

# Reflections on Collective Collections

**Brian Lavoie**

Senior Research Scientist

**Lorcan Dempsey**

VP, Membership and Research, OCLC

**Constance Malpas**

Manager Strategic Programs

Preprint accepted for publication in: *College & Research Libraries*  
Volume 81, number 6 (September 2020).



Please direct correspondence to:

OCLC Research  
[oclcresearch@oclc.org](mailto:oclcresearch@oclc.org)

Suggested citation:

Lavoie, Brian, Lorcan Dempsey, and Constance Malpas. 2020. "Reflections on Collective Collections."  
*College & Research Libraries* 81, no. 6 (September). Forthcoming.

OCLC Research Preprint. Accepted 13 December 2019. <https://doi.org/10.25333/t51w-b252>.

## Contents

Abstract.....	5
Acknowledgements.....	5
Introduction .....	5
What is a collective collection?.....	6
Collective collections: Building a body of evidence .....	8
General patterns in collective collections.....	9
1. Collective collections can be circumscribed, quantified, and visualized .....	9
2. Rareness is common .....	11
3. Coverage requires cooperation .....	12
4. The details are in the scale.....	15
5. Collective collections enable a distant reading of the published record.....	17
Conclusion.....	18

## Figures

Figure 1: Timeline of Selected OCLC Research Collective Collection Reports and Trends .....	8
Figure 2: US and Canadian Collective Print Book Collections .....	10
Figure 3: Scaling the Collective Print Book Collection: A CIC (now BTAA) Perspective .....	13
Figure 4: Scaling the Collective Print Book Collection: A Canadian Perspective .....	14
Figure 5: RLUK Collective Print Book Collection: Duplication Rates .....	17

## Abstract

Collective collections are multiple local collections described and/or managed as a single collection. Constructing, understanding, and operationalizing collective collections is an increasingly important aspect of collection management for many libraries. This article presents some general insights about collective collections, drawn from a series of studies conducted by OCLC. These insights identify salient characteristics of many collective collections and serve as a starting point for developing collective collection-based strategies for such library priorities as shared print, digitization, and group-scale discovery and fulfillment.

## Acknowledgements

The authors are deeply grateful to several anonymous reviewers, as well as editor Wendi Kaspar, for comments and suggestions that greatly improved the paper. We also thank the many collaborators, including OCLC staff and members of the OCLC library cooperative, who have contributed to our collective collections research.

## Introduction

*Collective collections* are the combined holdings of a group of libraries, analyzed and possibly managed as a unified resource. Collective collections are an important concept in a library environment that favors collaboration and coordination, and where libraries seek to create value through collective action and shared capacities. Library infrastructure, services, and expertise are increasingly deployed at scales above the level of a single institution. Collective collections are library collections at scale.

Operationalizing collective collections is a matter of growing importance for libraries. In the area of shared print, for example, organizations such as EAST (Eastern Academic Scholars' Trust) and WEST (Western Regional Storage Trust) have emerged to manage regional collective print collections, providing governance and coordination structures around collective print holdings.<sup>1</sup> Issues like securing print retention commitments, load balancing intra-group circulation of print publications, and determining the minimum number of copies needed to support preservation and access are important aspects of operationalizing collective collections organized around shared print initiatives. In these and other matters, intelligence derived from analyzing collective collections can be instrumental in developing effective solutions. More generally, the ability to create and analyze collective collections at a variety of scales increases *system-wide awareness*, and provides intelligence *to fuel data-driven decision-making* in a landscape where conscious coordination of library collections—that is, “deliberate engagement with—and growing dependence on—cooperative agreements”—is increasingly important.<sup>2</sup>

Given the importance of collective collections both as a source of intelligence about group-scale holdings, and as an operational scale for library services, staff at OCLC Research (a unit of OCLC) have been constructing and analyzing collective collections for more than a decade, using data from the WorldCat database.<sup>3</sup> These studies have looked at collective collections at many scales, ranging from small groups to regional, national, and even global scale. The studies have also examined collective collections in a variety of national and international settings. A key takeaway from this work is that a collective collection—no matter the scale or geographic context—is something that can be circumscribed in library data, quantified, and analyzed, which in turn supplies valuable intelligence in support of local collection management as well as new forms of library collaboration.

The majority of OCLC Research's collective collections work examines the print book holdings of various groups of libraries, with the purpose of illuminating and contextualizing issues having to do with new strategies for managing legacy print collections. Although these studies are sometimes rooted in and intended to inform the interests of particular groups or communities—for example, recent analyses of the collective print book collections of the Big Ten Academic Alliance (BTAA) and Research Libraries UK (RLUK) consortia supported their thinking about shared print strategies—a prevailing goal across all of the work is to draw out general insights about collective collections and their implications for library strategic interests and planning.

This article gathers together some of these insights about collective collections, building a general picture of library collections at scale and connecting it to current areas of library interest. The application of collective collections to shared print concerns and priorities is an important source of these insights, but they also draw on studies addressing mass digitization, copyright, and cultural patterns and trends in the published record. The foundation of all the studies is aggregated bibliographic and holdings data; a key purpose of this article is to underscore how library data can be used to construct, analyze, and derive practical intelligence from collective collections. Understanding the general features of collective collections, as well as how they manifest in different contexts, is a first step toward effectively managing library collections at scale.

### **What is a collective collection?**

The phrase *collective collection* was introduced by Lorcan Dempsey,<sup>4</sup> and has since become a term of art in library circles.<sup>5</sup> A collective collection is the combined collections of two or more institutions, viewed as a single, distinct resource, usually through aggregation and analysis of metadata about the collections. Put another way, a collective collection elevates the concept of a library collection to scales above a single institution, extending its boundaries to encompass the resources concentrated among a group of libraries; these resources are then treated as a distinct collection in their own right.

Collective collections are not simple summations of library holdings: the holdings are combined, and then duplicate holdings are removed to yield the collection of *distinct* publications held across the library grouping. For example, if multiple libraries hold the same edition of *A Tale of Two Cities*, this edition would only count once in determining the size of the collective collection.

The idea of scale is central to collective collections, and it can slide anywhere along a continuum bounded by the endpoints of *two libraries* and *all libraries*, depending on context. In this way, the combined collections of the Five College Consortium constitute a collective collection, as do the combined collections of all public libraries in the United Kingdom,<sup>6</sup> and the combined collections of all libraries everywhere.

It is important to note that the *collective* aspect of collective collections pertains strictly to scale, and makes no assumptions about the existence of shared ownership or stewardship arrangements, nor how or even whether the collective collection is accessed or disclosed. Nor do collective collections necessarily involve the entire library collection: they may be—and often are—limited to specific classes of materials, such as print books or digitized images. Indeed, the collective nature of the aggregations could manifest anywhere on a spectrum of progressively deeper levels: e.g., discovery, consolidation, or ownership.

Collective collections can exist either physically or virtually. For example, HathiTrust's corpus of digitized monographs<sup>7</sup> is a collective collection, physically gathered from contributors around the

world into a centralized digital curation facility. Similarly, the BTAA's Shared Print Repository<sup>8</sup> is another collective collection, in this case consisting of volumes of print journal backfiles, occupying a centralized storage facility at Indiana University. The HathiTrust and BTAA collective collections – as well as others – share the characteristics of being *physically* aggregated from contributions from multiple institutions and managed under a shared stewardship arrangement.

