating on behalf of regional partners, represents a significant constraint on the development of a system-wide, multi-regional preservation plan. It remains to be seen if existing organizational structures will adapt to serve the growing need for supra-institutional and supra-regional planning and governance, or if new organizations will be needed to bridge the gap between state- or province-level and larger-scale federal or national approaches to cooperative print management. To the extent that regional consolidation of print resources enables individual institutions to reduce costly duplication in infrastructure and management, while maximizing the value of the collective resource, it seems likely—if not inevitable—that appropriate cooperative infrastructure will emerge, whether through a pragmatic process of boot-strapping or a top-down initiative to institutionalize shared print management.

While a few library consortia can reasonably claim to represent the interests of some institutions within the mega-regions—the Committee for Institutional Cooperation (CIC) maps reasonably well to CHI-PITTS, the Association for Southeastern Research Libraries (ASERL) aligns with a large part of CHAR-LANTA, the Orbis-Cascade Alliance is a close match to CASCADIA—and while each of these consortia is actively pursuing regional collection management initiatives, none has explicitly embraced a membership model or governance structure that would enable a supra-regional approach to managing the aggregate print resource. Moreover, none of these consortia include non-academic (e.g. public or corporate) libraries in their membership, so the nature of the aggregate collection they can marshal as a cooperative resource is defined—and limited—by the collecting practices of academic institutions. One might argue that large-scale cooperative efforts are best organized within existing communities of interest: e.g., liberal arts colleges, research universities, or law schools. Indeed, one of the most successful approaches to scaling library service provision to a supra-regional level, the National Network of Libraries of Medicine (NN/LM) program, is explicitly limited to libraries serving the health services sector.³³

It remains to be seen what benefits might be achieved if the scope of cooperative print management schemes were broadened to include a wider range of library types, and partnerships are established among institutions that have not collaborated in the past. The present study suggests that unless multi-type partnerships are established, it will be difficult—if not impossible—to ensure that the remarkable breadth and diversity of the North American print book collection is preserved for future citizens and scholars. It is possible that existing organizational structures will alter or expand their remit to enable them to represent

33. The NN/LM includes more than six thousand health sciences libraries in the United States, encompassing university medical centers, teaching hospitals, and specialized research institutes. Network programs and services are coordinated by eight Regional Medical Libraries, which operate under contract to the National Library of Medicine. In 2011, the NN/LM initiated a regional print archiving program aimed at preserving at-risk medical journals (for more information about the MedPrint initiative, see NLM 2012).

the interests of other institutions in a common mega-region. But it is also possible that new organizations will emerge to fill the vacuum. This is essentially what has happened with the emergence of initiatives like WEST. Just as new cooperative structures have emerged to manage wildlife habitat, transportation systems, and high-speed computing networks on a regional scale, we can anticipate that management of library resources will increasingly require organizational structures that transcend existing political and institutional boundaries.

We do not claim that consolidation (physical or virtual) of print resources in North America should be organized at the mega-regional scale; rather, we have used the mega-regions framework to explore what a hypothetical regional consolidation of print resources might look like. Operationalizing a large-scale consolidation of regional print book collections will present many challenges; further study of the nature of these challenges and their potential solutions is needed. In particular, we hope this report will stimulate additional research on the factors that determine the appropriate scale of supra-institutional print management arrangements, as well as the cooperative and logistical infrastructure needed to sustain them.

References

- Bowker. 2012. "Bowker Releases Results of Global eBook Research." About us: Press. 27 March. http://www.bowker.com/en-US/aboutus/press_room/2012/pr_03272012.shtml.
- Conway, Paul. 2011. "Archival Quality and Long-Term Preservation: A Research Framework for Validating the Usefulness of Digital Surrogates". *Archival Science*. 11 (3-4): 293-309.
- COPPUL (The Council of Prairie and Pacific University Libraries. 2011. "Fact Sheet." http://www.coppul.ca/keydocs/Fact_Sheet_11.pdf.
- COSLA (Chief Officers of State Library Agencies). 2010. *COSLA: eBook Feasibility Study for Public Libraries Final Report*. Portland, Oregon: Pinpoint Logic. http://www.cosla.org/documents/COSLA2270_Report_Final1.pdf.
- Courant, Paul N., and Matthew "Buzzy" Nielsen. 2010. "On the Cost of Keeping a Book." In *The Idea of Order: Transforming Research Collections for 21st Century Scholarship*. (June): 81-105. Washington, DC: Council on Library and Information Resources. <http://www.clir.org/pubs/abstract/pub147abst.html>.
- CRL (Center for Research Libraries). 2012. "Print Archives Registry." *Global Resources Network*. <http://archivereg.crl.edu>.
- deGruyter, Lisa. 1980. "The History and Development of Rural Public Libraries". *Library Trends*. 28 (4): 513-23.
- Esposito, Joe. 2012. "PDA and Inter-library Loan" *The Scholarly Kitchen* (blog). 13 March. <http://scholarlykitchen.sspnet.org/2012/03/13/pda-and-inter-library-loan/>.
- Florida, Richard. 2008. *Who's Your City? How the Creative Economy is Making Where to Live the Most Important Decision of Your Life*. New York: Basic Books.
- Florida, Richard, Tim Gulden, and Charlotta Mellander. 2008. "The Rise of the Mega-Region." *Cambridge Journal of Regions, Economy and Society*. 1(3): 459-476.

