

Strength in Numbers: The Research Libraries UK (RLUK) Collective Collection

Constance Malpas
Research Scientist

Brian Lavoie
Research Scientist



© 2016 OCLC Online Computer Library Center, Inc.

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

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



June 2016

OCLC Research

Dublin, Ohio 43017 USA

www.oclc.org

ISBN: 978-1-55653-515-4

OCLC Control Number: 950534748

Please direct correspondence to:

Constance Malpas

Research Scientist

malpasc@oclc.org

Suggested citation:

Malpas, Constance, and Brian Lavoie. 2016. *Strength in Numbers: The Research Libraries UK (RLUK) Collective Collection*. Dublin, Ohio: OCLC Research.

<http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-strength-in-numbers-rluk-collective-collection-2016-a4.pdf>.

Cover image:

Map depicting the approximate locations of the 37 institutions comprising the RLUK membership. The ten RLUK members located in the London area are represented by a single large dot.

FOREWORD

Researchers and students all over the world benefit from the collections held by RLUK member libraries. These collections have grown over centuries: through purchases and donations, strategic subject building and happy accident, through collaboration and serendipity, all supported by a robust legal deposit system. We have exceptional individual collections, but we are becoming increasingly interested in the total collection within the UK. We are asking strategic questions about the preservation and storage of print books, the best use of library space, the range and effectiveness of digital surrogacy, and the nature of collecting.

Our colleagues at OCLC have harnessed the unique possibilities offered by WorldCat to provide us with a window into what they have coined the “collective collection” across RLUK institutions. Their discoveries, detailed in the report, both astound and defy expectations. The sheer breadth and scope of our collections are remarkable—we hold print books in almost every language and from every country conceivable, over the entire history of printing. What defies expectation is that there appears to be less duplication of content amongst RLUK members than might be supposed. This raises important questions when thinking about preservation and collection management, as well as highlighting the potential importance of further international collaboration. Another area for international collaboration is suggested by the analysis of the proportion of the print collection for which digital surrogates are available—and fully usable—through collections such as HathiTrust. It is clear that we have a lot more to do here.

We are very grateful to Constance Malpas and Brian Lavoie for their research expertise and the rigour that they have brought to this fascinating report. We are also grateful to our RLUK colleagues on the Advisory Group, and to many at both OCLC and RLUK member institutions who worked hard to ensure that our libraries were as fully represented within WorldCat as possible. The analysis in this report confirms that the robustness of the underlying bibliographic data is key. Good data will provide the foundation on which we can make good decisions in continued support of the researchers and students we serve.

David Prosser, PhD
Executive Director, RLUK

John MacColl, FRSE
Chair, RLUK
University Librarian & Director of Library Services, University of St Andrews

ACKNOWLEDGMENTS

The research described in this report would not have been possible without the support and collaboration of colleagues within OCLC, RLUK and the 11 libraries who served on the project advisory group. Advisory group members engaged in months of discussion that helped to shape and refine the analysis, participated in detailed reviews of data samples and preliminary findings, and provided thoughtful commentary on an early version of this report. We extend special thanks to the following individuals for providing valuable feedback on an earlier draft of this report: Janet Aucock (University of St Andrews), Sandra Bracegirdle (The University of Manchester), Suzy Cheeke (University of Bristol), Helen Faulds (University of St Andrews), John MacColl (University of St Andrews), Alasdair MacDonald (University of Edinburgh), Hannah Mateer (University of Edinburgh), Victoria Parkinson (Kings College London), David Prosser (RLUK) and Laura Shanahan (University of Edinburgh). Several OCLC colleagues were instrumental in producing the published version of the report and we are glad to acknowledge them here. Erin Schadt provided excellent editorial support; Jeanette McNicol wrangled many versions of the document into its final form; and JD Shipengrover refined and improved the graphics. We are also grateful to Paul Shackleton for his tireless work with RLUK libraries to improve coverage of UK collections in WorldCat, and to Andrew Hall for his advice and support over the course of the project. Finally, we wish to credit Lorcan Dempsey for his inspirational leadership in establishing and nurturing the program of research of which this work is a part.

CONTENTS

Introduction.....	8
Data Sources	9
The RLUK Collective Collection	10
Size and Scope	10
Places and Languages of Publication	12
Material Types	13
Print Books	14
Internal Overlap.....	20
Duplication in RLUK Library Holdings	21
Duplication of Intellectual Content: Works and Versions	22
Shared Strengths: Subject Coverage and Communities of Interest within RLUK.....	25
Duplication in RLUK Print Book Holdings.....	27
External Overlap	29
Overlap with the ARL Print Book Resource	30
Overlap with HathiTrust.....	32
Regional Analysis.....	33
Concluding Observations	36
Scale adds Scope and Depth.....	36
Coverage Requires Cooperation.....	37
Scarcity is Relative	37
Areas for Further Exploration	39
Appendices	41
Appendix 1: RLUK Libraries and WorldCat Symbols	42
Appendix 2: RLUK Advisory Group Members	44
Appendix 3: Editions Most Widely Held in RLUK Collective Collection	45
Notes	48

FIGURES

Figure 1. RLUK Collective Collection: Size (January 2016).....	11
Figure 2. Material Types in the RLUK Collective Collection (January 2016)	14
Figure 3. RLUK Collective Print Book Collection: Size (January 2016)	15
Figure 4. RLUK Collective Print Book Collection: Duplication Rates	16
Figure 5. RLUK Collective Print Book Collection: Language of Content (% of Publications)	17
Figure 6. RLUK and ARL Collective Print Book Collections: Age Distribution, by Decade of Publication (% of Publications).....	19
Figure 7. In-group Duplication of Titles in RLUK Collective Collection (January 2016)	21
Figure 8. WorldCat Duplication of Titles in RLUK Collective Collection (January 2016)	22
Figure 9. In-group Duplication of Print Book Titles in RLUK Collective Collection (January 2016) ...	27
Figure 10. WorldCat Duplication of Print Book Titles in RLUK Collective Collection (January 2016)..	28
Figure 11. Overlap between ARL and RLUK Print Book Collections (January 2016)	30
Figure 12. Overlap between Core and Rare Segments of ARL and RLUK Print Book Collections (January 2016)	31
Figure 13. Duplication of RLUK Print Book Titles in HathiTrust Digital Library (January 2016)	33

TABLES

Table 1. Top Ten Places of Publication in RLUK Collective Collection	12
Table 2. Top Ten Languages in RLUK Collective Collection.....	13
Table 3. RLUK Collective Print Book Collection: Top Ten Countries of Publication	18
Table 4. Top Ten Most Frequently Occurring FAST Headings in RLUK Collective Print Book Collection	20
Table 5. Works with the Largest Number of Editions (formats, versions) in RLUK libraries.	23
Table 6. Editions most Widely-held in RLUK Collective Collection (January 2016)	24
Table 7. Shared Subject Strengths in RLUK Collective Collection (February 2016).....	26
Table 8. Rare and Core Segments of the RLUK and ARL Collective Print Book Collections	29
Table 9. RLUK Collective Print Book Collection: Key Overlaps with ARL.....	32
Table 10. RLUK Collective Print Book Collection: England and Scotland Segments	35

INTRODUCTION

In a recent paper on the future of library collections, Dempsey, et al., (2014) emphasize the importance of scale in approaching collection management in a network environment, noting that libraries are “moving ‘above the institution’ and becoming increasingly embedded in networks of collaboration, cooperation and consolidation that are fundamentally changing the ways in which collections and related infrastructure are developed, managed and made accessible.”¹

The focus on scale has led to a growing interest in the capacities, infrastructure and resources of the *system*, however the system is defined (e.g., small group, consortium, region). This is especially evident in regard to collections: the boundaries of the local collection are becoming more fluid, and where once library collections were viewed as autonomous assemblies of material for local use, they are now often seen as components of a broader, system-wide library resource.

Decision-making across a wide range of library strategic interests, such as discovery, digitization, digital curation, managing down print collections or curating the evolving scholarly record, increasingly requires a thorough awareness of the surrounding network environment—whether in the context of multiple institutions acting collectively or a single institution taking local decisions in light of a group, regional or national setting. This, in turn, has increased the need for libraries to understand collections at scales beyond the local level in order to address areas of need both locally and within multi-institutional cooperative arrangements. In short, libraries require an evidence base from which to derive intelligence not only about their own collections, but also the *collective collection* of the libraries relevant to a particular decision-making and/or cooperative context.

OCLC Research introduced the concept of collective collections several years ago as an analytical construct for thinking about collections at scales above the institution. A collective collection—the combined collections of multiple institutions, viewed as a single, aggregate resource—supplies an invaluable perspective in situations where:

- activities and services extend across local collection boundaries;
- there is a need to gather and expose the aggregate library resource in the broader network environment;
- opportunity exists to optimize the supply and demand for library materials on a system-wide basis.

In general, collective collections are a key part of an emerging library environment that favors collaboration and coordination and where libraries seek to create value through collective action and shared capacities.

In light of the growing importance of collective collections, OCLC Research has undertaken a substantial amount of work aimed at sharpening the concept of a collective collection and charting the development of collective collections in a variety of contexts.² As a continuation of this strand of work, we are pleased to have had the opportunity to collaborate with Research Libraries UK (RLUK), a consortium of leading research libraries in the UK and Ireland, to characterize the general contours and specific features of the RLUK collective collection, both to understand its size, scope and cross-institutional overlap patterns, and to support deeper collaboration around collections management within the RLUK membership.

In this report, we paint—with admittedly broad brushstrokes—a portrait of the RLUK collective collection that represents it as a distinctive, cohesive aggregate resource, rather than as an assemblage of disparate institutional collections.³ The primary purpose of the analysis is to highlight some important aspects of this collective collection, with a special emphasis on those pertaining to print books. Cooperative management of the collective print book investment is an issue of keen interest to the RLUK membership and to UK higher education generally.⁴ The goal is also to provide a multi-scalar view of the RLUK collective collection, placing it in the context of institutional, group and global perspectives. We supplement the analysis by drawing in findings and observations from collective collections work we have done in other contexts, and identifying points of convergence and divergence with the RLUK collective collection. Finally, we frequently use the collective collection of the North American-based Association of Research Libraries (ARL) as a comparator for the RLUK collection, in order to provide additional perspective from a consortium of peer institutions vis-à-vis the RLUK membership.

It is important to emphasize that our characterization of the RLUK collective collection is only a sketch of the salient features of this resource. We cannot claim that it does justice to the rich nuance, diversity and depth present within the collections of the RLUK membership. Moreover, our view of the RLUK collective collection is not a complete one: the collections of some RLUK members are not included in the analysis, and for those collections that are included, not all of them are fully registered in WorldCat, the bibliographic database from which we constructed our characterization of the RLUK collective collection. Details on the scope and limitations of our data are found in the next section. Despite these limitations, however, we believe our portrait is sufficiently complete to provide an informative foundation for policy-making around issues of mutual interest within the RLUK membership.

As research libraries re-cast their collections for the 21st century networked environment, they will benefit from the ability to consider those collections at a system-wide level, where system can be defined at a variety of scales. Through the analysis of collective collections, supported by the unique global coverage of WorldCat, these system-wide views can be manifested in the form of finite, aggregate resources with properties that can be understood and acted upon in the context of a wide range of important library strategic interests. It is our hope that the analysis in this report will demonstrate the utility of collective collections in thinking about opportunities for furthering collaboration within the RLUK membership, as well as illuminate the key features of an important aggregate resource within the broader library system.

Data Sources

This study uses data from WorldCat®, which is a set of databases that includes 360 million bibliographic records and 2.4 billion library holdings, representing the most comprehensive global network of data about library collections and services. The analysis reflects the collections of RLUK members (and other libraries as noted) as they were registered in WorldCat in January 2016.

The RLUK collective collection was constructed by aggregating the holdings in WorldCat of all RLUK members, then removing duplicate holdings to yield the set of distinct publications that are held across the RLUK membership.

Our representation of the RLUK collective collection is not a complete one. The analysis, by necessity, is based on RLUK member collections *as they are represented in WorldCat*. Not all RLUK collections are fully up-to-date in terms of registration in WorldCat. Institutions included in the analysis were deemed to have a sufficient level of registration such that useful inferences could be drawn from their current representation in WorldCat. The collections of two RLUK members—the National Library of Wales and Queen’s University Belfast—were not included in the analysis because current library holdings are not represented in WorldCat. In addition, three RLUK members—University of Leicester,

University of Reading and Royal Holloway—joined RLUK after the project had commenced and were not included in the analysis. The remaining 32 RLUK members are included in the study.⁵ A list of RLUK institutions and WorldCat symbols included in our analysis is provided in appendix 1.

An advisory committee comprised of representatives from the RLUK executive and 11 RLUK institutions whose collections are included in the analysis was formed to provide guidance and consultation during the study. The advisory committee was an invaluable resource for improving our understanding of the RLUK consortium context against which the study was performed, as well as assisting in the refinement of the study's scope, and providing thoughtful feedback on preliminary results. We thank the advisory committee for its support and assistance. A list of advisory committee members is provided in appendix 2.

Throughout the analysis, we use the collective collection of the Association of Research Libraries (ARL) membership as a comparator for the RLUK collective collection. Significant differences exist between the two consortia—in particular, ARL consists of a North American membership and is significantly larger than RLUK (124 and 37 members, respectively). Nevertheless, in consultation with the RLUK advisory committee, we concluded that the ARL collective collection provides a valuable, peer-based context against which to compare the findings from our analysis of the RLUK collective collection.

Finally, the subject analysis performed in this study is based on FAST headings found in WorldCat record subject fields. FAST (Faceted Application of Subject Terminology)⁶ is an eight-facet vocabulary that streamlines and simplifies the Library of Congress Subject Headings schema. FAST's faceted structure is well-suited for use as the basis for subject analysis. Not all WorldCat records used in the analysis contain FAST subject headings, but sufficient coverage existed to warrant use of FAST as a common denominator for subject analysis across the RLUK collective collection.