Collective collections can also exist as a layer of data and services laid across distributed collections. For example, each member of the aforementioned Five College Consortium<sup>9</sup> maintains a local collection, but their combined collections are presented to member institutions' faculty and students as a distinct collective collection for discovery, access, and use.<sup>10</sup> In a similar way, the UK Research Reserve (UKRR) is a “collaborative distributed national research collection” which ensures that at least two copies of low-use print journal titles are available within the collections of UKRR member libraries.<sup>11</sup> In these examples, the collective collection is aggregated virtually, with its components physically distributed across multiple collections. Technologies such as Z39.50 and consortial borrowing applications help institutions merge distributed collections into virtual collective collections.

In OCLC Research's studies, collective collections are *constructs in data*: notional collections built to answer “what if” questions, provide useful intelligence to inform a wide range of library decision-making needs, and to detect patterns and trends within library collections at scale. The key difference between collective collections of this kind and the examples mentioned above is that in the latter cases, the collective collection is formally recognized and managed as a distinct resource; in contrast, collective collections constructed for research purposes are often putative, and not managed as a cohesive unit, either through physical or virtual aggregation.

WorldCat<sup>12</sup> is an important data resource in constructing collective collections. Using a combination of bibliographic data (i.e., *what* is in collections) and holdings data (i.e., *who* possesses at least one copy), a collective collection can be generated based on any group of libraries whose collections are represented in WorldCat. This data-driven view of a collective collection can then be mined for both customized and general intelligence touching on a range of library interests and decision-making needs.

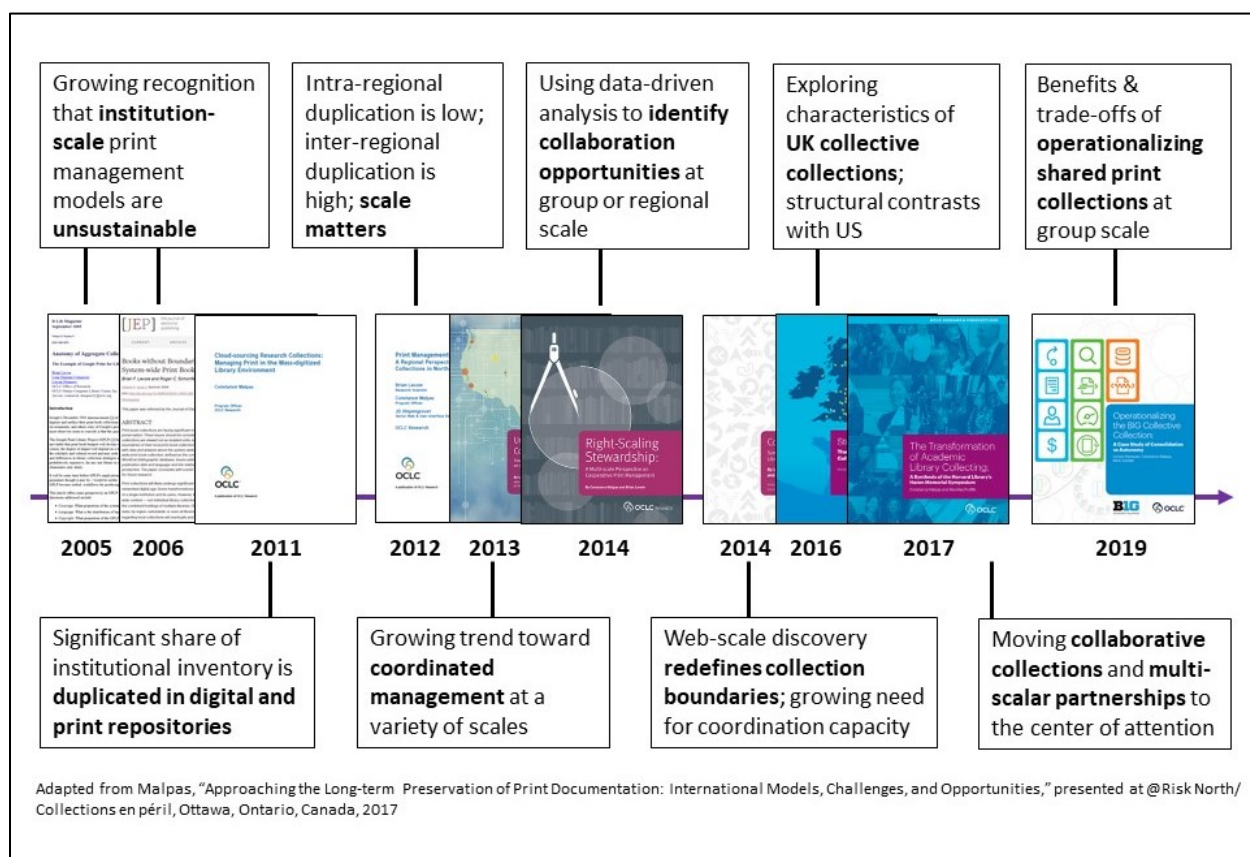
While WorldCat is well-suited as a data source for collective collections work, it is not perfect. Not all libraries have their collections represented in WorldCat, and for those that do, not all of these collections are represented comprehensively. Coverage of library collections in WorldCat is skewed toward North America. In light of this, it is important to take into account WorldCat's characteristics when interpreting findings derived from WorldCat data.

WorldCat's constraints illustrate a broader point: the limits of defining collective collections are set by data availability. As more and more library collections are fully and accurately registered in metadata aggregations like WorldCat, the opportunities to build and derive intelligence from collective collections increase. More data about individual collections means more data about the collective collections in which those collections are potentially embedded.

## Collective collections: Building a body of evidence

Taken together, the OCLC Research collective collection studies form a body of empirical evidence about library collections at scale. As mentioned, most of these studies focus on the print book holdings of academic libraries, paralleling a lively and ongoing discourse within the library community about the future of print collections. Figure 1 presents a timeline of some of our collective collection studies, linked to some of the adjacent thinking about print collections.

**Figure 1: Timeline of Selected OCLC Research Collective Collection Reports and Trends**



Around the time that the earliest of these studies were published, a view was emerging that legacy print collections, in their current form, were unsustainable for many academic institutions. The environment for academic libraries was challenging, with budgets stagnant or shrinking, at the same time that new demands were being placed on library resources in areas such as research support, data management, and publishing. Optimizing the use of space within the library was also a concern, in the face of increased demand for learning and work spaces.<sup>13</sup> But scarce resources sorely needed for new library focuses were often tied up in print collections with large physical footprints and declining use. Evidence accumulated showing significant overlap across collections, suggesting potential efficiencies in collection management strategies informed by a system-wide perspective.<sup>14</sup> The maturation of web-scale discovery blurred the boundaries of local collections; users could now search group-scale collections, sharpening the need for robust resource sharing networks and closer coordination of collection management strategies. All of this, in turn, led to strong interest on the part of academic libraries in



coordinating print collection management at a variety of scales, including small groups, consortia, regions, and even nationally.