- Gammon, Julia, and Michael Zeoli. 2003. "Practical Cooperative Collecting for Consortia: Books-Not-Bought in Ohio." *Collection Management*. 28 (1/2): 77-105. http://www.crl.edu/sites/default/files/attachments/pages/5_0.pdf.
- Gilson, Tom, and Katina Strauch. 2012. "ATG Interviews Rick Lugg and Ruth Fischer" *Against the Grain*. 24 (1): 34-36.
- HathiTrust Digital Library. 2012. "Update on February 2012 Activities." About. News and Publications. All Newsletters and Reports. 9 March. http://www.hathitrust.org/updates_february2012.
- Howard, Jennifer. 2011. "Building a Large-Scale Print-Journal Repository." Wired Campus (blog). *The Chronicle of Higher Education*. 8 February (5:33 pm). <http://chronicle.com/blogs/wiredcampus/building-a-large-scale-print-journal-repository/29558>.
- ITHAKA. 2012. "Preserving Scholarship." JSTOR. Preservation: About. <http://about.jstor.org/about-us/preserving-scholarship>.
- Lavoie, Brian, and Lorcan Dempsey. 2009. "Beyond 1923: Characteristics of Potentially In-Copyright Print Books in Library Collections". *D-Lib Magazine*. 15 (11/12). <http://www.dlib.org/dlib/november09/lavoie/11lavoie.html>.
- Long, Matthew P., and Roger C. Schonfeld. 2010. *Ithaka S + R library survey 2010 insights from U.S. academic library directors*. New York: Ithaka S + R. <http://www.ithaka.org/ithaka-s-r/research/ithaka-s-r-library-survey-2010/insights-from-us-academic-library-directors.pdf>.
- Lugg, Rick, and Ruth Fischer. 2008. "Future Tense—Weeding: The Time is Now." *Against the Grain*. 20 (4): 87-88. http://r2consulting.org/pdfs/2future_tense_lugg.pdf.
- Malpas, Constance. 2011. *Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment*. Dublin, Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2011/2011-01.pdf>.
- Malpas, Constance. 2009. *Shared Print Policy Review Report*. Dublin, OH: OCLC Research. <http://www.oclc.org/research/publications/library/2009/2009-03.pdf>.
- Maniatis, Petros, Mema Roussopoulos, TJ Giuli, David S. H. Rosenthal, and Mary Baker. 2005. "The LOCKSS Peer-to-Peer Digital Preservation System". *ACM Transactions on Computer Systems*. 23 (1): 2-50.

- MSCS (Maine Shared Collections Strategy). 2011. "Collaborating to Preserve Our Print Collections." University of Maine. <http://www.maineinfonet.net/mscs/>.
- Nelson, Mark. 2008. "E-Books in Higher Education: Nearing the End of the Era of Hype?" *EDUCAUSE Review* 43(2). <http://www.educause.edu/ero/article/e-books-higher-education-nearing-end-era-hype>.
- Nitecki, Danuta, Carol Jones, and Jeffrey Barnett. 2009. "Borrow Direct: A Decade of a Sustained Quality Book-Lending Service". *Interlending and Document Supply*. 37 (4): 192-198.
- NLM (US National Library of Medicine) "MedPrint—Medical Serials Print Preservation program." National Institutes of Health. Last updated 01 March. http://www.nlm.nih.gov/psd/print_retention_about.html
- OCLC. 2011. *Perceptions of Libraries, 2010: Context and Community*. Dublin, Ohio: OCLC. http://www.oclc.org/reports/2010perceptions/2010perceptions_all.pdf.
- OCLC. 2012. "Introduction to the WorldCat Collection Analysis Service: The OCLC Conspectus." *WorldCat Collection Analysis User Guide*. Sec. 1.2. http://www.oclc.org/us/en/support/documentation/collectionanalysis/using/introduction/introduction.htm#conspectus_WCA.
- . 2012a. "OCLC Establishes Strategic Partnership with Sustainable Collection Services." *News Releases*. 20 January. <http://www.oclc.org/news/releases/2012/20128.htm>.
- OCLC Research. 2011. "FAST (Faceted Application of Subject Terminology)" *Activities*. Last updated 18 November. <http://www.oclc.org/research/activities/fast/>.
- Rainie, Lee, Kathryn Zickuhr, Kristen Purcell, Mary Madden and Joanna Brenner. 2012. The Rise of E-reading. Washington, D.C.: Pew Research Center's Internet & American Life Project. <http://libraries.pewinternet.org/2012/04/04/the-rise-of-e-reading/>.
- Roche, Marilyn M. 1993. "ARL/RLG Interlibrary Loan Cost Study: A Joint Effort by the Association of Research Libraries and the Research Libraries Group." Washington, DC: The Association.
- Schonfeld, Roger C. 2011. "What to Withdraw? Print Collection Management in the Wake of Digitization". *Serials Librarian*. 60 (1-4): 141-145.
- University of Florida. 2010. "Shared Storage Facility for Library Materials." *George A Smathers Libraries*. Last updated 4 February. http://www.uflib.ufl.edu/communications/shared_storage_facility.html.

Whitehead (Whitehead Institute for Biomedical Research). 2012. "Compare Two Lists." Bioinformatics & Research Computing Tools. <http://jura.wi.mit.edu/bioc/tools/compare.php>.

Wikipedia. 2012. "List of Postal Codes in Canada." Last modified 6 May 2010. http://en.wikipedia.org/wiki/List_of_postal_codes_in_Canada.

———. 2012a. "List of ZIP Code Prefixes" Last modified 26 April 2012. http://en.wikipedia.org/wiki/List_of_ZIP_code_prefixes.

Wilkin, John P. 2011 "Bibliographic Indeterminacy and the Scale of Problems and Opportunities of 'Rights' in Digital Collection Building." *Ruminations*. February. <http://www.clir.org/pubs/ruminations/01wilkin>.