

The US and Canadian Collective Print Book Collection: A 2019 Snapshot

Brian Lavoie

Senior Research Scientist

© 2019 OCLC Online Computer Library Center, Inc.

This work is licensed under a Creative Commons Attribution 4.0 International License.

<https://creativecommons.org/licenses/by/4.0/>



September 2019

OCLC Research
Dublin, Ohio 43017 USA

www.oclc.org

ISBN: 978-1-55653-121-7

DOI: 10.25333/7zjv-jv94

OCLC Control Number: 1119391697

ORCID iDs

Brian Lavoie,  <http://orcid.org/0000-0002-7173-8753>

Please direct correspondence to:

OCLC Research

oclcresearch@oclc.org

Suggested citation:

Lavoie, Brian. 2019. *The US and Canadian Collective Print Book Collection: A 2019 Snapshot*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/7zjv-jv94>.

CONTENTS

Introduction	4
The US and Canadian Collective Print Book Collection	5
Regional perspective.....	6
Growth	7
Coverage	8
The State of the US and Canadian Collective Print Book Collection	9
Conclusion	11

INTRODUCTION

Imagine all of the print books held by libraries in the United States and Canada gathered together into a single collection. Surely such a collection would seem boundless, encompassing untold quantities of material, in comparison to which the holdings of even the largest libraries would be but a fraction.

Or . . . perhaps the scope of this collective collection would be limited by heavy overlap across individual collections. If so, perhaps the largest three or four library collections could collectively approximate the overall print book holdings of US and Canadian libraries?

Which of the two characterizations is right? Is this an answerable question?

This question is answerable, using the massive data aggregation of WorldCat. WorldCat is a shared registry of library collections around the world, in which nearly 450 million distinct publications are matched to more than 2.7 billion individual library holdings.¹ Within this data is an implicit map of the global distribution of information resources in libraries, which, in turn, can be used to circumscribe and explore collective collections clustered within the aggregate library resource.

In this position paper, we trace the contours of the US and Canadian collective print book collection—in other words, the collective print book holdings of all libraries in the US and Canada whose collections are registered in WorldCat. A collective collection perspective of this kind is valuable for two reasons. First, it demonstrates that the concept of a US/Canadian collective print book collection, far from being purely notional, can, in fact, be defined precisely and have its properties described using library data. Second, collective collections are growing in importance both as a source of intelligence about—and ultimately, the focus of—services that operate across collection boundaries, such as shared print management, group-scale discovery, and resource sharing.

In short, collective collections are all about scaling collections above the institution. Frequently, these collections are prospective, in the sense that they are not physically aggregated, or even brought together virtually through a layer of services. This is the case with the US/Canadian collective print book collection, so in this sense the findings presented below are a series of answers to “what if?” questions like the one that opened this position paper. Nevertheless, these questions help bring into focus the possibilities for creating value through services based on collections at scale.

1. As of January 2019. See OCLC. “WorldCat: Connect to the World’s Collected Knowledge.” <https://www.oclc.org/en/worldcat.html>.

The US and Canadian Collective Print Book Collection

As of January 2019, the US/Canadian collective print book collection accounts for **59.2 million distinct print book publications**,² based on **994.3 million holdings** in library collections across the United States and Canada.

The current size of the US/Canadian collection is a 5% increase over the previous year in terms of print book publications, and a 1% increase in terms of print book holdings. Growth in publications and holdings can be attributed to the acquisition of new print book publications by libraries, as well as WorldCat becoming ever more “aware” of the full extent of the existing collective print book collection as more libraries register their holdings in the database. While both print book publications and holdings exhibited increases, publications grew more robustly, with the result that the average print book publication now has slightly fewer holdings (16.8) than in 2018 (17.6). This suggests that, on average, a new addition to the collective print book collection is being duplicated in fewer local collections across the US and Canada.

A collection of more than 59 million distinct print books is certainly immense, although it may be difficult to place this in perspective. One way is to estimate the spatial footprint of such a collection. If a single copy of each distinct print book publication held by libraries in the US and Canada were physically collocated, approximately 1,400 miles (2,253 km) of shelving³ would be required to accommodate all of them—about the distance from New Orleans to Boston.

The number of distinct print book publications is the key unit of analysis for a collective collection. It shows the breadth and depth of the overall US and Canadian print book resource, with nearly 60 million different items.

It is worth noting that this is a narrow interpretation of the size of the overall US/Canadian collective print book resource, in that duplicate holdings across institutions have been eliminated; the nearly 1 billion print book holdings is a better estimate of the full print book inventory in US and Canadian libraries. But even this underestimates the total, because it does not account for libraries holding multiple copies of some print book publications.