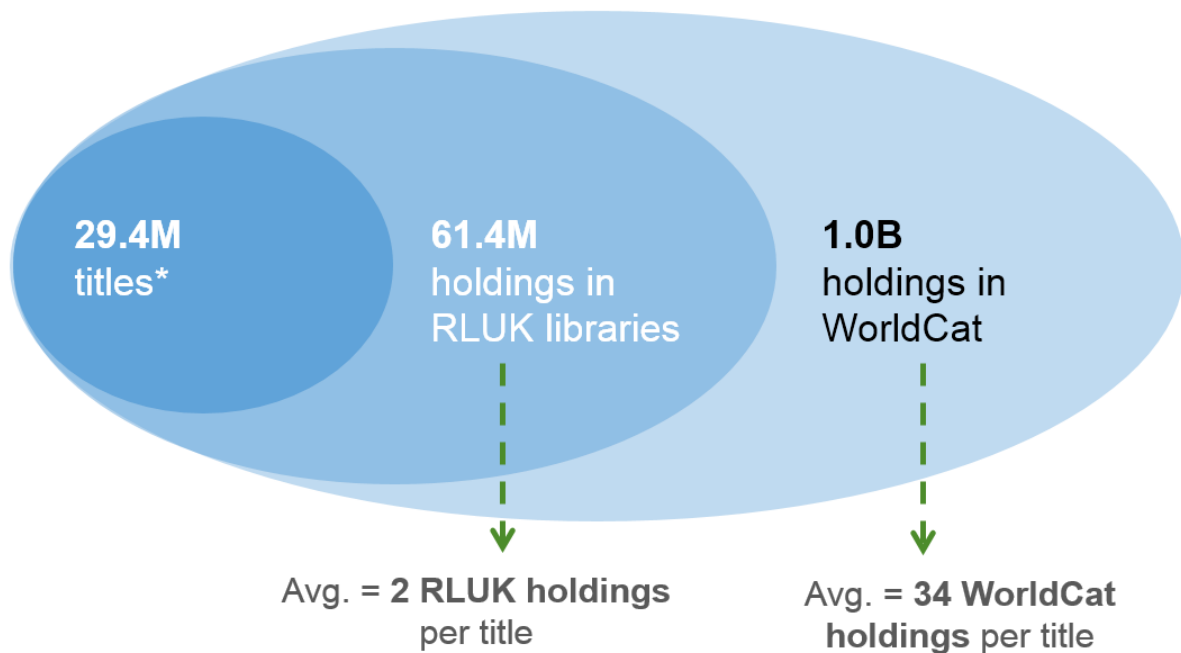
The RLUK Collective Collection

SIZE AND SCOPE

We used a January 2016 snapshot of WorldCat to generate a profile of the collective library resource of 32 RLUK consortium members. While patterns in print book holdings are the primary focus of this report, it is useful to consider the broader context of the complete set of RLUK library holdings, which includes books (including print, audio and electronic formats), serials (including government publications, journals and magazines), musical scores and other materials. We explored the breadth of the RLUK collective collection along several dimensions: the total number of de-duplicated titles, the aggregate number of RLUK library holdings for those titles and the aggregate holdings in WorldCat for the same titles. Taken in combination, these measures can be used to understand the size and scope of the RLUK collective collection as well as the intensity of library investment in that collection, within and beyond the RLUK membership.

Figure 1 provides a visual summary of the size of the RLUK collective collection. The collection comprises 29.4 million discrete titles representing 61.4 million holdings in RLUK libraries. Each of these title-level holdings may be represented by one or more local copies of a given title, so the aggregate inventory of RLUK libraries is substantially greater than this already impressive figure. While the average number of RLUK holdings per title is relatively modest (two), the total holdings in WorldCat suggest that at least some of the materials held in RLUK libraries are widely duplicated in the global library system. On average, a total of 34 libraries (including some number of RLUK libraries) have invested in each title in the RLUK collective collection. This suggests that while levels of redundancy within RLUK libraries may be relatively low, the library community as a whole has indirectly validated acquisitions decisions made in individual UK libraries by purchasing or licensing the same materials. We explore the level of duplication in RLUK holdings in more detail later in this report.

RLUK Collective Collection: Size



*distinct publications; de-duplicated OCLC numbers

Malpas and Lavoie. 2016.

FIGURE 1. RLUK COLLECTIVE COLLECTION: SIZE (JANUARY 2016)

The above figures provide a general idea of the overall *scale* of the RLUK collective collection. Some insight into the intellectual *scope* of the collection can be gained by comparing the number of titles therein to the number of creative works or intellectual expressions they embody. Here we borrow from the conceptual framework of the Functional Requirements for Bibliographic Records (FRBR) data model, which differentiates between intellectual expressions and their instantiation in specific editions or manifestations. OCLC Research has developed software that clusters related manifestations into FRBR-like work sets. We used these clusters to compare the number of titles in the RLUK collective collection to the number of creative works they embody; this provides an approximate measure of the intellectual breadth of the overall collection. Our analysis identified 19.5 million discrete works in the RLUK collective collection. Three quarters of these works are represented by a single manifestation; the remaining 25% are represented by two or more editions.

What are the implications of these findings? First, it is abundantly clear that the total volume of material held by RLUK libraries is very great, which speaks to the shared challenge of collections and space management in research institutions. At the same time, there is (on average) relatively low redundancy across the RLUK collective collection, which suggests that deduplication of local inventory will require significant coordination if preservation of the total scope of the RLUK collective collection is a priority. Moreover, given the very large proportion of titles that are represented by a single edition, it will be important to articulate a shared preservation strategy that strikes a reasonable balance between curating a cumulative record of many editions of a few works and stewarding a more comprehensive collection with more works, but (potentially) fewer editions. A balanced approach might include some prioritization of investment in works deemed to be of canonical importance across the university research library sector.

PLACES AND LANGUAGES OF PUBLICATION

The RLUK collective collection is global in scope with respect to both places and languages of publication. Titles from 333 different countries are represented, providing at least some coverage for almost 90% of the geographic sources accepted in standard cataloging practice.⁷ Unsurprisingly, the proportional representation of different countries of publication varies widely, ranging from a high of more than 9.2 million titles (31%) produced in England to a low of a single title from Saint Kitts-Nevis. Titles from the UK represent 37% of the RLUK collective collection. More than five million titles (17%) are described as having “no place, unknown or undetermined” publication location.⁸ Table 1 identifies the top ten national sources of publications in the collection.

TABLE 1. TOP TEN PLACES OF PUBLICATION IN RLUK COLLECTIVE COLLECTION

Country	Titles in RLUK Collective Collection	Percent of RLUK Collective Collection
UK	10,855,514	37%
USA	3,833,059	13%
Germany	1,734,736	6%
France	1,395,672	5%
Italy	648,558	2%
Netherlands	470,240	2%
India	429,912	1%
Russia	409,198	1%
Spain	295,675	1%
Canada	291,409	1%

The RLUK collective collection is also diverse from a linguistic perspective. English-language content accounts for about two-thirds of titles (68%), with the remaining third distributed over more than 400 other languages.⁹ A small part of the collection (1%) represents material without any linguistic content and a further 6% is in undetermined languages. The distribution of languages is highly skewed toward Western Europe, most likely an artifact of historical publishing trends. Even so, many recondite and even artificial languages such as Klingon are represented by a small number of titles. Table 2 identifies the top ten languages in the RLUK collective collection. Taken in combination, these languages account for 86% of publications in the collection.

TABLE 2. TOP TEN LANGUAGES IN RLUK COLLECTIVE COLLECTION

Language	Titles in RLUK Collective Collection	Percent of RLUK Collective Collection
English	19,895,139	68%
German	1,565,035	5%
French	1,554,480	5%
Spanish	511,019	2%
Italian	496,316	2%
Russian	457,158	2%
Latin	433,411	1%
Chinese	184,445	1%
Japanese	159,806	1%
Dutch	150,271	1%

MATERIAL TYPES

As noted, books in print format are the primary focus of this analysis given their importance in ongoing discussions of a national monographic strategy in the UK. Because university library collections comprise an ever-widening array of material types, it is useful to consider where monographs fit in the larger picture of RLUK library holdings. Library cataloging practice supports very rich and even elaborate description of material types and publication formats. Using a simplified taxonomy of material types, we examined the distribution of publications in the RLUK collective collection to evaluate the overall importance of books as a class of library resource. Books (including print, audio and electronic formats) represent 88% of the aggregate collection; serials (including a variety of continuing resources from government publications to research journals and magazines) constitute the next largest category representing 5% of titles. Musical scores account for a further 5% of the collection. The remaining 3% of titles represent a mix of maps, visual resources, sound recordings, archival resources and other materials.

Material Types in the RLUK Collective Collection (N = 29.4 M titles)

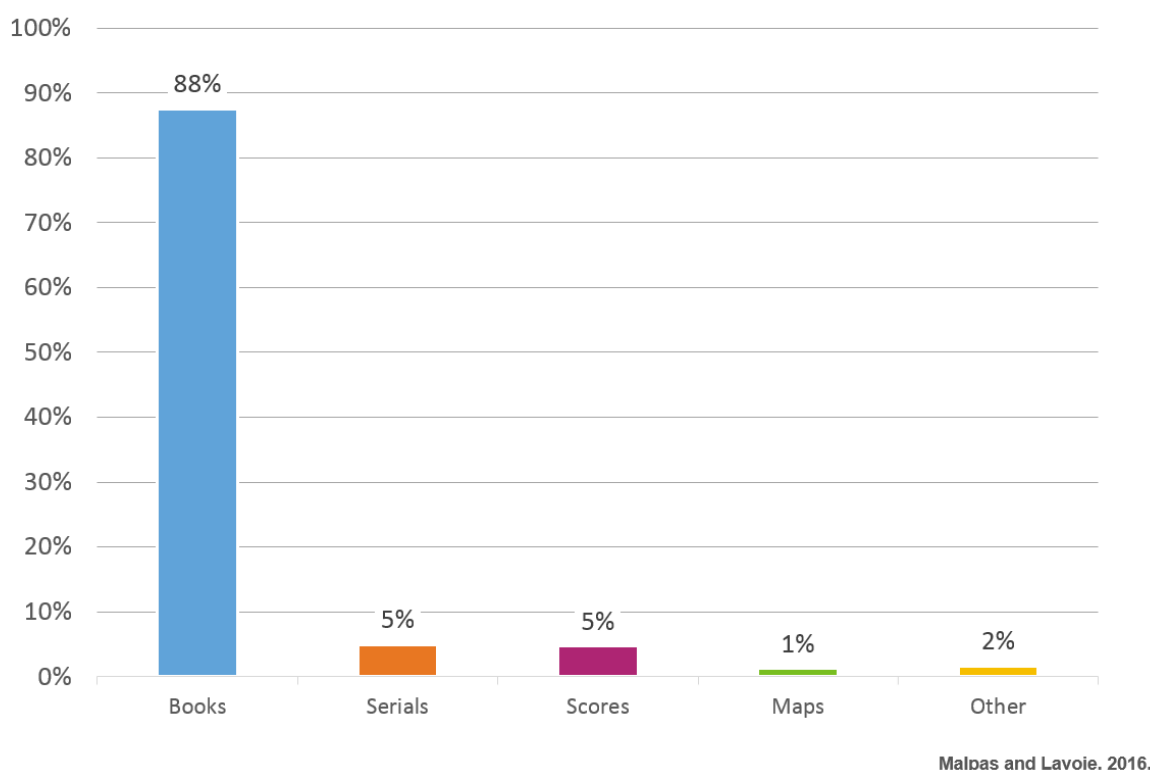


FIGURE 2. MATERIAL TYPES IN THE RLUK COLLECTIVE COLLECTION (JANUARY 2016)

Through initiatives like the UK Research Reserve,¹⁰ significant progress has been made in rationalizing management of retrospective print journal collections in university libraries. As figure 2 shows, the scope of the RLUK book collection (all formats) is much greater than any other segment of the collective library resource. This represents a significant opportunity for re-imagining the organization of library services in research universities, if not the higher education sector as a whole, to maximize the value of the aggregate resource as a shared asset while achieving some efficiency gains in local library operations.

The remainder of this study focuses primarily on print books in the RLUK collective collection.

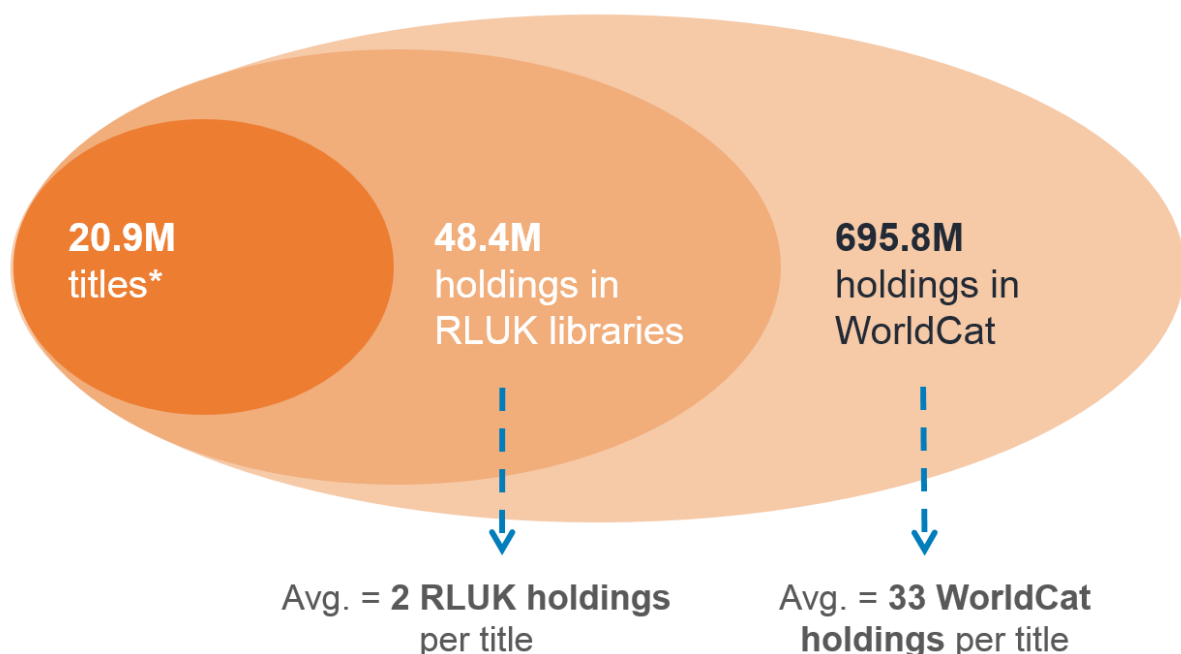
PRINT BOOKS

The RLUK collective collection, as it is represented in WorldCat, contains 20.9 million distinct print book publications, accounting for 71% of the overall resource. The print book intensity of the RLUK collection seems to be quite strong: in comparison, the ARL collective print book resource consists of 35.4 million distinct print book publications, which account for 49% of the overall ARL collective collection. It should be noted that ARL libraries have been registering their holdings in WorldCat for decades and have adopted more comprehensive record-loading practices than are in place for some RLUK libraries. For example, most if not all ARL libraries catalog rare books, manuscripts and archival collections in WorldCat, whereas only some RLUK libraries do so. Consequently, the RLUK collective collection may appear to be more print book intensive than the ARL collection simply because a less diverse range of library holdings are represented. Print books are an integral part of an academic library's identity, and yet declining use and the increasing availability of digital

surrogates have heightened interest in moving print book collections “above the institution” into some form of shared custodial arrangement, releasing local resources (including physical space) for other purposes. In this section, we examine the characteristics of the RLUK collective print book collection, with an emphasis on characteristics that are particularly relevant to the development of consortial-scale shared print programs.

The print book publications in the RLUK collective collection are linked to more than 48 million holdings across the RLUK membership, averaging about two holdings per publication. Median holdings for a print book publication held within the RLUK membership are even lower, at precisely one. Nearly 90% of the RLUK print book publications are held by fewer than five members, while in contrast, less than 1%— about 63,000 print book titles—are held by more than 20 RLUK members. Taken together, this suggests an RLUK print book resource that is thinly spread over the membership, with low levels of duplication and correspondingly high levels of scarcity in terms of in-group availability.