This was the context that motivated many of OCLC Research's studies of collective collections. The studies have focused on descriptive analysis of collective collections – size, duplication rates across collections, salient characteristics such as languages or subject strengths, etc. – with the purpose of providing a data-driven view of the collections around which collective action might take place. In addition, especially in later studies, OCLC Research has also explored the challenges of *operationalizing* collective collections. For example, robust collaborative infrastructure (i.e., governance or decision-making mechanisms) is needed to coordinate and support collective collections as a shared resource, but this infrastructure may not exist at the needed scale. OCLC Research's study of US and Canadian mega-regional collective collections<sup>15</sup> highlighted both potential efficiency gains through de-duplication, as well as the unique collecting strengths of each regional grouping of institutions, but noted the lack of collaborative infrastructure in place to support operationalizing collective collections at mega-regional scale. Another report, focusing on the collective print collection of the BTAA consortium<sup>16</sup>, found that even when a collaborative apparatus is in place (i.e., the consortium), important tradeoffs must be confronted, such as that between consolidation and autonomy.

Analysis of collective collections both reflects and encourages contemporaneous thinking about print collections. In addition, it builds up a body of empirical evidence that informs library decision-making about collective collections and strategies for managing them. Ohio State University's Karla Strieb observes:

It has become possible only recently to begin describing the collective collection that could be shared by libraries. Perhaps the most influential descriptive studies have come from OCLC Research, which has now shared reports outlining both levels of uniqueness as well as duplication among various aggregations of library collections. This growing body of computationally intensive analysis of the collective collection has also begun to clarify geographic distribution and other key characteristics of library collections relevant for decision making around coordinating activities. There is new understanding of collections at scale. Libraries can better assess past successes in coordination and cooperation and surface the outlines of new frontiers for collaborative activities as well as clarify potential efficiencies and opportunities.<sup>17</sup>

The remainder of this paper summarizes some of the key findings from OCLC Research's studies of collective collections.

## **General patterns in collective collections**

The evidence accumulated from OCLC Research's collective collections work has been drawn from studies of collective collections scoped at a variety of scales, and framed in different institutional and geographic contexts. Nevertheless, in considering this work as a whole, some general patterns emerge that describe in broad brushstrokes several salient features of collective collections that transcend context. This section describes these patterns, provides examples from the studies, and links them to some areas of strategic interest to libraries.

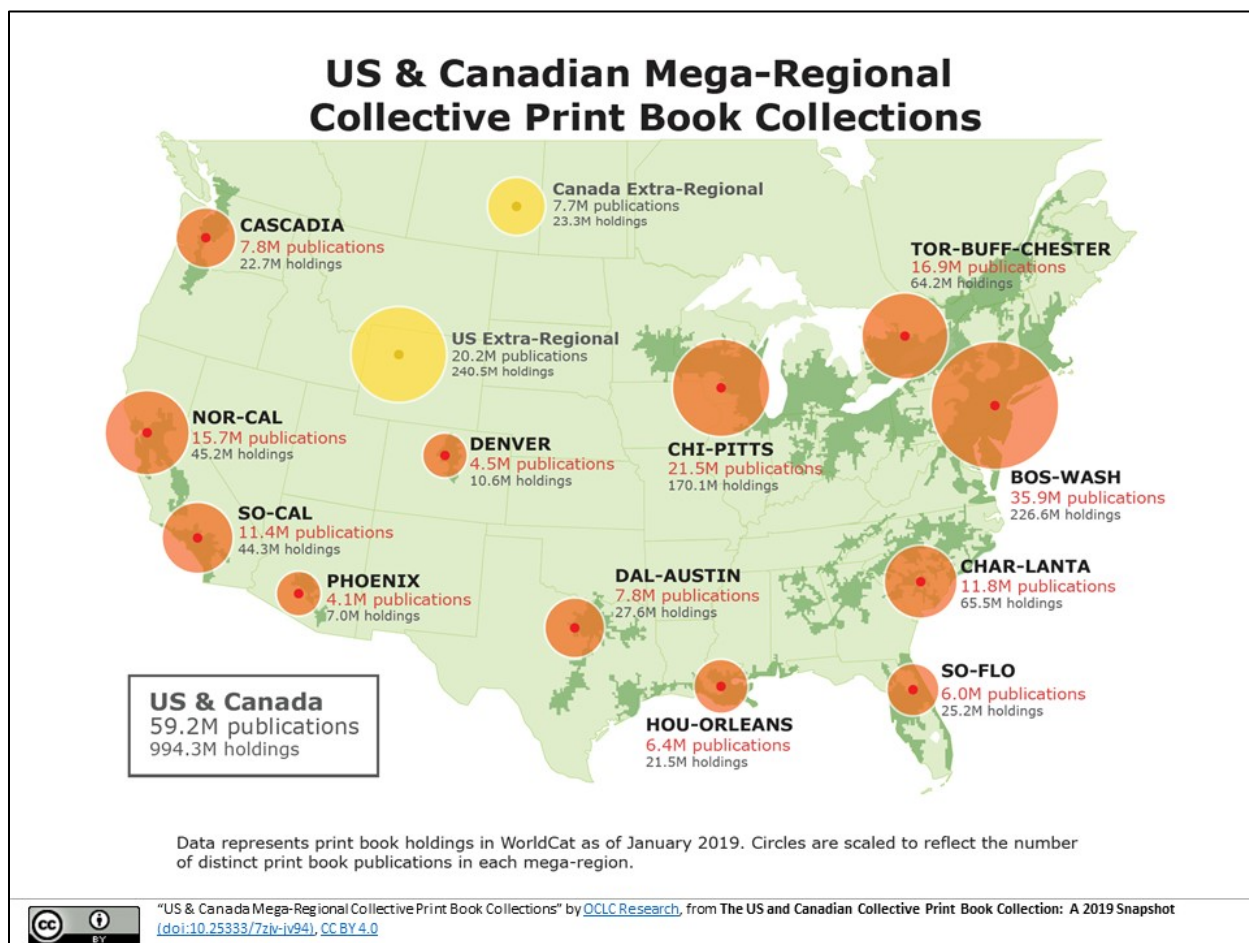
### **1. Collective collections can be circumscribed, quantified, and visualized**

The fact that many—indeed, most—collective collections are notional, existing solely as conceptual constructs in data, makes it natural to treat them as abstractions: useful perhaps as

a shorthand for referencing the aggregated holdings of a group of libraries, but falling short of a well-defined collection with distinct properties.

But to the extent that the data is accurate, comprehensive, and of good quality, one can be quite precise in describing the size and salient characteristics of collective collections—even those aggregated at very large scales. For example, consider this visualization of the US and Canadian collective print book collection and its associated mega-regional collections.<sup>18</sup> Here, the somewhat abstract idea of the aggregate print book holdings of all libraries across the US and Canada is made precise along several dimensions, including size (59.2 million distinct print book publications) and the overall number of print book holdings (994.3 million).

**Figure 2: US and Canadian Collective Print Book Collections**



We can also see the regional distribution of print books within the US and Canada, with the number of distinct print book publications and holdings specified for each of the twelve US and Canadian *mega-regions*—conglomerations of multiple urban centers and their hinterlands, bound together through infrastructure, mutual economic interests, and cultural similarities. For example, the collective print book collection in the BOS-WASH mega-region, which runs from Boston to Washington, DC in the northeastern United States, contains 35.9 distinct print book publications, based on 226.6 million holdings.

As Figure 2 illustrates, the totality of US and Canadian print book holdings can be aggregated into a well-defined collective collection with precise properties, as well as broken up into equally well-defined regional collective collections, also with precise properties.

The US and Canadian collective print book collection illustrates an important takeaway from the collective collections studies: collective collections of any scale can be constructed and analyzed as “real” things, with describable features and characteristics, from which practical intelligence can be derived. In this way, what might seem like a boundless ocean of materials can in fact be viewed as a tractable, well-defined reservoir. With appropriate data in hand, the distributed collections of libraries anywhere can be aggregated into a single, distinct resource—circumscribed, quantified, and even visualized as a collection at scale.