For our purposes, the number of distinct print book publications is the key unit of analysis for a collective collection. It shows the breadth and depth of the overall US and Canadian print book resource, with nearly 60 million different items. In addition, we see a great deal of global diversity in this collection: nearly 480 languages and dialects are represented among these materials, and the books themselves originate from more than 270 countries and territories.

2. For the purposes of this study, a print book publication is defined as a distinct print book imprint or edition of a work.

3. Assuming eight volumes per linear foot.

This brief description of the US and Canadian collective print book collection illustrates a broader point: we can define collective collections more or less precisely, even at grand scales—as with the multinational collection that is our focus here. We can speak about these collections not just as a concept, but as a circumscribed aggregation of materials, constructed in data, and with analyzable properties. For example, in speaking of the US and Canadian collective print book collection, we can say more than “it is a lot of books”—in fact it is 59.2 million distinct publications. While this number is certainly large, it is also finite. This is a deceptively simple conclusion with powerful implications: a finite collection with discernable boundaries lends itself to tractable and meaningful strategies regarding its features, management, and future evolution.

Regional perspective

In addition to circumscribing the entire US and Canadian collective print book collection, we can also identify and explore clusters of print book holdings within the collection. One approach is to look at regional print book holdings across the US and Canada, and to do this, we utilize the mega-regions framework developed by the urban studies expert Richard Florida.⁴ A mega-region is a conglomeration of multiple urban centers and their hinterlands, bound together through infrastructure, mutual economic interests, and cultural similarities. Figure 1 illustrates the 12 US and Canadian mega-regions defined by Florida in his work, along with their respective collective print book collections.

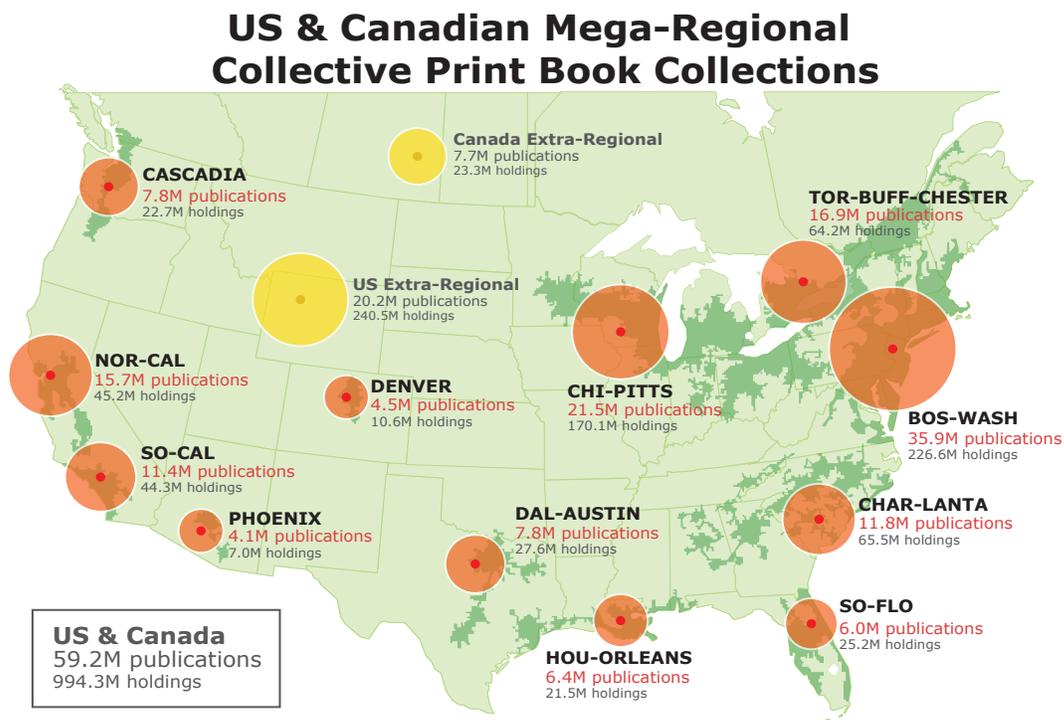


FIGURE 1: Map of US & Canadian mega-regional collective print book collections, as of January 2019.

4. For more information on Florida’s megaregions and their use in framing collective print book collections, see Lavoie, Brian, Constance Malpas and JD Shipengrover. 2012. *Print Management at “Mega-scale”: A Regional Perspective on Print Book Collections in North America*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C3133Z>.