RLUK Collective Print Book Collection: Size



*distinct publications; de-duplicated OCLC numbers

Malpas and Lavoie. 2016.

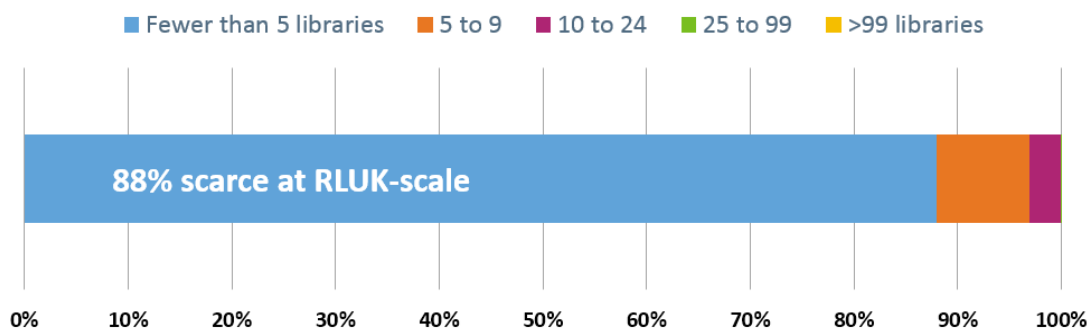
FIGURE 3. RLUK COLLECTIVE PRINT BOOK COLLECTION: SIZE (JANUARY 2016)

The ARL print book resource exhibits somewhat less scarcity than what is found in the RLUK collective collection. Print books account for almost 245 million holdings across the ARL membership; with 35.4 million distinct print book publications in the ARL collection, this results in an average of seven holdings per publication—three times higher than for the RLUK collection. Similarly, median holdings are greater in the ARL collection (two). Evidence of comparatively less scarcity in the ARL print book resource is also found in the lower percentage of the ARL print book resource that is held by fewer than five in-group libraries (73%). The larger scale of the ARL membership is likely the source of this disparity between the ARL and RLUK collections in regard to in-group availability: as the number of members increases, the degree of redundant collecting activity grows as well, even as the long tail continues to be built out at the other end of the distribution.

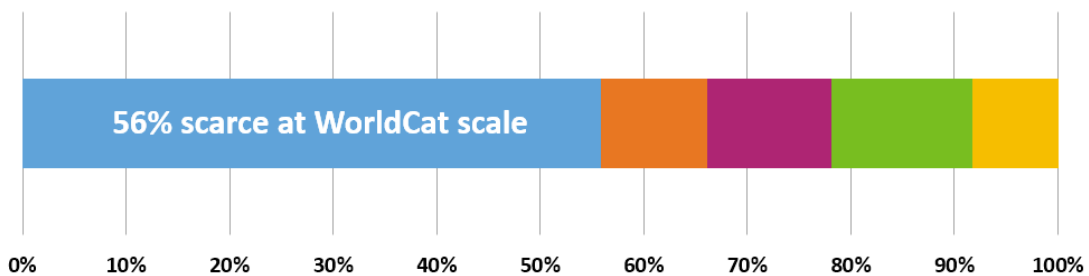
While in-group scarcity is a salient property of the RLUK-held print book publications, system-wide availability appears to be much greater. The print books held by RLUK members are, on average, held quite widely throughout the global library system as represented in WorldCat, accounting for more than 695 million global holdings and averaging about 33 holdings per publication. As noted above, nearly 90% of the RLUK collective print book collection is held by fewer than five libraries within the RLUK membership. But only 56% of the RLUK collection is held by fewer than five libraries worldwide, while nearly a quarter of the collection is held by 25 or more libraries. This suggests that the degree of scarcity associated with the RLUK print book resource is dependent upon the frame of reference in which it is placed. Some print book publications that appear scarce at group-scale—i.e., within the RLUK membership—may in fact exhibit much higher availability when evaluated at global scale—i.e., within the system-wide context represented by WorldCat.

RLUK Collective Print Book Collection: Duplication Rates

In-group Duplication of RLUK-held Print Book Titles (N = 20.9M)



WorldCat Duplication of RLUK-held Print Book Titles (N = 20.9M)



Malpas and Lavoie. 2016.

FIGURE 4. RLUK COLLECTIVE PRINT BOOK COLLECTION: DUPLICATION RATES

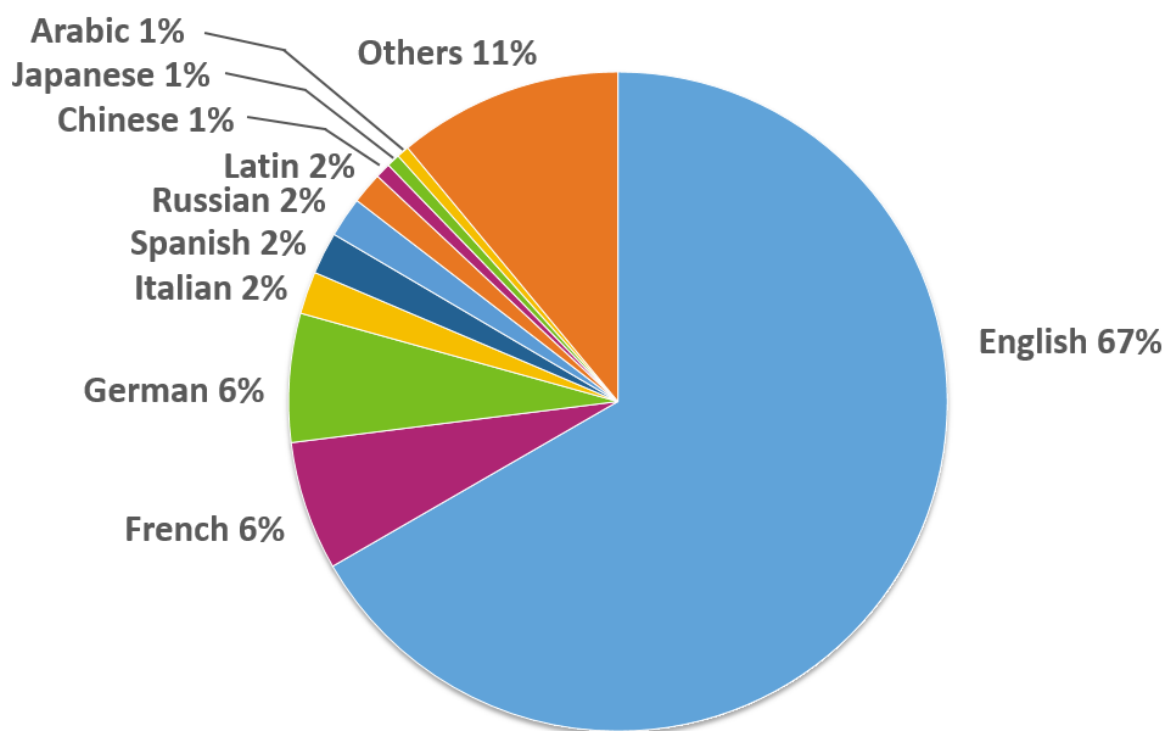
The same result is seen with the ARL collective print book resource, where 73% of the publications in the collection are held by fewer than five ARL members, but only 51% are held by fewer than five libraries globally. The 245 million in-group print book holdings attached to the publications in the ARL collection expand to nearly 1.1 billion when holdings external to ARL are included—an increase by a factor of more than four.

A print book resource containing 20.9 million distinct publications is of sufficient scale to exhibit considerable diversity in both language of content and country of publication. And, indeed, the RLUK collective collection contains print books in 467 languages originating from 254 countries, suggesting a rich diversity of publications in terms of both linguistic expression and publishing source. These

results are nearly identical to those from the ARL collective print book collection (475 languages and 255 countries of publication), even though the ARL collection is much larger. Of course, limits exist on the number of languages and countries that can be represented in a collective collection, since there are a finite number of languages and countries. Comparison of the RLUK and ARL results therefore suggest that scope increases (i.e., number of languages, number of countries) with the scale of the collective collection up to a point, after which additional scale tends to add to depth (i.e., more publications associated with particular languages or countries).

English-language materials predominate in the RLUK print book resource, accounting for two-thirds of the publications, with French (6%), German (6%), Italian (2%) and Spanish (2%) rounding out the top five languages in the collection. English-language print book publications account for more than three-quarters of all RLUK print book holdings, with average holdings (2.6) slightly higher than for the print book resource as a whole (2.3). The language profile of the ARL print book resource is slightly different from that of its RLUK counterpart: while English-language materials represent the largest segment of the resource, they constitute only 48% of the publications in the ARL collection.

RLUK Collective Print Book Collection: Language Of Content



Malpas and Lavoie. 2016.

FIGURE 5. RLUK COLLECTIVE PRINT BOOK COLLECTION: LANGUAGE OF CONTENT (% OF PUBLICATIONS)

Print books published in the UK are a plurality in the RLUK print book resource, but account for only about a third of the collection (35%), followed by print books published in the US (13%), Germany (6%), France (5%) and Italy (3%). Domestically published print books account for less than half of total RLUK print book holdings (47%), but like English-language books, average holdings are slightly higher (3.1) than for the print book resource as a whole. Like RLUK, domestically published materials (United States or Canada) are a plurality in the ARL collective print book collection, but fall far short of a majority (31%). Print books published in the UK represent the second-largest segment of the ARL print book resource (10%).

TABLE 3. RLUK COLLECTIVE PRINT BOOK COLLECTION: TOP TEN COUNTRIES OF PUBLICATION

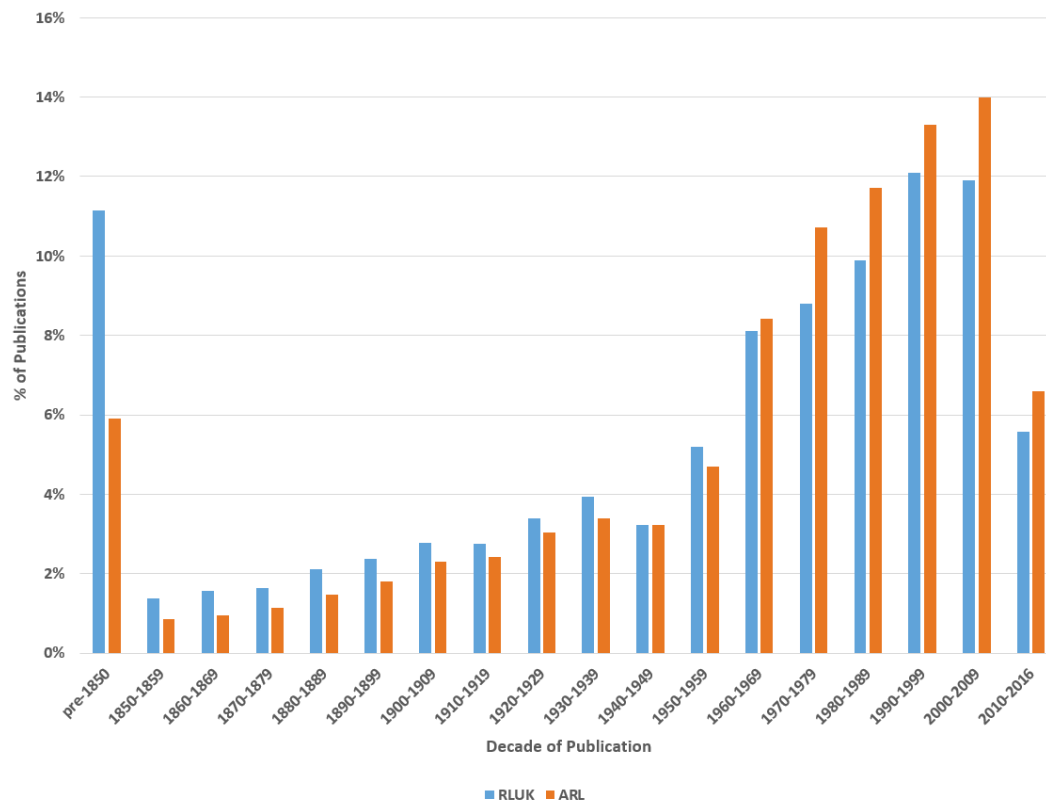
Country	Titles in RLUK Print Book Collection	Percent of RLUK Print Book Collection
UK	7,370,417	35%
USA	2,634,590	13%
Germany	1,351,147	6%
France	1,125,343	5%
Italy	559,758	3%
India	367,519	2%
Russia	367,061	2%
Netherlands	358,183	2%
Spain	253,493	1%
Switzerland	203,086	1%

The UK is home to a number of indigenous languages besides English, and these are reflected in the RLUK print book resource as well. Welsh-language books are the most prevalent, with nearly 50,000 publications, and more than three times as many as the next language, Irish. Other indigenous languages—Scottish Gaelic, Scots, Cornish and Manx—have comparatively light representation in the RLUK print book resource, each accounting for less than 1% of the publications. Perhaps unsurprisingly, the number of publications for each of these languages is less in the ARL collection than in RLUK; however, their relative ranking is identical in both collections.

The age of the RLUK collective print book collection skews young, with the majority of the collection published in the post-war period (more than 60% published since 1950), and almost a third of the collection published since 1990. A significant portion of the collection (11%) originates from the pre-1850 period, a threshold often used to demarcate materials that are considered historical artifacts.¹¹ In the post-1850 period, three dips occur in an otherwise steady upward progression in the number of print books published in each decade: 1910–1919, which presumably can be attributed to World War I;

1940–1949, which is likely a result of World War II; and 2000–2009. The origins of the latest dip may be partially attributable to cataloging lag, but may also be attributable to the emergence of digital content as a viable, and, in some cases, preferred option to print.¹²

RLUK and ARL Collective Print Book Collections: Age Distribution



Malpas and Lavoie. 2016.

FIGURE 6. RLUK AND ARL COLLECTIVE PRINT BOOK COLLECTIONS: AGE DISTRIBUTION, BY DECADE OF PUBLICATION (% OF PUBLICATIONS)

Like the RLUK collection, about a third of the ARL collective print book collection was published since 1990. However, several points of divergence exist between the two collections: a slightly higher fraction of the ARL collection (69%) was published since 1950, while the percentage of the ARL collection consisting of pre-1850 materials is significantly lower (6%). The ARL collection exhibits the same dip in the number of print books published in the period 1940–1949 as seen in the RLUK collection. Unlike the RLUK collection, however, the ARL print book resource does not have a dip in the 1910–1919 decade, nor in the 2000–2009 decade. The lack of a dip from 1910–1919 may reflect the relatively brief American engagement in World War I, compared to Europe; the lack of a dip from 2000–2009 is more difficult to explain, although it is interesting to speculate whether the growing availability of digital surrogates had different impacts on RLUK and ARL members’ collecting behaviors.