The ability to construct and analyze collective collections is increasingly important as libraries seek opportunities to leverage economies of scale and scope in collection management. Data about collective collections can support conversations about issues of interest to libraries, helping to make concepts and points of discussion more concrete. An early collective collection analysis was a study of the Google 5 digitization project, the precursor to the broader Google Books initiative.<sup>19</sup> The study looked at the collective print collection of the original five participants in the Google book scanning project announced in 2004; the aim was to describe the size and characteristics of the collection that would be produced if all five collections were comprehensively digitized. This work provided an empirical context for the many discussions that occurred around that time concerning the Google Book project specifically, and mass digitization generally, and helped make what had been abstract discussions of the project more precise.

Related to this was a later study of the collective collection of print books subject to US copyright law.<sup>20</sup> The purpose was to estimate the portion of the collective library print book resource that was likely to be encumbered by copyright restrictions, which was and still is an important question in regard to mass digitization and the legally permissible use of digitized surrogates. Later studies of the collective print book collections of two library consortia—the BTAA in the US,<sup>21</sup> and RLUK in the UK<sup>22</sup>—established both the size and salient features of each collection, as well as mapping out patterns of overlap and relative strengths in the print book holdings of the member institutions. These studies informed conversations within these consortia on potential shared print initiatives.

The key point is that collective collections can be constructed and analyzed—at virtually any scale—to inform a wide range of issues of interest to libraries. These collective collections aid in both understanding and visualizing the context or problem space with which they are linked, and supply practical intelligence to support decision-making.

## **2. Rareness is common**

One finding seen repeatedly in our collective collection studies is the paradoxical result that *rareness is common*—in other words, individual library collections are often sufficiently distinctive that when aggregated together, a rich and diverse long tail of scarce or uniquely held materials is built out within the collective collection. This finding underscores the idea that building collective collections may be just as much about identifying and leveraging distinctive local and collective strengths as it is about consolidation and reducing redundancy.

A good example is the finding that three-quarters of the collective print book collection of the membership of the Committee on Institutional Cooperation (CIC; later re-named the Big Ten

Academic Alliance (BTAA)) could be classified as “rare”—i.e., held by three or fewer CIC members.<sup>23</sup> Similarly, the analysis of the collective print book collection of the RLUK membership found that 88 percent of the collection was held by fewer than five libraries.<sup>24</sup> On a larger scale, at least three-quarters of the print book publications in the mega-regional collective print book collections were held by five or fewer institutions in that region. Moreover, 80 percent of the print book publications in the US and Canadian collective print book collection are held in five or fewer regions.<sup>25</sup>

Some of the apparent scarcity detected in these collective collections is undoubtedly an artifact of WorldCat’s development, and therefore a limitation of the data: not all collections are represented in WorldCat, and for those that are, not all of them are represented comprehensively. This will impact assessments of scarcity in analyses using WorldCat data, at times inflating the degree of scarcity of certain materials within a group of libraries. However, the “rareness is common” finding is so pervasive in the studies—witnessed at scales ranging from a few to hundreds of libraries—that it is clearly a persistent feature of collective collections.

The key implication of the rareness property is that it contributes to a stylized picture of a collective collection in which a core set of widely held materials is accompanied by a long tail of relatively scarce materials. Elucidating the features of the aggregated holdings of a group of libraries in this way affords a perspective on both the relative redundancy *and* distinctiveness of the individual collections comprising the collective collection. This is an important piece of intelligence that must be taken into account in a variety of decision-making contexts, especially in regard to shared print policies, in which issues such as the minimum number of copies needed to support the group, the efficient division of collecting responsibilities, and secure management of last copies, are considered.<sup>26</sup>

More specifically, description of the collective collection’s *core* (widely-held) materials creates opportunities for identifying and leveraging the collecting strengths of the group, highlighting shared institutional interests revealed by similarities in collecting patterns, and improving efficiency by reducing unwanted duplication across the collective collection. All of these points fall within the purview of many shared print programs, and in this sense, intelligence gleaned from analyzing collective collections can help advance library interests in this area.

On the other hand, understanding the nature of a collective collection’s *long tail* helps identify the collection strengths of *individual* institutions within the context of the group, and in doing so, creates opportunities for leveraging these local-scale strengths at group scale: for example, by implementing robust resource sharing arrangements that improve the circulation of scarce materials within the group. Again, it is the intelligence extracted from the collective collection that drives these opportunities. Descriptions of the core and the long tail illustrate how understanding the properties of collections at scale open up new possibilities for developing innovative approaches to collection management.

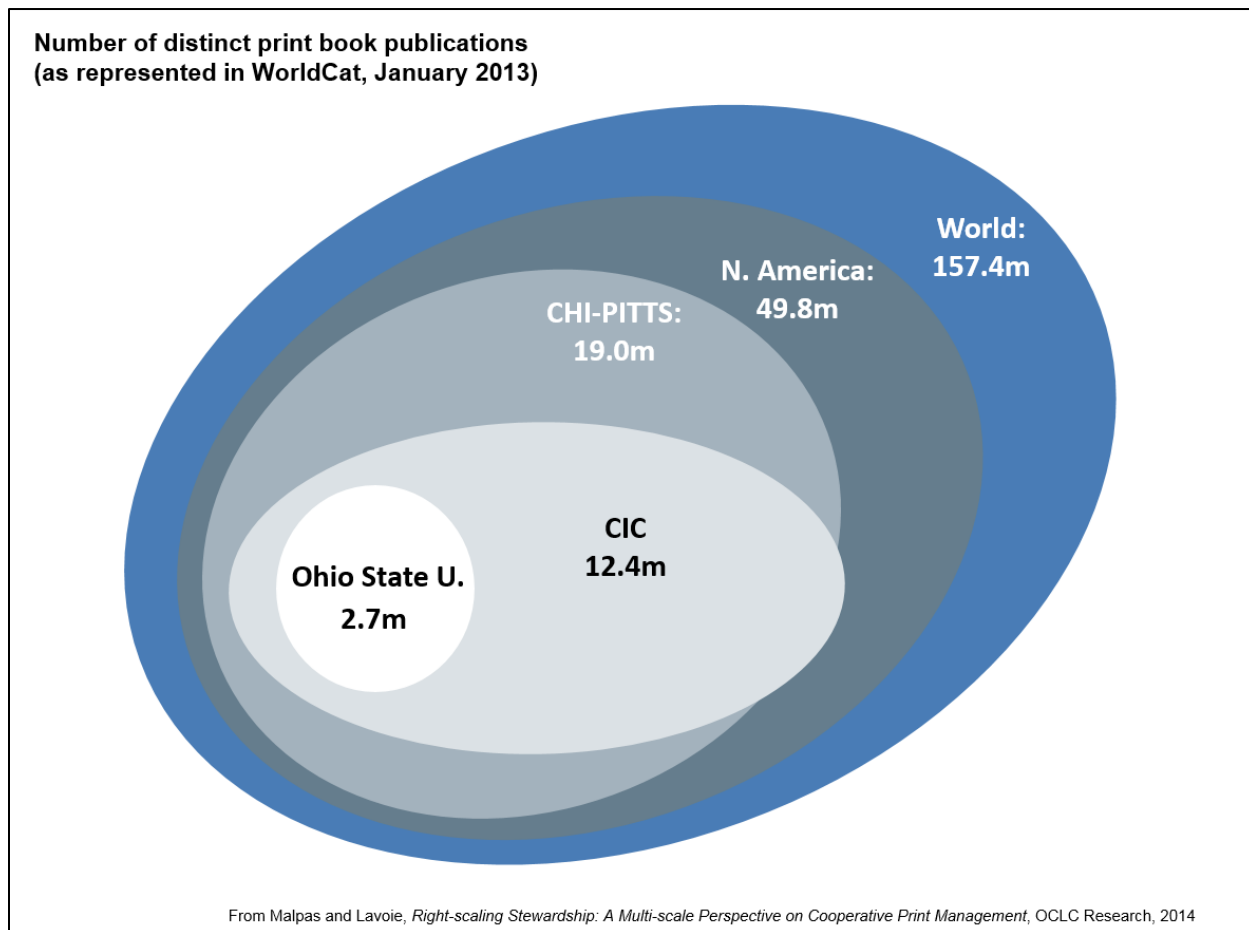
### **3. Coverage requires cooperation**