The largest regional collective print book collection, in terms of both publications and holdings, is BOS-WASH, situated in the northeastern United States and encompassing several of the largest population centers in the country. BOS-WASH alone accounts for nearly a quarter of the total print book holdings in the US and Canada; even more noteworthy, the regional collective print book collection could stand in for nearly two thirds of the overall US/Canadian collection: 61% of the print book publications available in the US and Canada are held by at least one institution in the BOS-WASH region—a remarkable geographic concentration of information resources. Another exceptionally large regional collection is that of CHI-PITTS (17% of print book holdings, 36% of print book publications), covering a large swath of the US Midwest along with parts of Canada.

Figure 1 also includes the collective print book collections of US and Canadian institutions, respectively, whose location falls outside of a mega-region—that is, extra-regional. Both collections are significant, and in the case of the US, quite large. Extra-regional print book holdings account for about a quarter of total holdings for the US and Canada. In terms of print book publications, the US extra-regional collective collection could provide coverage of about a third of the overall US and Canadian collection.

Growth

Between January 2018 and January 2019, the number of distinct print book publications grew in all of the mega-regions, with CHI-PITTS, DENVER, and TOR-BUFF-CHESTER exhibiting the highest growth rate (5%). In terms of print book holdings, positive growth was observed in most mega-regions, but two—CASCADIA and NOR-CAL—saw their total print book holdings shrink over the previous year. Most mega-regions mimicked the same pattern seen with the overall US/Canadian collective print book collection: growth in print book holdings lagged growth in print book publications, indicating that the degree of overlap in print book inventory in most regions is diminishing.

A similar pattern can be found in the US extra-regional collective print book collection. In this case, the number of distinct print book publications in the collection grew by 3%, while the number of print book holdings grew at a rate less than 1%. In the Canadian extra-regional collection, growth in holdings kept pace with publications, with both increasing by 3% between 2018 and 2019.

With the exception of two regions (CASCADIA and NOR-CAL), print book holdings increased in every region, as well as in the US and Canada overall. Yet the fact that growth in publications exceeded growth in holdings indicates that the average number of holdings per print book publication is falling. This suggests a small but perceptible shift in the make-up of print book collections: on average, they are becoming a little more dissimilar to one another. In other words, new print book acquisitions in one collection are now duplicated in fewer collections elsewhere in the region, or even in the US and Canada overall. So regional collective print book collections—as well as the US/Canadian collective print book collection—continue to grow, but they are also becoming more “dilute,” by which we mean a lessening degree of duplication across collections.

While the scenario just described aligns with our findings, other factors may be at work as well. For example, cataloging lag—a delay between the time a publication is acquired and the time it is cataloged and registered in library systems—may be responsible for some of the difference in growth rates of holdings and publications. The existence of cataloging delay would lead to current holdings understating the number of libraries actually holding a particular publication, and therefore could dampen the observed growth rate of holdings vis-à-vis growth in publications. Nonetheless, the difference in holdings and publications growth is an interesting trend to watch over time.

Coverage

As we have seen, the US and Canadian collective print book collection is large and diverse, drawing on the collections of thousands of libraries. But can the scale and scope of the US/Canadian collection be duplicated in the collective collection of a subset of libraries, with the remaining libraries simply duplicating the publications in this “core” collection? To test this proposition, we constructed the collective collections of three prominent consortia of research universities: the 14 members of the Big Ten Academic Alliance (BTAA), the 13 members of the Ivy Plus Library Confederation (IPLC), and the 124 US and Canadian institutions comprising the Association of Research Libraries (ARL). Can any of these collections reasonably approximate the US/Canadian collective print book collection, within which they are fully subsumed?

BTAA and IPLC are groupings of peer institutions: BTAA includes the membership of the Big Ten athletic conference, most of which are public universities. IPLC includes the eight members of the Ivy League athletic conference, as well as five other leading universities. All of the members of BTAA and IPLC are located in the US.

BTAA’s collective print book collection includes 13 million distinct publications, accounting for 37.7 million holdings. The BTAA collection can therefore represent 22% of the publications in the overall US/Canadian collection print book collection, based on 4% of US and Canadian print book holdings. The IPLC collective collection is a little larger (19.5 million distinct print book publications, 53.8 million print book holdings), covering 33% of the print book publications in the US/Canadian collection, while accounting for 5% of total US and Canadian print book holdings. From these findings, it is clear that neither the BTAA nor the IPLC groupings can themselves reasonably approximate the overall US/Canadian collective print book collection.