Like any large aggregation of academic library holdings, the RLUK collective print book resource contains materials representing a wide range of subjects. In order to gain a sense of the diversity and relative predominance of the various subjects found within the RLUK collection, all FAST subject headings were extracted from the WorldCat records representing these materials; these headings were then ranked by frequency of occurrence. More than 458,000 different FAST headings were

identified through this process. History is far and away the most frequently occurring FAST heading, occurring nearly 1.4 million times within the RLUK print book resource—nearly two-and-a-half times more frequently than the next most common FAST heading, Great Britain. Humanities and the social sciences figure prominently in the top ten most frequently occurring subject headings, suggesting that the print monographic literature has been an important channel of scholarly communication for these disciplines.

TABLE 4. TOP TEN MOST FREQUENTLY OCCURRING FAST HEADINGS IN RLUK COLLECTIVE PRINT BOOK COLLECTION

FAST Heading	Number of Occurrences
History	1,365,572
Great Britain	577,749
Criticism, interpretation, etc.	473,562
1900–1999	409,430
United States	391,766
Biography	388,214
England	370,386
Conference papers and proceedings	346,859
Politics and government	327,534
Fiction	294,307

It is interesting to note that the ARL print book resource, while containing a wider variety of FAST subject headings (about 774,000), nevertheless exhibits a top ten ranking of most frequently occurring headings that is strikingly similar to that of the RLUK print book resource. Eight of the top ten headings appearing in the ARL list also appear in the RLUK top ten list (albeit in differing order). History is the predominant heading for ARL as it is for RLUK, while United States is the second most frequently encountered heading. Like the RLUK print book resource, humanities and the social sciences account for a significant portion of the ARL top ten list.

Internal Overlap

The RLUK collective collection is a rich and varied resource encompassing a broad range of intellectual content and publication formats managed in repositories of various sizes and capacities, serving institutions with diverse missions. To examine how responsibility for stewardship of the collective resource might be distributed across the group, it is important to understand the existing distribution of the aggregate inventory. How much content is duplicated within the collective collection, and at what level of duplication? We approached these critical questions from two different vantage points. First, we examined duplication in RLUK holdings from a title-level perspective, calculating

aggregate holdings set on individual titles. Second, we examined duplication in the intellectual content held in RLUK libraries by exploring the degree to which multiple RLUK libraries have invested in different versions of the same works, i.e. acquiring different editions of a single title.

In this section, we begin by exploring duplication along these two dimensions within the RLUK collective collection as a whole, and then focus more closely on duplication patterns for print books.

DUPLICATION IN RLUK LIBRARY HOLDINGS

As noted previously, the RLUK collective collection comprises 29.4 million publications with a total of more than 61 million holdings in RLUK libraries. These figures suggest that duplication in RLUK holdings is very low, with approximately two libraries per title on average. This finding is partly explained by differences in local cataloging practices, which can produce bibliographic descriptions that are not easily reconciled, so that holdings which might otherwise be grouped together are instead scattered across different versions of a bibliographic resource describing the same publication.¹³ Figure 7 below provides a high-level view of title-level duplication in the RLUK collective collection, using thresholds we have found to be useful in similar analyses in the past. Few would disagree that a title held by fewer than five libraries in a group of 32 is relatively scarce; likewise, titles held by more than two-thirds of the group will almost certainly be regarded as common. The upper limit of duplication in the RLUK collective collection was 31, or about 90% of the library symbols included in the study. Fewer than 30 titles were duplicated at this level. We take a closer look at some of these later in this section; a full list is provided in appendix 3.

In-group Duplication of Titles in RLUK Collective Collection (N = 29.4M titles)

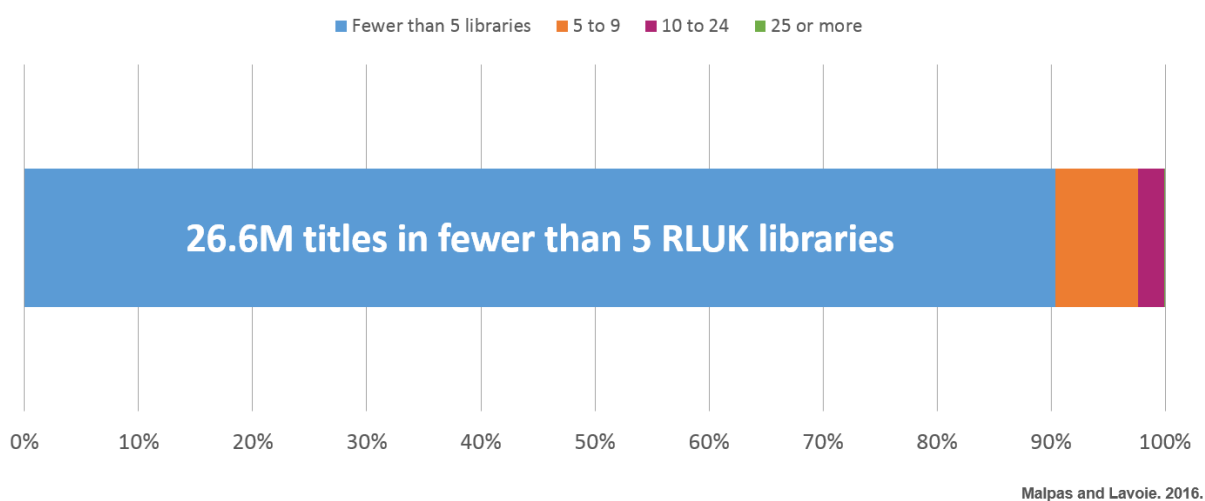


FIGURE 7. IN-GROUP DUPLICATION OF TITLES IN RLUK COLLECTIVE COLLECTION (JANUARY 2016)

It is readily apparent that the aggregate resource is unevenly dispersed, with 90% of titles held by just a few libraries in the RLUK group. If duplication is measured at the global scale of WorldCat, a different picture emerges. As figure 8 shows, the proportion of titles that are held by fewer than five libraries is reduced to about 60% when libraries outside of the RLUK group are included, suggesting that even for specialized “long tail” resources that are not widely produced or collected by libraries, there is latent redundancy in the global library system that can be leveraged in support of shared preservation goals.

WorldCat Duplication of Titles in RLUK Collective Collection (N = 29.4M titles)

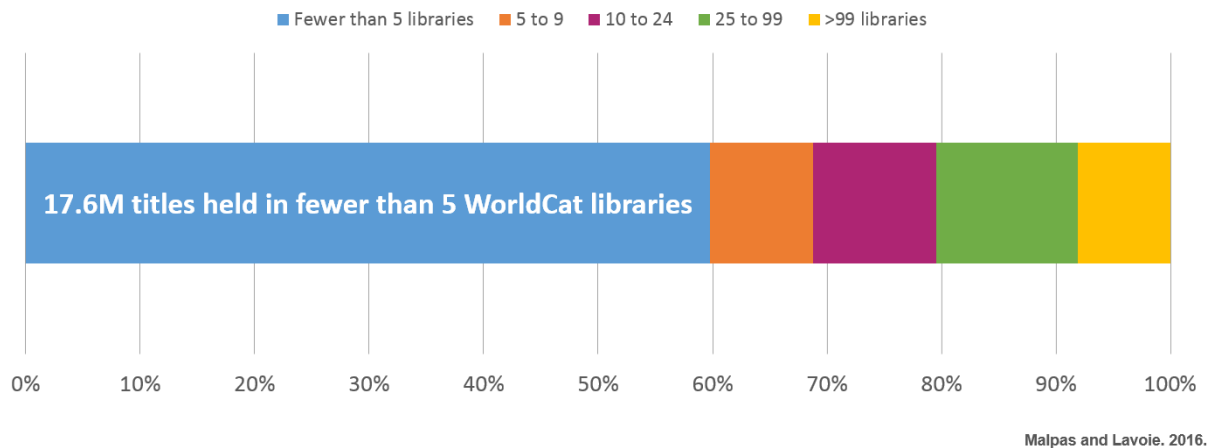


FIGURE 8. WORLDCAT DUPLICATION OF TITLES IN RLUK COLLECTIVE COLLECTION (JANUARY 2016)

From an operational perspective, RLUK libraries cannot strike arrangements with every library in WorldCat to ensure that desired levels of duplication are maintained beyond the RLUK membership. But it is still instructive to see that 30% of titles that might otherwise be deemed at risk because aggregate RLUK holdings are low are duplicated in other libraries. Equally important is the fact that while only about 15,000 titles in the collective collection are held by 25 or more RLUK libraries, more than six million titles are duplicated in 25 or more libraries in WorldCat. Thus, by aggregating holdings on a greater scale, it is possible to reduce the overall risk profile while simultaneously increasing the potential yield on library space recovery within RLUK. We explore some of these external sources of duplication later in this report.

DUPLICATION OF INTELLECTUAL CONTENT: WORKS AND VERSIONS

A closer examination of FRBR-like work sets in the RLUK collective collection provides additional insights into the range of intellectual content it comprises and the degree to which RLUK libraries exhibit shared patterns of investment in acquiring similar content in different editions and formats. As reported above, we identified 19.5 million works, or title clusters, in the aggregate RLUK resource, each containing one or more related versions of the same intellectual content held in one or more RLUK libraries. The vast majority of these works are represented by a single bibliographic description, representing a publication for which a single version is available within the RLUK membership.¹⁴ Some of these are widely held by RLUK libraries. For example, Barbara Kirshenblatt-Gimblett's *Destination Culture: Tourism, Museums, and Heritage* (Berkeley: University of California Press. 1998) is held in almost 90% of RLUK libraries. However, most of the works with a single version in RLUK libraries belong to the long tail of the scholarly record and are not widely duplicated within the group.

In the context of cooperative stewardship, titles that are represented by many versions or formats in RLUK libraries are of particular interest. The fact that multiple institutions have acquired different versions of the same intellectual content provides an indirect measure of its worth to the collective, as determined by library selectors, bibliographers and faculty. Conversely, if a single member library has acquired multiple versions of a publication while others have collected fewer or even none, it may represent a specialized local interest that is not widely shared across the group. A deeper understanding of the intellectual content that is shared by RLUK may be a useful complement to the traditional approach of counting the number of titles or copies that are duplicated within the group.

By examining the intellectual works that are most widely represented within the RLUK collective collection, in terms of the number of editions or formats held anywhere within the group, we can arrive at some idea of the content that has been most collected by these university libraries. Table 5 below lists ten of the largest work set clusters in the RLUK collective collection.¹⁵ Classics from the canon of Western literature dominate the list, including works of popular fiction and non-fiction published in multiple editions over several centuries. Notably, publications by UK authors are particularly well represented, a pattern we have observed in other analyses of regional collections; publications of and about places tend to be well represented in aggregate collections of libraries in a given geographic region.¹⁶

TABLE 5. WORKS WITH THE LARGEST NUMBER OF EDITIONS (FORMATS, VERSIONS) IN RLUK LIBRARIES.

Work	Versions in RLUK Collective Collection	Aggregate RLUK Library Holdings
Dante. <i>Divine Comedy</i>	956	1,642
Horace. <i>Opera Omnia</i>	930	1,423
John Bunyan. <i>Pilgrim's Progress.</i>	926	1,473
James Thomson. <i>The Seasons.</i>	909	1,433
Henry Fielding. <i>Tom Jones.</i>	904	1,570
Oliver Goldsmith. <i>Vicar of Wakefield.</i>	903	1,416
William Makepeace Thackeray. <i>Works.</i>	815	1,352
John Milton. <i>Works.</i>	792	1,389
Julius Caesar. <i>Commentaries on the Gallic War.</i>	741	1,008
Daniel Defoe. <i>Adventures of Robinson Crusoe.</i>	674	1,071

It is not especially surprising that titles with a long publication history would tend to be well-represented in university library collections; the sheer number of formats and editions produced over many years is a testament to their enduring cultural relevance and interest as objects of cultural production. In the context of collaborative collection management, the relative abundance of these titles is an indication of their importance and not necessarily a sign of undesirable redundancy.

Comparing the best represented works in the RLUK collective collection to the best represented manifestations or editions, it appears that titles with a longer pedigree (greater age and more diversity of publication formats) have a relative advantage when it comes to overall popularity or presence in the library record. The aggregate holdings of RLUK libraries on classic works like Dante's *Divine Comedy* or Fielding's *Tom Jones* significantly exceeds group holdings on even the most widely duplicated publications in the collective collection. This is easily seen by comparing the figures reported in table 5 to the individual editions that are most widely held in RLUK libraries. Table 6 lists a sample of the editions most widely held by RLUK libraries. These are a subset of the 27 publications

we identified as having the most edition-level holdings in RLUK. (A complete list is included as appendix 3.) Even for the most popular title, a reference source documenting the administrative history of the United Kingdom, aggregating all of the RLUK holdings for all editions produces a figure that is substantially less than the aggregate holdings on any one of the ten largest work sets presented in table 5. It is fair to say that *The Adventures of Robinson Crusoe* is more widely available in research university library collections in the UK than even the most important works of contemporary scholarship. This is a by-product of the centuries long publication history of *Robinson Crusoe* (or *Pilgrims Progress*, etc.), compared with modern scholarly works such as *The World Turned Upside Down*, which was issued in three print editions between 1972 and 1978 and multiple printings through the 1990s.

TABLE 6. EDITIONS MOST WIDELY-HELD IN RLUK COLLECTIVE COLLECTION (JANUARY 2016)

OCLC number	Title	Date of most widely held edition in RLUK	No. of RLUK library symbols on most widely held edition	Aggregate RLUK holdings for all editions and versions
24009791	E B Fryd et al. <i>Handbook of British Chronology.</i>	1986	31	97
447768	Christopher Hill. <i>The World Turned Upside Down: Radical Ideas During the English Revolution.</i>	1972	31	76
1884048	Mary Douglas. <i>Implicit meanings: Essays in Anthropology.</i>	1975	31	61
8430152	Julia Kristeva. <i>Powers of Horror: An Essay on Abjection.</i>	1982	31	56
12052097	David Lowenthal. <i>The Past is a Foreign Country.</i>	1985	31	38
24847759	Hal Varian. <i>Microeconomic Analysis.</i>	1992	31	71
21375348	Thomas Laqueur. <i>Making Sex: Body and Gender from the Greeks to Freud.</i>	1990	31	36
22597117	Elazar Barkan. <i>The Retreat of Scientific Racism: Changing Concepts of Race in Britain and the United States Between the World Wars.</i>	1992	31	35
23869294	Terence Ranger; Paul Slack (Eds.). <i>Epidemics and Ideas: Essays on the Historical Perception of Pestilence.</i>	1992	31	37
24373115	L D Schwarz. <i>London in the Age of Industrialisation: Entrepreneurs, Labour Force, and Living Conditions, 1700-1850.</i>	1993	31	36

RLUK libraries have literally put great stock in Christopher Hill's scholarship on the English Revolution and in the other seminal works listed in table 6.¹⁷ By examining the RLUK collective collection from the perspective of both works and versions or editions, we can better appreciate the degree to which research libraries have expressed a shared interest in the intellectual content of a publication, even if institutional choices with respect to which edition to acquire or retain have differed. Thus, while the 1992 edition of Hal Varian's *Microeconomic Analysis* has accrued the most holdings in RLUK libraries, earlier editions (1978 and 1984) are still retained by many RLUK partners. Julia Kristeva's *Power of Horror* is most widely held in an English language edition published in 1982, but a 1980 edition in French is held by 16 RLUK libraries.