The “rareness is common” property of collective collections has some important implications for collection management decision-making undertaken with a system-wide perspective. Typically, no single group member, or even a subset of members, can cover the full scope of the collective collection. In other words, if one or more member’s holdings are removed from the collection, the scope of the collection will be diminished—there will be at least some publications that the other members of the group will not be able to duplicate with their own holdings. Therefore, no institution or set of institutions can rely on the rest of the group to fully steward the breadth and

scope of the collective collection; instead, coverage of the collective collection requires the cooperation of the entire group.

This principle appears at a variety of scales. Imagine a series of collective collections, each encompassing a larger and larger group of members. In this case, one can ask whether a group of libraries can cover the collective holdings of a larger group of libraries in which the first group is embedded. By way of illustration, consider this figure from the study of the CIC (now BTAA) collective print book collection.<sup>27</sup>

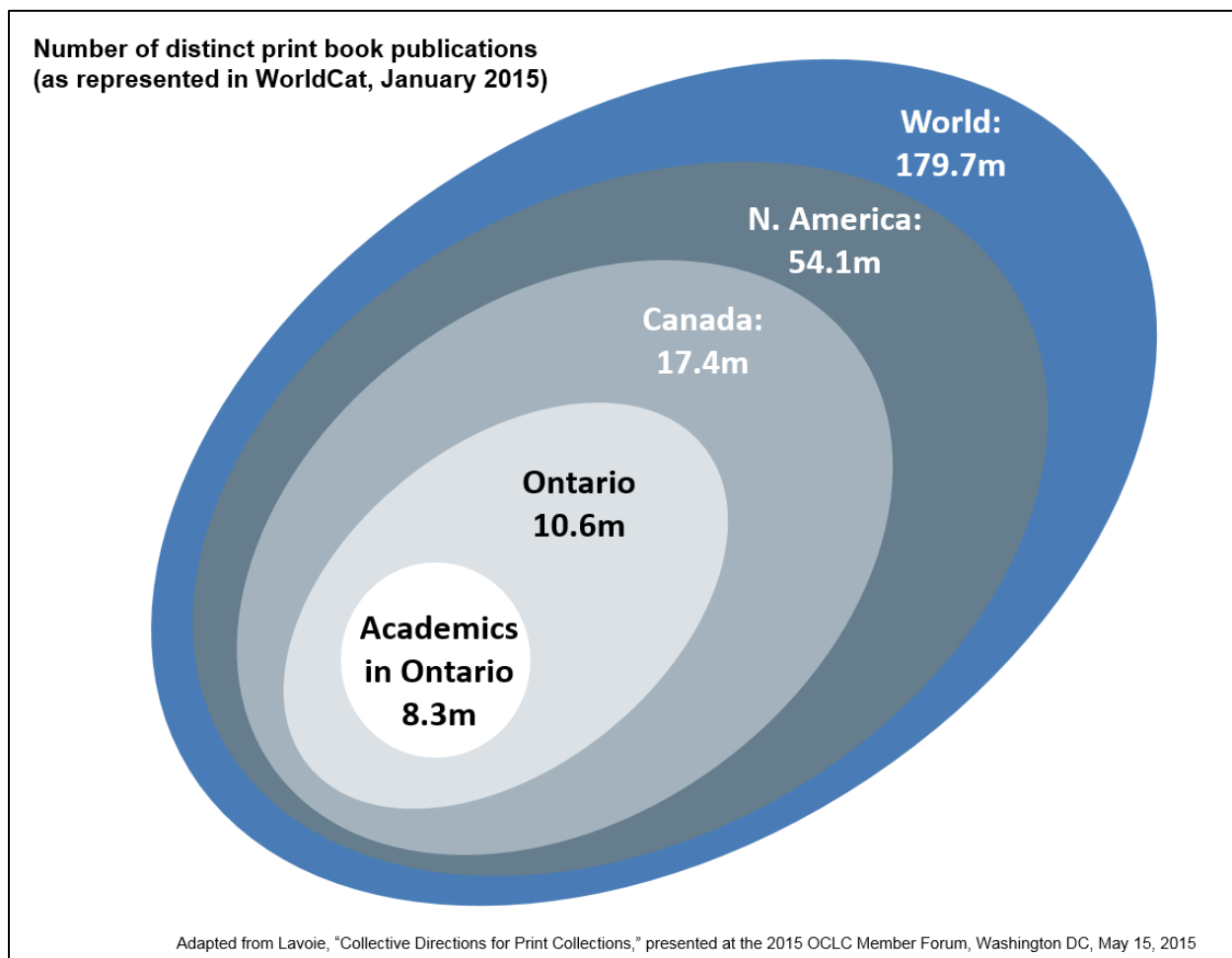
**Figure 3: Scaling the Collective Print Book Collection: A CIC (now BTAA) Perspective**



As the collections move from local to global scale, the size of the print book resource rapidly increases with the scale of aggregation – and in doing so, illustrates that coverage requires cooperation. As the picture shows, The Ohio State University can only cover a fraction of the materials in the CIC collective print book collection.<sup>28</sup> CIC covers only part of the CHI-PITTS regional collective print book resource (CHI-PITTS extends from roughly Chicago to Pittsburgh, and is the mega-region in which most of the CIC membership are located<sup>29</sup>), which in turn covers only part of the North American print book resource.<sup>30</sup> Finally, the North American collective print book collection covers only part of the global collective print book collection.

The same pattern is evident in another context. Consider the Canadian print book resource<sup>31</sup>:

**Figure 4: Scaling the Collective Print Book Collection: A Canadian Perspective**



Beginning in the center, the figure shows the collective print book collection of all the academic libraries in the province of Ontario. Next, the scale increases to the entire province of Ontario, including the academic libraries as well as public libraries, school libraries, and other collecting organizations. This is followed by the collective print book collection of all of Canada, and then of North America<sup>32</sup> and the world. As in the previous figure, it is evident that smaller groupings of libraries cannot provide full coverage of the holdings of larger groupings. For example, the provincial print book resource is not approximated by the collective collection of Ontario's academic libraries.