Taken together, the print book collections of the ARL membership (which includes universities in both the US and Canada) form a collective collection of 36.6 million distinct publications, accounting for 258.7 million holdings. ARL libraries alone can therefore cover 62%, or nearly two-thirds, of the overall US and Canadian collective collection, while representing 26% of US and Canadian print book holdings. These findings highlight the particularly large and rich collections of the ARL membership.

If access to print books was restricted only to the collections of ARL libraries, one would still have access to about two-thirds of the distinct print book publications available across the US and Canada.

Note we are not saying that two-thirds of the US/Canadian collective print book collection resides only in ARL library collections; because the holdings of other libraries overlap with those of the ARL membership, most of these publications are available elsewhere as well. What we are saying here is that if access to print books was restricted only to the collections of ARL libraries, one would still have access to about two-thirds of the distinct print book publications available across the US and Canada. A similar interpretation should be applied to the findings regarding BTAA and IPLC coverage of the US/Canadian collective collection.

They also underscore another important point: even the collections of the leading US and Canadian research universities taken together cannot adequately approximate the entire US and Canadian print book resource. In short, the US and Canadian collective print book collection owes its size and diversity not just to the contributions of the largest collections, but to all collections. This becomes especially important when linked to our earlier finding that collective print book collections seem to be becoming more dilute, suggesting that the opportunity for every local collection to make a unique contribution to the collective collection will only grow.

The State of the US and Canadian Collective Print Book Collection

From our findings reported here, we can make several general observations about the state of the US and Canadian collective print book collection:

- The collection includes 59.2 million distinct print book publications, based on 994.3 million holdings, and is growing.
- The collection is becoming more dilute, in terms of holdings duplication.
- The collection is globally diverse in terms of country of publication and language of content.
- The collection's largest regional concentration is in the northeastern US. In most regions within the US and Canada, growth in print book publications exceeded growth in print book holdings.
- The collection's size and scope are a result of contributions from all libraries in the US and Canada, not just the leading research institutions.

Despite trends that seem to chip away at the traditional preeminence of print materials, the collective print book resource in the US and Canada is both vast and growing. However, the pattern by which the resource is distributed over the thousands of US and Canadian library collections seems to be shifting, as the average number of library holdings per print book publication is falling. We have characterized this as the collective collection becoming more dilute, in the sense that even as the collection expands in size, duplication of print book holdings is diminishing.

The breadth and scope of the US and Canadian collective print book collection are not derived from a subset of the largest local collections, but are instead obtained from the contributions of thousands of libraries all over the US and Canada.

It is remarkable to see the hundreds of countries of publication and languages of content represented within the US/Canadian collective print book collection. This finding, coupled with the sheer size of the collection, hints at the rich diversity of the publications in this collective collection.

Though the collection is physically located in the US and Canada, the boundaries of its content clearly extend beyond those two countries. Looking within the US/Canadian collective print book collection, we identified a number of regional concentrations of print books across the US and Canada. The largest by far is the BOS-WASH region in the northeastern US, with over 35.9 million distinct publications based on 226.6 million print book holdings. This means that over 60% of the print book publications available in the US and Canada are also available within this region alone. We also found that in most regions, growth in print book publications exceeded growth in print book holdings; this suggests that regional collective print book collections are also becoming more dilute in terms of holdings patterns.

The vast print book holdings available in various regions within the US and Canada are still not enough to reasonably approximate the overall US/Canadian collective print book collection. The largest regional collection—BOS-WASH—can provide coverage of less than two-thirds of all print book publications available in the US and Canada. Similar results are found when analyzing the collective print book collections of groupings of prominent research universities: the BTAA consortium could account for 22% of print publications available in the US and Canada; the IPLC consortium, 33%; and the ARL membership, 62%. These findings indicate that the breadth and scope of the US/Canadian collective print book collection are not derived from a subset of the largest local collections, but are instead obtained from the contributions of thousands of libraries all over the US and Canada.

CONCLUSION

In addition to highlighting key features of the US and Canadian collective print book collection, this position paper also provides an illustration of how collective collections can be operationalized by constructing them in library data. The key point is that collective collections of any scale can be circumscribed in sufficiently large aggregations of bibliographic and holdings data, such as WorldCat. Understanding collective collections is increasingly important both as the focus of library services, and as a source of data-driven intelligence for library decision-makers.

For more information about our work please visit our website at:
oclc.org/research



6565 Kilgour Place
Dublin, Ohio 43017-3395

T: 1-800-848-5878

T: +1-614-764-6000

F: +1-614-764-6096

www.oclc.org/research

ISBN:978-1-55653-121-7
DOI:10.25333/7zjv-jv94
RM-PR-216308-WWAE 1909