Whether a collaborative stewardship arrangement should seek to preserve the broadest possible record, retaining multiple editions of works that have been broadly collected, or prioritize investment in the editions that have proven especially popular will depend on many factors; here, we intend only to suggest that a FRBR-informed perspective on the RLUK collective collection may be helpful in identifying patterns of common institutional interest, acknowledging that title-level holdings are more meaningful when considered in the context of broader collecting trends.

SHARED STRENGTHS: SUBJECT COVERAGE AND COMMUNITIES OF INTEREST WITHIN RLUK

Beyond identifying intellectual content that is shared by many RLUK libraries, a FRBR-informed approach to analyzing the RLUK collective collection can provide insights into topical subject areas that represent distinctive strengths of the aggregate resource. We used a February 2016 dataset comprising all of the WorldCat library holdings set on records with FAST and VIAF¹⁸ identifiers to explore subject areas in which RLUK libraries collect broadly both individually and collectively. By comparing the number of FRBR works associated with a given subject area in WorldCat to the FRBR work sets that are held in RLUK libraries, we can arrive at an approximate measure of the breadth and coverage of the global publication record by topic and library.

We identified nearly 3,000 FAST and VIAF headings for subjects that are well represented in individual RLUK libraries, relative to other subject areas in the local collection. These represent locally distinctive strengths and usually reflect specialized collecting interests. Thus while the Wellcome Library collects broadly in the area of medical history, it has particularly comprehensive holdings related to the history of AIDS, with broad coverage of literature related to events (e.g., World AIDS day; FAST ID: 1180717) and organizations (e.g., the Terrence Higgins Trust, FAST ID: 672072) that have shaped social understanding of this disease. Likewise, the University of Exeter Library collects European literature broadly, but has notable strength in Swiss fiction in both French (FAST ID: 1140580) and German (FAST ID: 1140582).

By looking above the institution-scale strengths of individual RLUK libraries, we can identify subject areas of shared interest that might be leveraged in collaborative stewardship arrangements. Three RLUK libraries (Trinity College Dublin, University of Exeter and the British Library) rank within the top 25 most comprehensive collections of Swiss literature in French, with each holding more than 50% of the related works in WorldCat. There will be some duplication across these collections, of course, but with respect both to the range of works and the editions represented there is a degree of complementarity. This has two important implications. First, libraries actively building collections in this area can look to RLUK peers with established expertise in acquiring, describing and managing related resources. Second, there may be opportunities for selective de-duplication of editions within these work sets, creating additional room for extending the range of the collective resource in a subject area that is or has been a focus of collecting activity for all of these institutions. Intelligence about retrospective collection development choices can inform prospective collection management strategy.

Table 7 presents an illustrative list of subjects in which multiple RLUK libraries have collected relatively comprehensively compared to other subject areas in the local collection. The scope of institutional coverage within RLUK varies quite widely for some topics, suggesting that the intensity of local interest (or collecting capacity) is not evenly distributed, even in subject areas that are demonstrably of some interest to several libraries in the group. Thus, while Oxford University libraries provide access to more than half of the global literature on the transmission of texts (FAST ID: 1154863), a longstanding area of scholarship in intellectual history, the University of Birmingham provides coverage of about a fifth of the related works. This is arguably a much better scenario than having six (or more) RLUK libraries offer equivalent coverage of this specialized literature at the expense of other materials that enrich the local (and aggregate) collection. Likewise, the fact that at least 12 RLUK libraries collect in the area of evidence-based medicine (FAST ID: 917247) suggests that there is a community of interest that might be leveraged in a cooperative stewardship arrangement.

TABLE 7. SHARED SUBJECT STRENGTHS IN RLUK COLLECTIVE COLLECTION (FEBRUARY 2016)

Subject (FAST)	Works in WorldCat	RLUK Centers	Institution-scale (Minimum)	Coverage of Works (Maximum)
Great Britain	1,065,373	17	6%	29%
Anglo-Saxons	1,349	16	21%	46%
Labour Party (Great Britain)	4,134	12	15%	35%
Evidence-based medicine	1,912	12	12%	18%
Personal narratives—British	8,030	9	10%	40%
Scottish poetry	2,965	9	14%	45%
Transmission of texts	1,023	6	21%	51%
Middle East--Persian Gulf Region	3,855	6	10%	17%
Revolution of 1688 (Great Britain)	1,375	5	21%	26%
Popish Plot (1678)	2,076	3	28%	31%

An understanding of shared collection strengths within RLUK may help to delineate potential partnerships around curation of specialized topical concentrations within the aggregate resource. Similarly, an understanding of singular institutional collection strengths can be helpful in identifying *de facto* stewards of materials in specific areas. As an example, SOAS (University of London) Library has an unrivalled collection of Hausa fiction produced by authors from a cultural/language diaspora of Northern Nigeria. Hausa literature is a topic of special interest to researchers at SOAS and so it is not surprising that the library has collected intensively in this area.¹⁹ Several RLUK institutions are home to prominent research centers in African Studies, and broader disciplinary interests in cultural anthropology, British colonial history, etc., are shared across the UK higher

education community. However, SOAS is the only RLUK library with substantial holdings to support the study of Hausa fiction and this distinctive capacity is not only good for the reputation of the university as a center of research in African Studies, it also contributes to the overall strength of the RLUK consortium.

Having now explored several dimensions of duplication within the RLUK collective collection as a whole, we turn to an examination of overlap within the subset of print book holdings.

DUPLICATION IN RLUK PRINT BOOK HOLDINGS

The RLUK collective collection of print books numbers 20.9 million titles with 48.4 million holdings distributed across the 32 libraries included in our analysis. The holdings distribution pattern is similar to that of the overall collective collection, with a majority of titles held by a relatively small number of libraries. Figure 9 provides a view of title-level duplication within the print book collection, using the same thresholds employed in our analysis of the RLUK collective collection as a whole.

In-group Duplication of Print Book Titles in RLUK Collective Collection (N = 20.9M titles)

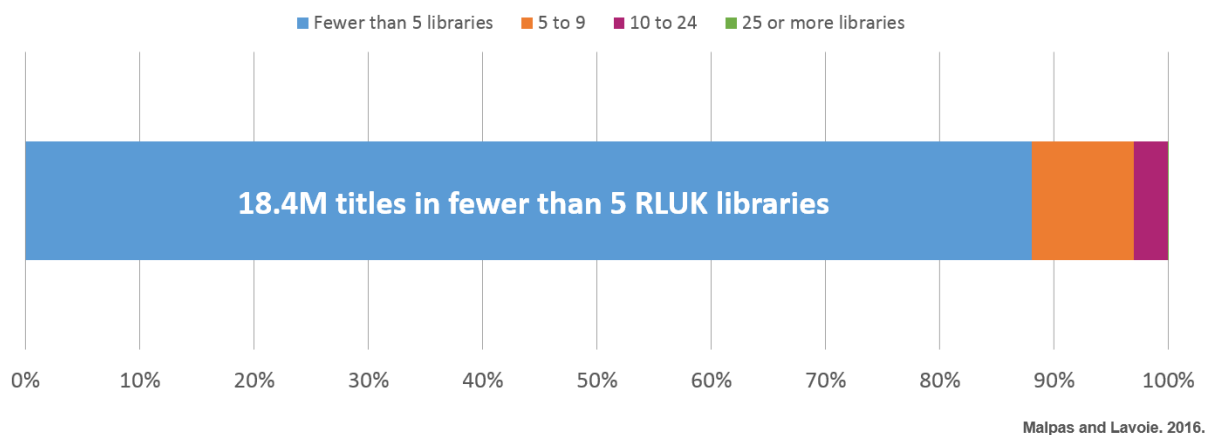


FIGURE 9. IN-GROUP DUPLICATION OF PRINT BOOK TITLES IN RLUK COLLECTIVE COLLECTION (JANUARY 2016)

As we observed with the RLUK collective collection overall, increasing the scale at which duplication is measured reduces the proportion of materials that are distinctive or scarce. Figure 9 demonstrates this by recasting the scale of analysis to all of WorldCat. At group scale, 88% of print book titles in the collective collection are scarce (held by fewer than five libraries); if we include holdings of other libraries represented in WorldCat, the proportion of scarce materials is reduced to 56%. This suggests that strategic partnerships between RLUK libraries and other libraries with a shared stake in stewardship of print books could substantially reduce the burden that RLUK libraries would otherwise shoulder alone by distributing preservation responsibilities across a broader number of institutions. The feasibility of such arrangements will be determined by the level of trust that exists or can be cultivated among RLUK and other consortia, and the tradeoffs of reliance on external partnerships, including logistical factors in the supply of print books from remote locations.

WorldCat Duplication of Print Book Titles in RLUK Collective Collection (N = 20.9M titles)

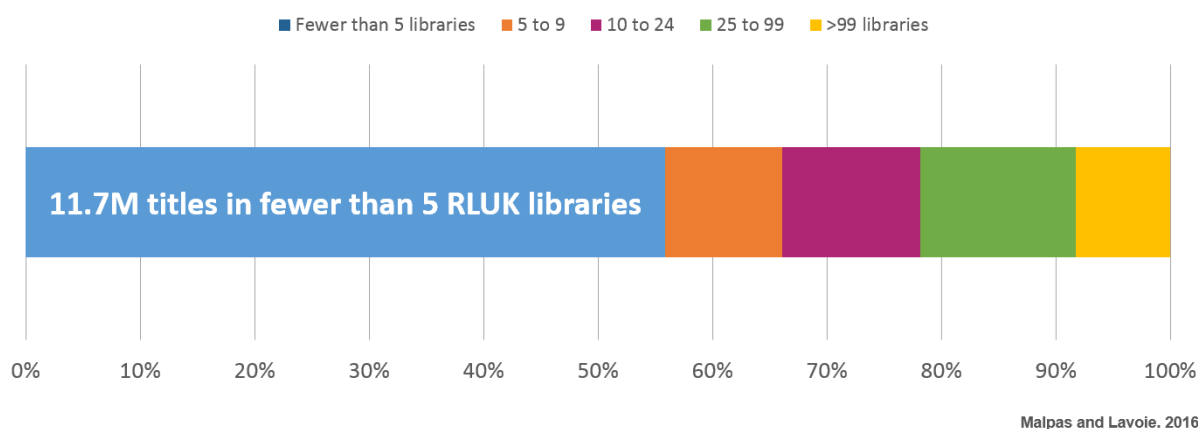


FIGURE 10. WORLDCAT DUPLICATION OF PRINT BOOK TITLES IN RLUK COLLECTIVE COLLECTION (JANUARY 2016)

For convenience, our analysis of group and global duplication of print books in the RLUK collective collections has relied up to this point on relatively arbitrary thresholds: fewer than five libraries, ten to 24 libraries, more than 99 libraries, etc. This facilitates comparison of duplication rates across populations of very different scale—i.e., the 32 RLUK libraries included in this study and the tens of thousands of libraries whose holdings are represented in WorldCat. From an operational perspective, it is more useful to sort the RLUK collective collection into classes of material that are scarce or common within the context of the consortium since strategies for managing these resources as shared assets are likely to be quite different.

In a prior collective collection project,²⁰ we segmented a consortium-level print book resource along fixed proportions of group holdings rather than absolute numbers to define rare and core materials. Table 8 provides a similar view of the RLUK’s collective print book collection, dividing the membership of 32 libraries into quarters and assessing duplication of holdings along these lines. Print book publications that are held by 24 or more of RLUK libraries (75% of group) are designated as core while those held by eight or fewer are designated as rare within the group. While the overall distribution is not markedly different from the pattern shown in table 8, with a large majority of titles held by a small number of RLUK libraries and a much smaller proportion—less than 1% of the collection—held by many RLUK members, it has the advantage of being more easily compared to the distribution of print book holdings in other groups.

TABLE 8. RARE AND CORE SEGMENTS OF THE RLUK AND ARL COLLECTIVE PRINT BOOK COLLECTIONS

Held by ...	RLUK		ARL	
	Number of Titles	Percent of Collection	Number of Titles	Percent of Collection
25% or less of membership (Rare)	20,216,838	97%	33,598,008	95%
25% to 50% of membership	536,758	3%	1,138,770	3%
50% to 75% of membership	115,987	< 1%	522,340	1%
75% or more of membership (Core)	5,417	< 1%	171,215	< 1%

The fact that only a very small part of the RLUK print book collection (less than 1%) is held by three-quarters of the group is less surprising than it may seem. Libraries that support research institutions tend to acquire highly specialized materials that are produced in relatively small quantities; many also have substantial historical collections that cannot be duplicated except in facsimile. Aggregating these individual long-tail collections simply extends the tail. Nonetheless, it is noteworthy that 96% of the RLUK print book collection—more than 20 million titles—is held by fewer than a quarter of the RLUK membership. This has important implications for cooperative collection management, as it suggests that stewardship responsibility is already highly distributed across the group, with individual institutions managing distinctive portions of the collective resource, albeit without explicit coordination.

We observed a similar long-tail effect in a prior study exploring the aggregate print book collection of a consortium about a third the size of RLUK: 76% of titles in the collective collection were held by a quarter or fewer of the group of 13 libraries. For purposes of comparison with RLUK, we produced a similar analysis of rare and core titles in the collective collection of ARL libraries. The results are strikingly similar to those observed with the RLUK collection: 95% of titles are held by a quarter or fewer of the membership; less than 1% of titles fall within the core segment.

Taken in combination, these findings suggest that even very large aggregations of print book collections in research libraries exhibit relatively low levels of duplication. This would seem to pose a challenge to cooperative stewardship arrangements, since the incentives to contribute to the preservation of content acquired by other organizations may be quite low in the absence of direct, local benefit in library space recovery or some other positive tradeoff such as expedited on-demand access to inventory held elsewhere in the system. Simply put, the business case for collaborative stewardship of a heterogeneous and geographically dispersed inventory of mostly low-use monographs must be more attractive than the status quo. Looking above the consortium to identify additional sources of duplication can help limit the share of preservation responsibility that is shouldered by individual RLUK libraries.

External Overlap

While in-group overlap provides an important perspective on both redundancy and distinctive strengths associated with RLUK collections, other perspectives are also important in framing a view of the relative distinctiveness of the RLUK print book resource. In this section, we examine two such perspectives: overlap of the RLUK collective print book collection with 1) the ARL collective print book collection, an important peer consortium; and 2) the digitized monographs in the HathiTrust digital

repository, a significant corpus of digitized content. Our focus here is on print books, since comparison of the collective print book resource with peer resources and collections of digitized content provides valuable insight touching on a number of library strategic interests, including print retention policies, cooperative print management programs and digitization strategies.

OVERLAP WITH THE ARL PRINT BOOK RESOURCE

The ARL collective collection, compiled from the local collections of 124 member libraries, includes 71.7 million distinct publications of all descriptions, 35.4 million of which (49%) are print books. These print book publications account for 244.8 million holdings within the ARL membership and 1.1 billion holdings globally.