The significance of this finding speaks to the element of uniqueness found in the contributions of local collections to the collective collection. Recall that the size of a collective collection grows only when publications are added that were not already present in the collection – duplicate holdings are not counted. So the fact that the collective collection of all Ontario libraries is larger than that limited to Ontario academic libraries means that non-academic libraries hold many publications that are not duplicated in the collections of the academic libraries. Similarly, the CIC

consortium can only account for a portion of the print books available in the CHI-PITTS region. The implication is that if Ontario's provincial – or CHI-PITT's regional – print book resource is to be preserved, the cooperation of libraries beyond the academic cohort accounting for the largest local collections will be needed.

These findings suggest that the scale of cooperation must grow as the scale of the collective collection grows. Additionally, the idea that coverage requires cooperation amplifies the potential importance of looking beyond legacy associations in organizing cooperative endeavors. For example, both the CIC and Canadian data suggest that the goal of comprehensive stewardship of the print book resource at regional or provincial scale will not be fully achieved by a single consortium of geographically proximate academic institutions. Instead, cooperation would be needed from other institutions in the region or province, such as public libraries and other kinds of cultural heritage institutions. With new forms of collective collections comes the opportunity for new groupings of cooperating institutions.

An example of the “coverage requires cooperation” principle in action is the Eastern Academic Scholarship Trust, or EAST.<sup>33</sup> EAST is a membership-based shared print initiative involving the collective print book collection of 60 academic and research libraries. Each of these members must commit to retain certain monographs in their collections, in particular “those titles that the library holds that are unique to the EAST collective collection as well as additional copies of titles identified as those frequently used.” The idea is that new members “have significant or unique monograph collections, which can supplement the existing retention commitments made by EAST members.”<sup>34</sup> Therefore, as the scale of the EAST collective collection grows through the addition of the collections of new members, the scale of cooperation grows as well, as each new member commits to retaining the portions of its collection that are unique relative to the holdings of other EAST members. EAST has constructed a membership model around these principles, and in 2018, become self-supporting through membership dues.

#### **4. The details are in the scale**

The report on the collective collection of the CIC membership concludes with the following observation:

If there is one principle that warrants special emphasis, it is that scale impacts nearly all the fundamental characteristics of a collective print resource and the cooperation needed to sustain it. As our findings indicate, scale shapes the scope and depth of the collective print resource; the degrees of redundancy and distinctiveness attached to that collective resource at both local and consortial level; and the scope of cooperation needed to achieve reasonable thresholds of coverage and access. In this sense, “right-scaling” stewardship of the collective print investment becomes the central question of any shared print strategy ...<sup>35</sup>

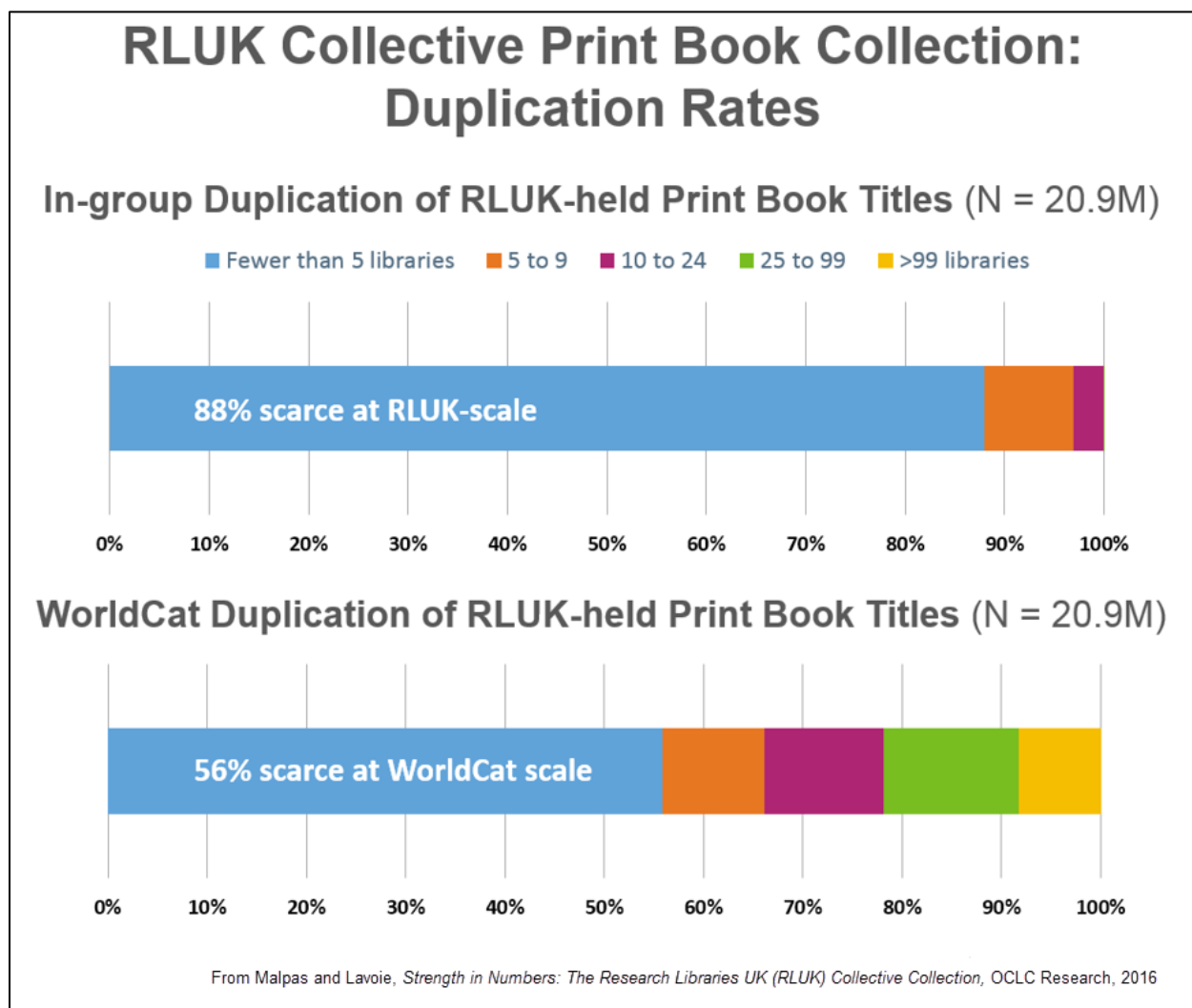
One way that scale matters is that it drives the scope and depth of the collective collection. Again, this is an implication of the “rareness is common” theme, where the long tail keeps getting longer as more and more members contribute their holdings to a collective collection, and in doing so, add more and more materials that were not previously present in the collection. The collective collection expands in scope, in the sense that the range of subjects, formats, material types, and so on expands. But the collective collection can also expand in depth, as more materials are added *within* each of these categories. In this sense, scaling up carries with it the potential to fuel both horizontal (scope) and vertical (depth) growth in the collective collection.

In a similar way, scale drives the global diversity of the collection. Larger collective collections tend to be more diverse in terms of the country of origin of published materials, as well as the language of content. For example, a 2012 study of the collective print book collections of the US and Canadian mega-regions found that while all of the regional collective collections were quite diverse in terms of country of origin and language of content, the largest regional collections exhibited the highest proportions of these materials: more than sixty percent of the books in the largest regional collection (BOS-WASH), for example, were published outside the US and Canada, and nearly half were in languages other than English.<sup>36</sup>

An important corollary to the idea that collective collections build out a long tail of relatively rare (not widely held) and diverse materials is that the degree of uniqueness associated with a collective collection depends heavily on the scope of comparison. This is seen in the study of the collective collection of the RLUK membership.<sup>37</sup> In the figure below, the top graph divides the print books in the RLUK collective collection into segments according to how widely held they are across the group. The data indicates that nearly 90 percent of the print book publications in this collection are held by fewer than five libraries in the group (the blue bar). The bottom graph also divides the print books in the RLUK membership's collective collection into segments according to how widely held they are, except this time the frame of reference is at global scale, represented by WorldCat. Here only 56 percent of the RLUK membership's collective collection is held by fewer than five libraries in the world. In general, materials that appear scarce at one scale can be widely held at larger scales.