Comparison of the 20.9 million publications in the RLUK print book resource to the ARL collection yields an overlap of 8.8 million publications, amounting to 42% of the RLUK collection and 25% of the ARL collection. Figure 11 provides a visual representation of the overlap between the ARL and RLUK print book collections. The overlapping publications tend to skew toward the more widely held materials in the RLUK print book resource. Whereas nearly 90% of the overall RLUK print book resource is held by fewer than five RLUK members, 9% by five to nine libraries, and 3% by more than nine libraries, these percentages shift appreciably in the context of the segment overlapping with the ARL collection. Of the ARL collection overlap, 79% are held by fewer than five RLUK members, 15% by five to nine members, and 6% by more than nine members. This indicates that materials that appear in both the RLUK and ARL print book resources exhibit somewhat heavier collecting activity within the RLUK membership vis-à-vis the rest of the RLUK collection, which, in turn, suggests that the RLUK membership converge in the collecting decisions they share with ARL members.

Overlap between ARL and RLUK Print Book Collections

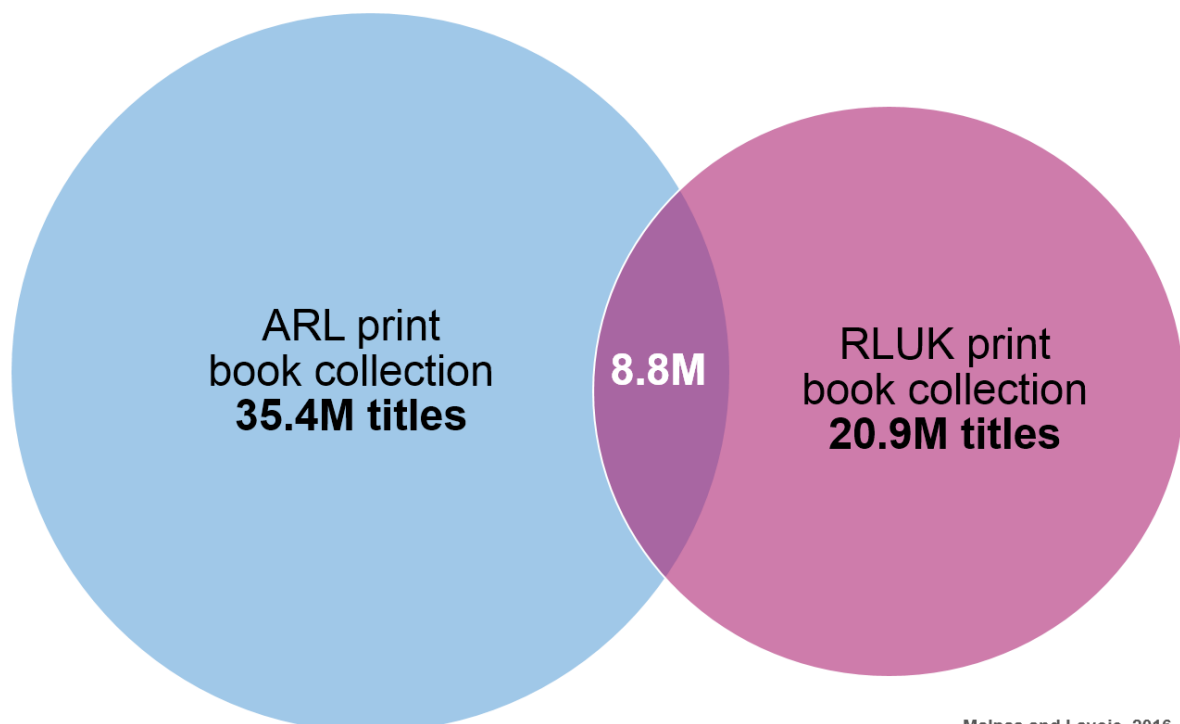
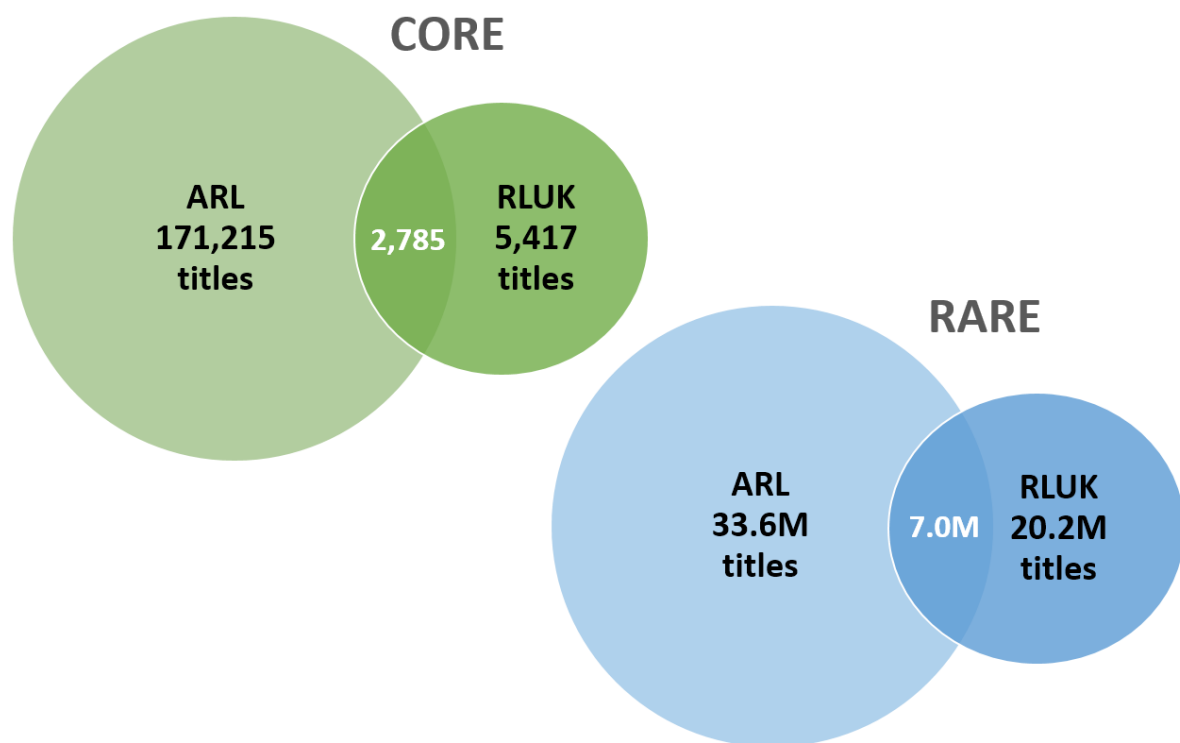


FIGURE 11. OVERLAP BETWEEN ARL AND RLUK PRINT BOOK COLLECTIONS (JANUARY 2016)

The previous section described the rare and core segments of the RLUK collective print book resource; these segments of the RLUK collection can be compared to their counterparts in the ARL collection. An overlap of seven million publications is found across the rare segments of the two collections, or 35% of the RLUK rare segment. For the core segments, we find an overlap of 2,785 publications, or 51% of the RLUK core segment. Figure 12 provides a visual depiction of the overlap for the rare and core segments. These results suggest a significant correlation between RLUK and ARL in regard to materials that, from an in-group perspective, are either relatively rare or relatively core within the membership. Put another way, print book publications that are rare in the RLUK consortium setting have a strong likelihood of being rare in the ARL setting as well, while print books that are core within RLUK have an even stronger likelihood of being core within ARL. The latter results suggest that there might be a “global core” of heavily collected print book publications that manifests across consortial and geographical settings.

Overlap between Core and Rare Segments of ARL and RLUK Print Book Collections



Malpas and Lavoie. 2016.

FIGURE 12. OVERLAP BETWEEN CORE AND RARE SEGMENTS OF ARL AND RLUK PRINT BOOK COLLECTIONS (JANUARY 2016)

Comparison of the RLUK core segment with the ARL rare segment, and vice versa, also yields interesting results. Nearly 67,000 publications were found in both the RLUK rare segment and the ARL core, indicating a small corpus of books that are lightly collected within RLUK, yet heavily collected within ARL. Conversely, 323 publications were in both the RLUK core and the ARL rare segment, i.e., heavily collected within RLUK, yet lightly collected within ARL. Table 9 summarizes the comparative overlap of the ARL and RLUK print book collections in their entirety and for the rare and core segments of each.

TABLE 9. RLUK COLLECTIVE PRINT BOOK COLLECTION: KEY OVERLAPS WITH ARL

Overlap	Number of Publications	Percent of RLUK Corpus
RLUK/ARL	8,764,942	42%
RLUK Rare/ARL Rare	6,989,687	35%
RLUK Core/ARL Core	2,785	51%
RLUK Rare/ARL Core	66,740	< 1%
RLUK Core/ARL Rare	323	6%

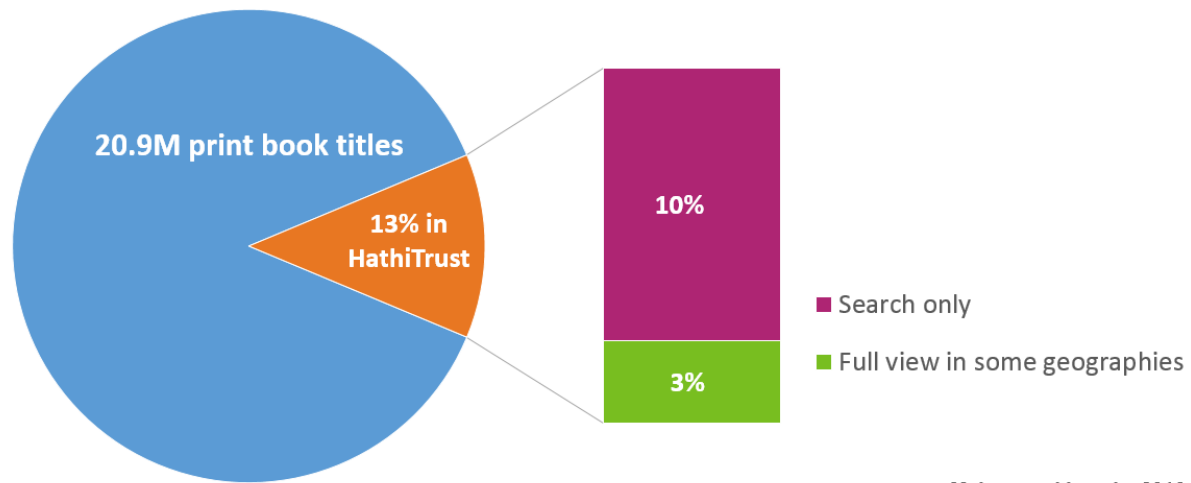
These results suggest that while there may indeed be a global core of sorts—print book publications that are collected heavily across all consortial/geographical contexts—the in-group core is at least partially idiosyncratic to the group itself. We see this in the extreme with the rare/core comparisons between the RLUK and ARL collections. What is core in one collective collection may have only a negligible presence in another. An interesting implication of this pattern is that a national monograph strategy focused on resources that are distinctive to UK (or US) university libraries might be usefully supplemented with super-consortial agreements to preserve content representative of the global core, which is valued by research libraries everywhere.

OVERLAP WITH HATHITRUST

Another important context against which to compare the RLUK collective print book collection is the HathiTrust Digital Library, a large corpus of digital surrogates built from the print book collections of research universities in several countries. Availability of a digital surrogate in HathiTrust (or another digital repository) can be an important factor in local print retention decisions. We used a January 2016 snapshot of the HathiTrust Digital Library to measure duplication between the RLUK print book collection and the HathiTrust collection. At the time of our analysis, the HathiTrust Library encompassed 6.8 million titles, of which 2.2 million (32%) were designated as public domain.

Comparison of the RLUK print book resource with the contents of the HathiTrust repository reveals that 13% of the print book publications in the RLUK collective collection is duplicated in HathiTrust. This is comparable to the overlap between HathiTrust and the ARL print book collection, which we calculated to be 15%. The majority of the titles in the overlap for both groups are humanities-related, in keeping with the overall predominance of humanities-related materials in HathiTrust as a whole. Significantly, the global library holdings profile for RLUK titles duplicated in HathiTrust is quite high, with an average of 114 WorldCat holdings per title. By comparison, global holdings on ARL print book titles duplicated in HathiTrust are 36% lower, with an average of 72 holdings per title.

Duplication of RLUK Print Books in HathiTrust Digital Library



Malpas and Lavoie. 2016.

FIGURE 13. DUPLICATION OF RLUK PRINT BOOK TITLES IN HATHITRUST DIGITAL LIBRARY (JANUARY 2016)

More than three-quarters of the RLUK print book titles duplicated in HathiTrust are in copyright or subject to other rights restriction, while 22% are either in the public domain or viewable as full-text in some geographies. This suggests that a very small fraction of the overall RLUK print book resource is available without restriction through the HathiTrust library. Although HathiTrust is but one repository, it is a significant corpus of digitized monographs, and its relatively small overlap with the RLUK print book resource suggests that ample scope for further digitization exists across the RLUK print book collections. The fact that the vast majority of the overlap with HathiTrust encompasses materials that are still under some form of copyright or use restriction serves as a reminder that encumbrances on access and use are likely to be an important issue with future digitization initiatives: recall that the vast majority of the RLUK print book resource is likely still in copyright, with more than 60% published during or after 1950.

Beyond providing an estimate of potential availability of content in digital format, quantifying the overlap between the RLUK collective collection and HathiTrust may be important to assessing print preservation risks. As part of its Shared Print Monograph Archiving program, HathiTrust is developing plans for a distributed print archive of monographic publications (i.e. books) that have already been digitized.²¹ Hence the duplication rate between RLUK and HathiTrust is a measure of the potential value to UK libraries of print archiving efforts in the US.

Regional Analysis

In the context of ongoing discussion of a UK national monograph strategy, it is useful to consider the distinctive characteristics of the RLUK print book resource held in different UK countries. To what degree does the print book collection of RLUK libraries in Scotland resemble the aggregate resource of RLUK members in England? Could cooperative preservation and access agreements be struck amongst subsets of RLUK members in these regions, such that a greater share of the whole is secured for future generations of students and researchers? For print books especially, it is important to acknowledge that traditional library logistics—moving physical inventory from shelf to user—is a limiting factor in the optimal scale of cooperative management. The fact that a book exists somewhere in the system is no guarantee of value without a robust and relatively frictionless

fulfillment service to bring the book to its reader. Put another way, the appropriate scale of collaboration may not be dictated by the size of the group but by the local and regional contexts in which its members are embedded.