Figure 5: RLUK Collective Print Book Collection: Duplication Rates



Similar results were found with the collective collection of Ontario academic libraries: more than 80 percent of the materials in the collection were held by only a few institutions in the group; in contrast, only 20 percent of the collective collection was similarly scarce in the context of WorldCat.<sup>38</sup>

Scale impacts many key aspects of the collective collection, including the length of the long tail, the overall diversity of the content in terms of country and language, and the redundancy as well as the distinctive strengths associated with the collective collection. In this sense, scale is an important variable in organizing cooperative arrangements around activities such as shared print, collection development, digitization, and resource sharing.

## 5. Collective collections enable a distant reading of the published record

As collectors and stewards of the world's published output, libraries are uniquely positioned to provide insight into broad patterns and trends emerging from world literature, scholarship, and other forms of creative expression. This is because the global collective collection—the

collective collection at its highest scale, encompassing library collections everywhere—is the best approximation available of the published record, or society’s cumulative published output.

Digital humanities scholars such as Franco Moretti have developed a keen interest in “distant reading”<sup>39</sup>—analyzing large aggregations of digitized text for patterns and insights that would be obscured by the more traditional practice of “close reading” (reading books, or passages from books, one at a time). Similarly, the analysis of large aggregations of library data—i.e., descriptions of the publications in library collections—can serve as another form of distant reading, allowing an examination at scale of the properties of large swathes of publications.

An example of how library data, and the concept of collective collections, can be employed in service of humanities research is the series of “national presence” reports that OCLC Research has published over the last few years. These reports—focusing on Scotland,<sup>40</sup> New Zealand,<sup>41</sup> Ireland,<sup>42</sup> and Canada,<sup>43</sup> respectively—have sought to identify and explore the national presence of a given country in the published record: in other words, the collective collection consisting of all materials published *in* that country, *by* the people of that country, and/or *about* that country.

For example, one report identified the Irish presence in the published record, revealing a collection of approximately 900,000 distinct works. Using library holdings as a metric of popularity, the report found that *Gulliver’s Travels* was the most popular work by an Irish author, and that Jonathan Swift was the most popular Irish author. The report also tracked changing interests in particular Irish authors over time, and described the global diffusion of the Irish presence in the published record as it is manifested in library collections around the world.<sup>44</sup>

The key point is that this type of humanities research begins with a collective collection—the aggregated holdings of all libraries everywhere. The collective collection is constructed using WorldCat, which registers the collections of thousands of libraries world-wide. This global collective collection approximates the published record, in the sense that the collecting activity of libraries everywhere has brought most of the world’s published output into library collections.<sup>45</sup> From this collective collection-inspired view of the published record, the subset of interest—for example, the Irish contribution—can be carved out for analysis.

The global collective collection’s approximation of the published record opens up fascinating opportunities for new frontiers in humanities work, using data about library collections to explore trends and patterns in cultural, literary, and intellectual development, as shaped by contingent historical and/or political circumstances.

## **Conclusion**

Looking back over OCLC Research’s body of work on collective collections, it is evident that collective collections generate value to libraries in three ways.

First, collective collections aid local decision-making by making it “system-aware”—where the system can be a group, a consortium, a region, or even all libraries everywhere. Knowledge about the collective collection helps libraries orient their local collection management decisions—such as acquisitions, retention, and de-accessioning—within a broader context. In this sense, the rising importance of collective collections illuminates a shift in the strategy of managing collections, in which local collections are seen not just as assemblies of materials for local use, but also as pieces of a larger system-wide resource.

Second, collective collections help libraries cooperate with one another in mutually beneficial ways. Understanding the scope and depth of the collective collection helps groups of libraries identify individual strengths and group redundancies. It opens pathways toward making the ever-present long tail of the collective collection more visible and more accessible. And it helps libraries manage down their legacy print investments in ways that can potentially release resources for other uses, while at the same time securing the ongoing availability of this important corpus of materials.

Finally, collective collections help libraries project an aggregated presence into non-library spaces, generating a critical mass of resources that exceeds the visibility any single collection might obtain. For example, the national presence studies illustrate an opportunity to project the global collective collection—in the form of aggregated library data—into the humanities research space, where interesting questions can be asked and answered through a distant reading of world literature. More generally, data about collective collections helps consolidate a fragmented library presence and project it into non-library domains.

The value of collective collections is inextricably linked to data comprehensiveness and quality. If a collection is not fully registered in places where it can be aggregated with data about other collections, then portions (if not all) of that collection will be for all intents and purposes invisible in places where value-creating activity built around collective collections is taking place.

Strategies for developing, managing, and disclosing library collections are experiencing fundamental shifts, moving from a perspective that is largely local and autonomous to one that is system-aware and cooperative. The concept of collective collections is a natural outgrowth of this new approach to collection management.

Collective collections have progressed to the point where the concept is now quite familiar, with many examples to point to, and a growing awareness of the wide range of library interests in which collective collections could play an important role. As the concept of collective collections continues to evolve, the next frontier may be the development of general strategies for *operationalizing* the collective collection: in other words, framing collective collections as shared resources, supported by shared services, residing in shared stewardship infrastructures, and managed within a robust set of cooperative arrangements. In this way, collective collections may become the fundamental unit of interest in a new set of network-enabled approaches to collection development and management.

## Notes

<sup>1</sup> Shared print is an area where operationalizing collective collections is especially important. For example, OCLC's Sustainable Collection Services offers data and tools to help libraries visualize and understand their local collection in the context of a broader group-scale collective collection (See OCLC's "Sustainable Collection Services" page, <https://www.oclc.org/en/sustainable-collections.html>).

<sup>2</sup> For a detailed discussion of conscious coordination, see Brian Lavoie and Constance Malpas, *Stewardship of the Evolving Scholarly Record: From the Invisible Hand to Conscious Coordination, 2015*, (Dublin, Ohio: OCLC Research), <https://doi.org/10.25333/C3J63N>.

<sup>3</sup> A compendium of select reports from OCLC Research's collective collection work was published in 2013. *Understanding the Collective Collection: Towards a System-wide Perspective on Library Print Collections* went on to receive a 2014 Association for Library Collections & Technical Services (ALCTS) Presidential Citation. The compendium is available at: <https://doi.org/10.25333/C3GP8R>.

<sup>4</sup> "Lorcan Dempsey", Wikipedia, accessed August 3, 2018, [https://en.wikipedia.org/wiki/Lorcan\\_Dempsey](https://en.wikipedia.org/wiki/Lorcan_Dempsey)

<sup>5</sup> For example, at the 2018 Electronic Resources and Libraries conference, one finds the session "Managing the Collective Collection: Tools and Strategies for Collaborative Collection Management," <https://erl18.sched.com/event/CrgW/s038-managing-the-collective-collection-tools-and-strategies-for-collaborative-collection-management>. Similarly, the 2017 Big Ten Academic Alliance Library Conference aimed to explore "how member research libraries can move from legacy CCD activities to a more holistic environment that leverages robust discovery, digitization, delivery, and shared service environments to advance and shape the collective collection," <https://www.btaa.org/about/calendar/conferences/library/2017/home>.