The England-based subset of the RLUK print book resource consists of 18.5 million distinct print book publications held by RLUK members located in England, constituting 88% coverage of the overall RLUK collection and accounting for 77% of all RLUK print book holdings. The extensive coverage of the RLUK print book resource supplied by the England-based collection is unsurprising, given that the vast majority of RLUK members are located there. But it does suggest that most of the RLUK print book resource can be accessed through the collections of RLUK members located in England. However, it is important to note that nearly a quarter of the RLUK collective print book collection is *not* available in England, which speaks to the importance of a UK-scale cooperative effort (supplemented by Ireland's Trinity College Dublin) to secure the full extent of the RLUK legacy print book investment. As with the overall RLUK print book resource, the publications in the England-based segment are, on average, held quite widely throughout the global library system, with an average of about 37 global holdings per publication. Forty-five percent of the England-based collection overlaps with the ARL collective collection, closely approximating the overall RLUK/ARL print book overlap (42%).

Since only a few RLUK members are located in Scotland, the Scotland-based collection is much smaller than its England counterpart, consisting of 5.9 million distinct print book publications. The Scotland-based collection constitutes 28% coverage of the overall RLUK collective print book collection, and accounts for 15% of all RLUK print book holdings. Apart from the disparity in size, several key differences between the England- and Scotland-based collections are worth noting. The publications in the Scotland-based collection are, on average, more heavily collected in the global library system, with an average of 65 global holdings per publication, compared to 37 for the publications in the England-based collection. Also, the Scotland-based collection has a higher percentage overlap with the ARL collective collection than the England-based collection—51% compared to 45%. In short, the Scotland-based collection aligns more closely with collecting decisions made within a peer consortium (ARL) and the global library system as a whole. This suggests that RLUK libraries in Scotland may have a relatively greater degree of freedom in identifying potential partners for cooperative stewardship beyond the RLUK group, while RLUK libraries in England will need to sort out preservation responsibilities within the boundaries of the national monograph collection.²²

The England- and Scotland-based collective print book collections are especially interesting because they both include the print book holdings of a national library: the British Library (which is the national library of the United Kingdom) and the National Library of Scotland (a research library with a specialization in Scotland-related materials), respectively. It is important to emphasize that the inclusion of the British Library's holdings in the England-based collection, and the National Library of Scotland's holdings in the Scotland-based collection, is based strictly on the location of each institution, and has nothing to do with the scope of their remit.

While the British Library and the National Library of Scotland share some characteristics with other RLUK members, in other respects—such as mission, collecting strategy and preservation responsibilities—they are quite different. Consequently, it is instructive to consider the impact on the England- and Scotland-based collections when the holdings of the British Library and the National Library of Scotland are removed.

Removing the British Library's holdings reduces the England-based collection to 10.7 million print book publications, a reduction of 42%. Similarly, excluding the National Library of Scotland's holdings reduces the Scotland-based collection to 2.7 million print book publications, a reduction of 54%. Table 10 summarizes the coverage of print book publications and holdings for the England- and Scotland-

based collections, as well as overlap with the ARL collection, with and without the holdings of the national libraries. Note that for consistency, Library of Congress holdings were removed from the ARL print book collection when comparing to the England- and Scotland-based collections with British Library and National Library of Scotland holdings removed.

TABLE 10. RLUK COLLECTIVE PRINT BOOK COLLECTION: ENGLAND AND SCOTLAND SEGMENTS

	England	England (no BL; no LOC)	Scotland	Scotland (no NLS; no LOC)
Publications	18,454,718	10,710,887	5,872,838	2,704,664
Coverage of RLUK collection	88%	51%	28%	13%
Percent of RLUK print book holdings	77%	44%	15%	7%
Percent overlapping with ARL print book collection	45%	61%	51%	67%

Clearly, both national libraries are important contributors to the print book resources in the regions in which they are located. More significantly, however, both national libraries also appear to be important contributors to the *distinctiveness* of their respective regional collections, both at the peer consortial level (ARL), and at the level of the global library system. For example, when British Library holdings are excluded, the England-based collection’s overlap with the ARL collective collection increases from 45% to 61%; similarly, when the National Library of Scotland’s holdings are excluded, the Scotland-based collection’s overlap with ARL increases from 51% to 67%. Similar results are obtained in regard to the global collective collection when national library holdings are removed: in the England-based collection, average global holdings per publication rise from 37 to 54; for Scotland, from 65 to 99. These results highlight the important and distinctive contributions that the British Library and the National Library of Scotland provide not just to the print book resource in the regions in which they are located, but to the RLUK collective collection as a whole.

CONCLUDING OBSERVATIONS

The preceding analysis describes the characteristics of the RLUK collective collection. It is useful to conclude by placing those characteristics in a broader perspective. In 2014, OCLC Research published *Right-scaling Stewardship: A Multi-scalar Perspective on Cooperative Print Management*,²³ which describes the collective print book collection of the Committee on Institutional Cooperation (CIC) membership, a consortium of research universities located chiefly in the American Midwest. The *Right-scaling Stewardship* report proposes some general observations, derived from the analysis of the CIC collective collection, that are likely to be applicable in most circumstances involving collective collections. This section considers the findings from the RLUK study in light of these observations. Since the observations were obtained from an analysis of the CIC print book resource—and, indeed, most of our collective collection work to date focuses on print books—attention is restricted to print books in this section as well.

SCALE ADDS SCOPE AND DEPTH

A pervasive finding in nearly all of the collective collections work conducted by OCLC Research is that local collections tend to be more distinct from one another than *a priori* intuition would predict. When comparisons are made at the publication (i.e., FRBR manifestation) level, surprisingly little overlap is found across collections. Elevating the unit of analysis to the work level reduces the cross-collection distinctiveness somewhat, but what remains is still quite significant. The relatively small degree of overlap across academic library collections is encountered repeatedly in collective collection studies, leading to the paradoxical conclusion that “rareness is common.” Moreover, this property appears to manifest in collective collections involving a small number of institutions, and those that include many. In all cases, as individual collections are aggregated (and duplicate holdings removed), a rich long-tail of scarce or rare publications is built out. In this sense, moving from local to collective scale adds scope and depth to the resulting aggregate resource, and, significantly, this scope and depth tends to be a far more prominent feature of the collection in comparison to consolidation through reduced redundancy.

The observation that scale adds scope and depth is certainly borne out in the RLUK collective print book collection. As we have seen, the vast majority—nearly 90%—of the 20.9 million distinct print book publications in the RLUK collection are held by fewer than five RLUK members; the median holdings per publication is only one. In contrast, less than 1% of the collection is held by more than 20 RLUK members. The low level of overlap across RLUK collections is the primary reason that the aggregate resource is so large: even with consolidation through removal of duplicate holdings, the collective collection is of sufficient dimensions to average a contribution of more than 650,000 publications per RLUK member. From these contributions, a collective resource of deep, rich diversity is formed: as the preceding analysis indicates, the RLUK print book resource encompasses an extremely wide diversity of subjects, languages and countries of publication that transcends the collection of any single RLUK member.

Key Implication: Elevating local collections above the institution into a collective collection perspective is not just about finding opportunities to consolidate holdings, release space and remove redundancy. While these are certainly important considerations, it should not be forgotten that collective collections are equally—perhaps even primarily—about building a collective resource that leverages the individual strengths of its component local collections. As RLUK members consider the analysis of their collective print book holdings presented in this report, they will hopefully find opportunities to preserve and amplify each member collection’s distinctive features within the context of cooperative arrangements built around the collective resource.

COVERAGE REQUIRES COOPERATION

The distinctiveness of academic library print book collections—the fact that rareness is common and embeds a prominent long tail in most collective collections—has important consequences for strategies aimed at managing the collective print book holdings of a group of institutions. A recurring theme in the collections analysis conducted by OCLC Research has been that collective collections are indeed *collective*—each local collection supplies a contribution to the collective collection that is not duplicated elsewhere in the group. This has important ramifications in regard to stewardship of the collective resource, in particular because it suggests that a cooperative effort to manage a group-scale collective collection cannot be confined to the institutions with the largest collections. Put another way, it is a mistake to assume that the collective collection of a particular group of institutions is fully subsumed within—or even approximated by—a small subset of collections within the group, with the remaining collections offering nothing more than duplicate holdings. Rather, full coverage of the collective collection requires full group participation in the stewardship effort.

The regional analysis of RLUK print book holdings illustrates this point well. As we saw, both England and Scotland each include a national library with a large print book collection within its geographical boundaries. Yet neither the British Library nor the National Library of Scotland provides full coverage of the portion of the RLUK print book resource located in England or Scotland, respectively. Similarly, neither the England-based collection nor the Scotland-based collection are subsumed within one another: the England-based collection is much larger than the Scotland-based collection, so clearly is not subsumed by the latter, while a third of the Scotland-based collection is not held by institutions located in England. Finally, neither the England- nor the Scotland-based collection covers the full extent of the RLUK print book resource. In short, neither the British Library nor the National Library of Scotland can cover the full extent of the print book resource located in their respective regions; neither regional collection is covered by the other regional collection; and neither regional collection can cover, or even closely approximate, the full RLUK print book resource.

Key implication: Collective collections are collective in two senses: in their construction and in their curation. In most group settings, each participating institution can supply a unique contribution from its local holdings to the collective resource. The corollary to this, however, is that no single institution or subset of institutions in the group can supply coverage sufficient to replicate, or even closely approximate, the collective resources of the full group. As a consequence, cooperation must be collective in managing as well as building the collection; moreover, the scale of cooperation must grow as the scale of the collective collection grows. In the same way, securing the long-term future of the 20.9 million distinct print book publications held across RLUK institutions today will likely require the collective efforts of not just a few, but all of the RLUK members.

SCARCITY IS RELATIVE

An important property of collective collections is that scarcity is intrinsically linked to scale, a finding encountered in a variety of collective collections studies. Publications that appear relatively scarce in the context of an academic consortium could have a very different profile when considered at higher scales of aggregation, e.g., regional, national or even global. While this result is hardly surprising—a frame of reference that includes more libraries is likely to increase the number of holdings for most publications—it nevertheless highlights a key decision-making parameter surrounding collective collections: what is the relevant scale to inform choices on print retention, identification of last copies, minimum levels of availability and other aspects of cooperative print management? In other words, should these issues be addressed strictly on the basis of in-group resources or are higher scales relevant as well?

The linkage between scarcity and scale is evident within the RLUK collective print book collection. Nearly 90% of the RLUK collection is held by fewer than five RLUK members, compared to less than 1% of the collection held by more than 20 libraries. Yet only 56% of these same print book publications are held by fewer than five libraries globally, with nearly 25% held by 25 libraries or more. Similarly, more than 48 million RLUK in-group holdings are attached to the RLUK print book collection; these holdings expand to more than 695 million global holdings on these publications—an increase by a factor of 14. The RLUK membership averages two holdings per publication for the print books in the RLUK collection; the global library system averages 33 holdings per publication for these same publications. It is clear that the degree of scarcity found when the RLUK print book resource is examined at consortial scale (i.e., the RLUK membership) is quite different from what is found when that same collective collection is placed in a different frame of reference (the global library system).

Key implication: The evidence demonstrating that scarcity is a function of scale highlights the importance of determining the relevant frame of reference for decision-making around collective collections. Should the context in which the collective collection is placed be ring-fenced by group membership—i.e., with only the in-group perspective relevant for decision-making—or should it be elevated to higher scales, with a recognition that the collective collection is embedded in a broader *system* of collective collections that needs to be accounted for as in-group policy is developed? As the RLUK membership considers opportunities for deeper collaboration around print management and other areas of mutual interest, an important element of that process will be determining when inclusion of external perspectives—other consortial collective collections, regional collections, national collections or even the global collective collection—is a necessary ingredient for effective policymaking regarding the RLUK collective resource.

* * *

The RLUK collective collection described in this report is purely notional, manifested only in aggregated library data. It is an analytical construct that provides a high-level view of a collective resource that is, in reality, scattered across 32 research libraries. But while the collective collection itself may exist only as an abstraction, the benefits from analyzing it are intended to be wholly practical: to inform decision-making aimed at advancing RLUK community interests around their collective holdings. In a recent blog post, OCLC President and CEO Skip Prichard discusses the benefits of analyzing “big data,” noting that data leads to insight and insight leads to action.²⁴ That is precisely how analysis of collective collections can benefit libraries. WorldCat is a big data resource uniquely positioned to create aggregated views of library collections at any scale. That, in turn, provides insight that helps libraries take practical action to address areas of need around their collective holdings. In doing so, many of these libraries will bring their notional collective collections to practical fruition.

In closing, we offer these practical conclusions for consideration by RLUK libraries as they begin to develop and operationalize a shared strategy for managing print books:

- The collective print book collection in RLUK libraries represents a major asset for the UK higher education community as a whole. Stewardship efforts to secure this resource will deliver benefit to generations of students and scholars in the UK and can be an important differentiator of the UK educational and research offer *if it is managed as a collective asset*. As noted above, the relative scarcity of print book holdings in the UK means that preservation will require collaboration on a large scale.
- Selection and prioritization for preservation can be usefully informed by an understanding of the authors and works (in addition to editions) that are best represented in the aggregate resource. This is content that is demonstrably important to the research library community as

a whole; many libraries will benefit from some rationalization of collective holdings so that the scope of the total resource can be enlarged and local collecting interests can be specialized according to institutional need.

- An appreciation of individual and shared collection strengths of RLUK libraries can help to sort curatorial alliances along lines that fit with local institutional interests and university research priorities.
- While rareness is common in research collections, there is substantial overlap between the RLUK resource and print books in North American research libraries. This represents a significant opportunity for cross-consortial agreements to leverage collective stewardship interests. As evidenced by a recent joint meeting of major international research library organizations, there is growing interest in above the consortium coordination of preservation and stewardship activities. RLUK is well placed to advance this work through multi-national partnerships with other research library groups.

Three key areas for action were identified in the course of this project. While they are described here in terms specific to RLUK, they are equally relevant to other groups formulating strategy for preservation of print book collections:

- Agree on an appropriate FRBR-informed strategy that balances the interest in preserving representative specimens of a broad range of works and the interest in providing a complete record of publication history for at least some titles. For example, there may be greater interest in preserving complete editorial runs of works of historical material, acknowledging that new editions of more recent scholarship may not be represented (or best preserved) as print books.
- Concentrate some attention to normalization of cataloging practice within the RLUK so that measures of distinctive holdings are less likely to be affected by artifacts in metadata. This is an area where further collaboration between OCLC and RLUK partners will be mutually beneficial.
- Improve awareness of core data operations in the union catalog environment as these have a direct bearing on how library holdings are aggregated for display and use in management systems. It is important that participating libraries understand policies and procedures that affect the matching and merging of contributed bibliographic and holdings data, and that the aggregator (OCLC, in this instance) promote awareness of how those policies are established and evolve in consultation with a global community. One area of further work will be the creation of more robust documentation of data ingest, management and export options for new contributors to WorldCat.

AREAS FOR FURTHER EXPLORATION

This report represents the culmination of a narrowly scoped research collaboration between RLUK and OCLC focused on developing a shared understanding of the collective collection of RLUK libraries as they are represented in WorldCat. Predictably, the research analysis and community consultation identified a number of potentially interesting and useful lines of inquiry for further exploration. We identify a few of them here in the interest of stimulating additional research:

- Given the distinctive role of the five UK deposit libraries as stewards of the UK publication record, it could be instructive to examine whether there is a core of materials that is duplicated in both legal deposit and other RLUK libraries, potentially relieving print management pressures in non-deposit libraries. We explored this in a preliminary fashion and

found that the vast majority of “scarce” resources (publications held in fewer than five RLUK libraries) are represented in one or more legal deposit collections; however, we did not test duplication rates in legal deposit for publications that are widely held by RLUK libraries. More detailed analysis of the legal deposit collections may be particularly helpful in the context of the National Monograph Strategy.