<sup>6</sup> Nick Ismail, "UK Public Libraries to Become a 'Single Digital Presence'?" *Information Age*, August 30 2017, <http://www.information-age.com/uk-public-libraries-single-digital-presence-123468260/>.

<sup>7</sup> the website for HathiTrust, <https://www.hathitrust.org/>.

<sup>8</sup> "Shared Print Repository – Menu," Big Ten Academic Alliance, accessed July 24, 2018, <https://www.btaa.org/library/shared-print-repository/introduction>.

<sup>9</sup> the website for the Five Colleges, "Five College Consortium: Home" page, <https://www.fivecolleges.edu/>.

<sup>10</sup> Five Colleges; Libraries Catalog; "Basic Search of All Five Colleges," <https://fcaw.library.umass.edu/>.

<sup>11</sup> the website for the UK Research Reserve, "Introduction," <http://www.ukrr.ac.uk/>.

<sup>12</sup> OCLC; OCLC's "WorldCat" page, <https://www.oclc.org/en/worldcat.html>.

<sup>13</sup> See, for example. Lizanne Payne, "Winning the Space Race: Expanding Collections and Services With Shared Depositories," *American Libraries* (September 23, 2014), <https://americanlibrariesmagazine.org/2014/09/23/winning-the-space-race/>

<sup>14</sup> It should be noted that while the potential for achieving efficiencies through collective approaches to print management has been broadly asserted, the benefits gained through strategies such as space re-allocation will depend on local circumstances.

- <sup>15</sup> Brian Lavoie, Constance Malpas and JD Shipengrover. *Print Management at "Mega-scale": A Regional Perspective on Print Book Collections in North America*, (Dublin, Ohio: OCLC Research), 2012, <https://doi.org/10.25333/C3133Z>.
- <sup>16</sup> Lorcan Dempsey, Constance Malpas, and Mark Sandler, *Operationalizing the BIG Collective Collection: A Case Study of Consolidation vs Autonomy*, (Dublin, Ohio: OCLC Research), 2019, <https://doi.org/10.25333/jbz3-jy57> .
- <sup>17</sup> Karla Strieb, "Collaboration: The Master Key to Unlocking Twenty-First Century Library Collections," in *Shared Collections*, ed. Dawn Hale, (Chicago: American Library Association, 2016), 5.
- <sup>18</sup> Brian Lavoie, *The US and Canadian Collective Print Book Collection: A 2019 Snapshot*, (Dublin, Ohio: OCLC Research), 2019, <https://doi.org/10.25333/7zjv-jv94>.
- <sup>19</sup> Brian F. Lavoie, Lynn Silipigni Connaway and Lorcan Dempsey, "Anatomy of Aggregate Collections: The Example of Google Print for Libraries," *D-Lib Magazine* 11 no. 9 (September 2005), <http://www.dlib.org/dlib/september05/lavoie/09lavoie.html>.
- <sup>20</sup> Brian Lavoie and Lorcan Dempsey, "Beyond 1923: Characteristics of Potentially In-copyright Print Books in Library Collections," *D-Lib Magazine* 15, no. 11/12 (November/December 2009), <http://www.dlib.org/dlib/november09/lavoie/11lavoie.html>.
- <sup>21</sup> Constance Malpas and Brian Lavoie, *Right-scaling Stewardship: A Multi-scale Perspective on Cooperative Print Management*, (Dublin, Ohio: OCLC Research), 2014, <https://doi.org/10.25333/C33059>.
- <sup>22</sup> Constance Malpas and Brian Lavoie, *Strength in Numbers: The Research Libraries UK (RLUK) Collective Collection*, (Dublin, Ohio: OCLC Research), 2016, <https://doi.org/10.25333/C3N33J>.
- <sup>23</sup> Malpas, *Right-scaling Stewardship*.
- <sup>24</sup> Malpas, *Strength in Numbers*.
- <sup>25</sup> Lavoie, *Print Management at "Mega-scale"*.
- <sup>26</sup> For more on long tails and library collections, see Lorcan Dempsey, "Libraries and the Long Tail: Some Thoughts about Libraries in a Network Age," *D-Lib Magazine* 12 no. 4 (April 2006), <http://www.dlib.org/dlib/april06/dempsey/04dempsey.html>.
- <sup>27</sup> Malpas, *Right-scaling Stewardship*.
- <sup>28</sup> Although the focus here is on Ohio State's BTAA membership, that is not to say that this is the only membership-based collective collection in which Ohio State participates. For example, Ohio State also is a member of the state-wide OhioLINK consortium, whose collective holdings can also be viewed as a collective collection.
- <sup>29</sup> Note that the CIC collection is not fully nested in the CHI-PITTS regional collection; this is because not all CIC members are located in the CHI-PITTS region.
- <sup>30</sup> For the purposes of this study, our view of North American print book holdings was limited to the US and Canada.
- <sup>31</sup> OCLC Research, unpublished data.
- <sup>32</sup> Again, our view of North American print book holdings was limited to the US and Canada.
- <sup>33</sup> the website for EAST, <https://eastlibraries.org/>.

<sup>34</sup> the website for EAST, “Criteria for Joining EAST” page, [https://eastlibraries.org/sites/default/files/BLC\\_Uploads/Criteria%20for%20Joining%20EAST%20Final%203.29.19.pdf](https://eastlibraries.org/sites/default/files/BLC_Uploads/Criteria%20for%20Joining%20EAST%20Final%203.29.19.pdf).

<sup>35</sup> Malpas, *Right-scaling Stewardship*, 52 (see n. 18).

<sup>36</sup> Lavoie, *Print Management at “Mega-scale”*.

<sup>37</sup> Malpas, *Strength in Numbers*.

<sup>38</sup> OCLC Research, unpublished data.

<sup>39</sup> Franco Moretti, *Distant Reading*, (London: Verso), 2013

<sup>40</sup> Brian Lavoie, *Not Scotch, but Rum: The Scope and Diffusion of the Scottish Presence in the Published Record*, (Dublin, Ohio: OCLC Research), 2013, <https://doi.org/10.25333/C3SH0S>.

<sup>41</sup> Lavoie, Brian. *Kiwis in the Collection: The New Zealand Presence in the Published Record*, (Dublin, Ohio: OCLC Research), 2014, <https://doi.org/10.25333/C3QP8X>.

<sup>42</sup> Brian Lavoie and Lorcan Dempsey, *An Exploration of the Irish Presence in the Published Record*, (Dublin, Ohio: OCLC Research), 2018, <https://doi.org/10.25333/C3WS6R>.

<sup>43</sup> Brian Lavoie, *Maple Leaves: Discovering Canada through the Published Record*, (Dublin, Ohio: OCLC Research), 2019, <https://doi.org/10.25333/ek4v-ag09>.

<sup>44</sup> Lavoie, *An Exploration of the Irish Presence in the Published Record*.

<sup>45</sup> This assertion must be qualified by acknowledging that some historical and contemporary cultures may not be well-represented in the types of published materials that libraries collect, or even if they are, that they might still be under-represented in library collections. See the discussion in Section II of some of WorldCat’s limitations in terms of coverage and comprehensiveness.