- For reasons of convenience, the collective collection of ARL libraries was used as a benchmark against which the RLUK collective collection could be compared. From an operational perspective, RLUK institutions are likely (in the near term, at least) to regard libraries in the UK or European continent as potential partners for managing national and regional print collections. Further research could explore the potential benefits of coordinating curatorial activities in RLUK and other UK and European library groups.

APPENDICES

APPENDIX 1: RLUK LIBRARIES AND WORLDCAT SYMBOLS

Analyses in this report are based on holdings contributed to WorldCat by the following libraries as of January 2016.

Library	Symbol	WorldCat Holdings
British Library, Document Supply	BRI	4,578,408
British Library, St. Pancras	BLSTP	9,848,579
British Library, Boston Spa	UKM	6,249,535
British Library, Boston Spa	BLX	129,948
Cambridge University Library	CUD	4,455,878
Cardiff University Library	RDF	725,600
Durham University Library	DHA	909,950
Imperial College London Library	LIP	376,768
King's College London Library	KIJ	1,193,103
London School of Economics and Political Science Library	LSD	857,316
National Library of Scotland	NLE	4,815,576
Newcastle University Library	EUN	509,930
Queen Mary, University of London Library	MM9	295,171
Sir Duncan Rice Library, University of Aberdeen	ADU	1,146,683
SOAS, University of London Library	LOA	781,572
Trinity College Library Dublin Library	ERD	3,307,773
University College London Library	LUN	1,518,903
University of Birmingham Library	SXB	997,109
University of Bristol Library	BUB	782,124
University of Edinburgh Library	EUX	1,806,604
University of Exeter Library	EZ9	651,738

Library	Symbol	WorldCat Holdings
University of Glasgow Library	QCL	1,278,805
University of Leeds Library	ERL	1,074,402
University of Liverpool Library	LVT	1,107,991
University of London Senate House Libraries	UEJ	709,454
University of Manchester Library	EUM	1,116,806
University of Nottingham Library	UVN	125,299
University of Oxford Libraries	EQO	5,833,178
University of Sheffield Library	SHS	752,703
University of Southampton Library	S2H	24,122
University of St Andrews Library	AUD	853,596
University of Warwick Library	U3W	826,725
University of York Library	UKUOY	649,213
Victoria and Albert Museum National Art Library	AVA	667,423
Wellcome Library	EUW	459,452

APPENDIX 2: RLUK ADVISORY GROUP MEMBERS

Institution	Name	Position
King's College London	Victoria Parkinson	Metadata Coordinator
National Library of Scotland	Gill Hamilton	Digital Access Manager
RLUK	David Prosser	Executive Director
University of Birmingham	Frances Machell	Head of Collection Management
University of Bristol	Suzy Cheeke	Collections Librarian
University of Edinburgh	Alasdair MacDonald	Metadata Coordinator
	Hannah Mateer	Collections Lifecycle Manager
	Laura Shanahan	Head of Collections Development and Access
University of Glasgow	Rosemary Stenson	Head of Cataloguing
University of Manchester	Sandra Bracegirdle	Head of Collection Management
University of Oxford	Michael Williams	Head of Storage & Logistics, Bodleian Libraries
University of St Andrews	Janet Aucock	Head of Metadata and Content Acquisition
	Helen Faulds	Collections Manager
University of York	Ruth Elder	Library Collection Space Manager
	Sue Elphinstone	Collection Development Manager
	Sarah Thompson	Head of Collections
Wellcome Library	June Tomlinson	Cataloguing and Metadata Services Manager

APPENDIX 3: EDITIONS MOST WIDELY HELD IN RLUK COLLECTIVE COLLECTION

Listed here are the publications with the most edition-level holdings in the RLUK collective collection. As of January 2016, a total of 31 RLUK library symbols (holdings) were set on each of the OCLC numbers listed below.

OCLC number	Title	Date of most widely held edition in RLUK	Versions in RLUK Collective Collection	Aggregate RLUK holdings for all versions
24009791	E B Fryde et al. <i>Handbook of British Chronology.</i>	1986	11	97
447768	Christopher Hill. <i>The World Turned Upside Down: Radical Ideas During the English Revolution.</i>	1972	10	76
1884048	Mary Douglas. <i>Implicit Meanings: Essays in Anthropology.</i>	1975	10	61
8430152	Julia Kristeva. <i>Powers of Horror: An Essay on Abjection.</i>	1982	6	56
12052097	David Lowenthal. <i>The Past is a Foreign Country.</i>	1985	5	38
24847759	Hal Varian. <i>Microeconomic Analysis.</i>	1992	5	71
21375348	Thomas Laqueur. <i>Making Sex: Body and Gender from the Greeks to Freud.</i>	1990	4	36
22597117	Elazar Barkan. <i>The Retreat of Scientific Racism: Changing Concepts of Race In Britain and the United States Between the World Wars.</i>	1992	4	35
23869294	Terence Ranger and Paul Slack (Eds.) <i>Epidemics and Ideas: Essays on the Historical Perception of Pestilence</i>	1992	4	37
24373115	L D Schwarz. <i>London in the Age of Industrialisation: Entrepreneurs, Labour Force, and Living Conditions, 1700-1850.</i>	1993	4	36
25631046	Patrick Karl O'Brien; Roland E Quinault (Eds.). <i>The Industrial Revolution and British Society.</i>	1993	4	37
26807179	Malcolm Barber. <i>The New Knighthood: A History of the Order of the Temple.</i>	1994	4	44

OCLC number	Title	Date of most widely held edition in RLUK	Versions in RLUK Collective Collection	Aggregate RLUK holdings for all versions
33838719	David Held. <i>Democracy and the Global Order: From the Modern State to Cosmopolitan Governance.</i>	1995	4	36
20356956	E A Grosz. <i>Jacques Lacan: A Feminist Introduction.</i>	1990	3	35
23255546	Terry Eagleton. <i>Ideology: an Introduction.</i>	1991	3	49
26851176	Averil Cameron. <i>The Mediterranean World in Late Antiquity, AD 395-600.</i>	1993	3	37
10273619	P N Johnson-Laird. <i>Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness.</i>	1983	2	34
16681349	Robert Lumley. <i>The Museum Time-machine: Putting Cultures on Display.</i>	1988	2	32
16681707	Lynda Nead. <i>Myths of Sexuality: Representations of Women in Victorian Britain.</i>	1988	2	34
20758938	Felicity Heal. <i>Hospitality in early modern England.</i>	1990	2	38
3016556	P N Johnson-Laird; Peter Cathcart Wason (Eds.). <i>Thinking: Readings in Cognitive Science.</i>	1977	1	31
6447296	Daniel R Headrick. <i>The Tools of Empire: Technology and European Imperialism in the Nineteenth Century.</i>	1981	1	31
12972377	Stephen M Stigler. <i>The History of Statistics: The Measurement of Uncertainty Before 1900.</i>	1986	1	31
14188031	Catherine Gallagher; Thomas Walter Laqueur (Eds). <i>The Making of the Modern Body: Sexuality and Society in the Nineteenth Century.</i>	1987	1	31

OCLC number	Title	Date of most widely held edition in RLUK	Versions in RLUK Collective Collection	Aggregate RLUK holdings for all versions
26319151	Donella H Meadows, Dennis L Meadows, and Jørgen Randers. <i>Beyond The Limits: Global Collapse or a Sustainable Future.</i>	1992	1	31
34046260	Ivan Hannaford. <i>Race: The History of an Idea in the West.</i>	1996	1	31
36135824	Barbara Kirshenblatt-Gimblett. <i>Destination Culture: Tourism, Museums, and Heritage.</i>	1998	1	31
24009791	E B Fryde et al. <i>Handbook of British Chronology.</i>	1986	11	97
447768	Christopher Hill. <i>The World Turned Upside Down: Radical Ideas During The English Revolution.</i>	1972	10	76

NOTES

1. Dempsey, Lorcan, Constance Malpas, and Brian Lavoie. 2014. "Collection Directions: The Evolution of Library Collections and Collecting" *portal: Libraries and the Academy* 14,3 (July): 393-423 (p. 37). OCLC Research Preprint: <http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-collection-directions-preprint-2014.pdf>.
2. Dempsey, Lorcan, Brian Lavoie, Constance Malpas, Lynn Silipigni Connaway, Roger C. Schonfeld, JD Shipengrover, and Günter Waibel. 2013. *Understanding the Collective Collection: Towards a System-wide Perspective on Library Print Collections*. Dublin, Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2013/2013-09.pdf>.
3. It is important to note that the aggregate resource of RLUK libraries has not been developed or managed as a shared collection, so any patterns that are revealed by analysis reflect the cumulative effect of individual institutional decisions. With the exception of a few legal deposit institutions (the British Library, Bodleian Libraries of the University of Oxford, Cambridge University Library, the National Library of Scotland, and the Library of Trinity College, Dublin) that collectively uphold responsibility for preserving publications produced in the UK, the collecting interests and priorities of individual RLUK libraries are dictated by institutional need. We also note that while our analysis treats RLUK member collections monolithically, important distinctions within collections may exist in practice: for example, portions of the collection that primarily support research and those that primarily support teaching and learning. It is not possible to reflect these distinctions in our data, but we acknowledge that they often have varying implications for decision-making around print retention, digitization and other collection management practices.
4. Sector-wide interest in the aggregate UK book resource is reflected in recent efforts to elaborate a National Monograph Strategy for UK higher education. See <https://www.jisc.ac.uk/rd/projects/national-monograph-strategy>.
5. These 32 RLUK institutions are represented by 35 OCLC library symbols in WorldCat (see appendix 1).
6. For more information on FAST, as well as several FAST tools and interfaces, see <http://www.oclc.org/research/themes/data-science/fast.html>.
7. The MARC21 Cataloging Standard includes 380 official country codes, including 48 discontinued codes and one code used for titles where the place of publication is not identified. See: https://www.loc.gov/marc/countries/countries_code.html. The RLUK data set extracted from WorldCat includes 333 valid country codes and a single invalid code (for one title). A small number of titles, representing .01% of the collective collection, lacked any place of publication data. It is not unusual for country codes to include some typographical errors, which may result in inaccurate counts of geographic representation. For instance, the single title published in Saint Kitts was erroneously coded as a publication from Saint Pierre and Miquelon, an archipelago 2,000 miles north of the Antilles. No attempt was made in this study to validate place of publication data from bytes 15-17 in the 008 against geographic information elsewhere in the bibliographic record.
8. Some of the apparent gaps in descriptive cataloging of the RLUK collective collection in WorldCat may be due to imperfect or incomplete conversion of records from UKMARC to MARC21 format.

9. Of the 515 authorized MARC21 language codes, 476 (92%) are represented in the RLUK Collective Collection. A complete list of MARC21 language codes can be found here: https://www.loc.gov/marc/languages/language_code.html. Twenty titles in the collection lacked any language of publication data.
10. See <http://www.ukrr.ac.uk>.
11. It is not clear if the greater proportion of pre-1850 titles in RLUK libraries compared to ARL libraries reflects the greater age of some of the UK universities and a longer history of collecting activity or is merely an artifact of cataloging backlogs or less comprehensive record-loading.
12. There is a fourth dip in 2010–2016, but the data for this decade is obviously incomplete.
13. Calhoun, Karen, and Glenn Patton. 2011. *WorldCat Quality: An OCLC Report*. Dublin, Ohio: OCLC Online Computer Library Center, Inc. http://www.oclc.org/content/dam/oclc/reports/worldcatquality/214660usb_WorldCat_Quality.pdf.
14. In some instances what appears to be a work with a single edition is merely an artifact of local cataloging practices. For instance, there is a single OCLC accession number (636229098) and single OCLC Work identifier (499672520) associated with a seven-page publication by Herbert E. Roaf (1906) titled "A Contribution to the Study of the Digestive Gland in Mollusca and Decapod Crustacea," In *Biochemical Journal* 1(8-9):390-7, with a single holding at University of St Andrews Library. A notation in the bibliographic record indicates that this one of numerous scientific papers bound in a single volume. We identified nearly 26,000 Work identifiers associated with articles or other bibliographic analytics in the RLUK collective collection, representing less than 1% of the aggregate resource.
15. Excluded from this list are several important religious works (including the Bible and the *Book of Common Prayer*) and government publications that are widely collected by RLUK institutions but which seemed to be less representative of independent library selection patterns than informal or, for legal depository libraries, official deposit mandates.
16. For example, a list of the best-represented works in Dutch university libraries features classics from the canon of Dutch literature alongside global favorites like the *Divine Comedy*. In a prior study of the North American print book collection, we also found that regional collections have a distinctive regional character emphasizing events, people and places of local significance. See Lavoie, Brian, Constance Malpas, and JD Shipengrover. 2012. *Print Management at "Megascale": 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>. (p. 33.)
17. It is interesting to note that the most widely held editions in the RLUK collective collection were all produced between 1970 and 1998. Whether this reflects a historical trend in scholarship (lines of inquiry or interpretation that were compelling to broad interdisciplinary audiences) or an artifact of library budgets (pre-dating the shift in acquisition resources to journal subscriptions) and acquisition strategies (more attention to benchmarking and competitive collection development) is unclear and would require research that is beyond the scope of this study.
18. See <https://viaf.org/>.
19. A group of SOAS researchers and faculty members has, for example, produced a bibliography of Hausa popular literature. Interestingly, the bibliography is based on the personal collection of a faculty member (Professor Graham Furniss, OBE) rather than SOAS library holdings, see "Hausa Popular Literature Database." <https://www.soas.ac.uk/africa/research/hausadb/>.
20. Malpas, Constance, and Brian Lavoie. 2014. *Right-scaling Stewardship: A Multi-scale Perspective on Cooperative Print Management*. Dublin, Ohio: OCLC Research. <http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearchcooperative-print-management-2014.pdf>.

21. Information on the HathiTrust Shared Print Monograph Archiving program is here https://www.hathitrust.org/print_monograph_archiving.
22. It is worth noting that university and research libraries in Scotland are exploring Scotland-scale print preservation and management models alongside UK-wide efforts. For example, the Scottish Confederation of University and Research Libraries (SCURL), the National Library of Scotland (NLS) and the Scottish Library and Information Council (SLIC) have developed a Scottish Collections Policy aimed at ensuring that at least one copy of certain classes of material is retained in Scotland as part of a national heritage collection. See <http://scurl.ac.uk/what-we-do/publications/scottish-collections-policy/>.
23. Available at: <http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-cooperative-print-management-2014.pdf>.
24. "Transforming Data into Impact" *Next* (blog). Posted 8 February 2016. <http://www.oclc.org/blog/main/transforming-data-into-impact/>.