Print Archiving and Shared Print in North America: A Preliminary Analysis and Status Report

September 10, 2015

Center for Research Libraries

Background

The present report was prepared as the basis for discussion at the June 25, 2015, meeting, Preserving America's Print Resources II: a North American Summit. The PAPR II summit was held to consider the strength of the safety net that North American libraries have created over the past twelve years through archived and shared print efforts, and to promote a common vision for the further development and growth of those efforts. The summit was also intended to help CRL determine its priorities for supporting the archiving and sharing of print serial collections.

In July 2003 CRL hosted the Preserving America's Print Resources (PAPR) conference, which explored the challenges faced by libraries in sustaining print collections at a time when digital resources were fast becoming the preferred sources for researchers. Since 2003, CRL has put in place a number of activities and resources to support the preservation of print serials.

- It has sponsored the semi-annual Print Archives Network Forum, a venue for the sharing and discussion of ideas and best practices for shared print and print archiving.
- In 2010 CRL formed the Global Resources Law Partnership with the Law Library Microform Consortium, which brings together a combined CRL and LLMC membership of over 400 North American libraries to digitize and preserve critical primary legal publications from the U.S., Canada, and all world regions.
- CRL created a partnership with the Linda Hall Library of Science, Engineering and Technology in 2012, to preserve, provide access to, and further develop Linda Hall's extensive collections of current and historical print serials.
- Also in 2012, in cooperation with the California Digital Library and with support from the Andrew W. Mellon Foundation and the Institute of Museum and Library Services, CRL brought online the Print Archives and Preservation Registry, a database of holdings, terms of service, and conditions of archiving of institutions committed to the preservation of print serials.
- At about the same time, CRL developed and brought online the International Coalition of Newspapers (ICON) database, a registry of print, microform and digital holdings of published newspapers.

Print sharing and archiving efforts launched since 2003 by various organizations and consortia have also gained traction in the U.S. and Canada. Yet, as the center of gravity for libraries continues to shift from print to digital, the need for rational and sustainable provisions for print is growing.

The present report represents an examination of the results of our collective efforts to date, based on the copious, but still incomplete and inconsistent, data on archiving gathered by CRL in the past few years. The analysis of this data, presented here, identifies in broad terms the scope, strengths and gaps in our coverage of print serials to date, and identifies some remaining obstacles to the comprehensiveness and effectiveness of the archiving efforts. The current version of the report has been revised to incorporate comments and information from the PAPR II summit.

Summary Findings of the Analysis

Given the incomplete and inconsistent data available, the findings in this report should be considered somewhat provisional. However, the data we have strongly suggests that much work is yet to be done to prevent the loss of important print serial literature collected and preserved by U.S. and Canadian research libraries over the past century as the focus of library investment shifts toward digital resources. The analysis suggests that the scale and scope of print archiving must be increased radically. Existing efforts now preserve relatively little humanities and social science materials, and pay little attention to materials that are both of high value and at risk. The analysis also suggests that disclosure of more detailed information about archive programs and library commitments is necessary to enable due diligence by libraries that wish to depend on the archives. In the absence of such disclosure it will be difficult for libraries to demonstrate that consequential decisions and actions on management of print are made on an objective, empirical basis.

On the Scope, Strengths and Gaps in the Materials Archived

Specifically, the available data tells us the following about the contents of the registered archives programs:

- ***Only a small portion of the universe of published serials is covered by the existing archiving and shared print programs.***

  Measured in terms of the percentage of serial titles published in the past 100 years, the number of serial titles archived by credible programs is small to insignificant. While millions of volumes are being archived by established print archiving and sharing programs, estimates of the percentage of titles archived range from fifteen percent to less than two percent.

- ***Redundancy among the archives of print serials is rare and sporadic.***

  Of the 71,527 unique titles registered in PAPR, only 13,513 (19%) appear to be held by more than one archiving program. In general, the amount of duplication across all archives is minimal and, ironically, is concentrated on materials that are also available in electronic form.

- ***Science and Technology titles are the most commonly archived.***

  Titles in LC class Q (Science) make up 25% of the titles archived; followed by Technology (T) at 22% and Social Science (H) a distant third at 9%.

- ***The overwhelming majority of titles archived are in English.***

  English-language serials represent 78% of the titles archived in programs registered in PAPR; but only 60% of the titles listed in Ulrich's.

- ***Serials published in the U.S. are more widely archived than non-U.S. titles.***

  Analysis of archived titles by country of publication suggests that titles published in the U.S. are more heavily represented among the archived titles (46.7%) than in Ulrich's (28.9%).

- ***Most shared print and print archiving efforts focus on secondary, rather than primary, literature.***

  Not surprisingly, scholarly and scientific journals dominate the holdings of most print archive and shared print programs, while trade journals, popular serials, and industrial literature are archived by relatively few programs.

- ***Historical and at-risk materials are not well covered.***

  A disproportionate number of the serial titles archived, as a percentage of the total number of serials published in a given year, date from the late twentieth century.

On The State of the North American Archiving Programs in General

Specifically available data on the nature and practices of the registered archive and shared print programs suggests that:

- ***Actionable information about archived holdings is scarce.***

  The information available about the terms of archiving and shared print programs falls short of the granularity and detail needed to support sound risk assessments and decision-making on the preservation, retention and disposition of serial materials.

- ***Actionable information about archive programs is scarce.***

  Information about the commitments made by archiving libraries to verify and maintain archived runs of serials, and information about the measures in place to fulfill those archiving obligations, is often unavailable or lacks sufficient detail to inform the decisions of stakeholder libraries.

- ***Volume is more common than value as a criterion for selecting materials for archiving.***

  Because recovering space occupied by print journals is an imperative for many libraries most programs are designed to achieve coverage of the largest amount of material, in terms of space reclaimed or the number of volumes archived.

- ***Most programs archive materials "in place", as opposed to extracting and isolating archived materials in dedicated environments.***

  Reliance upon extant runs of titles currently in campus stacks or remote storage, rather than assembling and isolating holdings for the purpose of archiving, is a common practice. This is the least expensive and least disruptive approach to creating designated archives of titles.

- ***The programs vary considerably in the benefits and incentives they provide.***

  The benefits archiving programs provide to participating libraries vary from one program to another. A variety of incentives for participation exist, some weaker and shorter-lived than others.

The Data and Its Limitations

Scope of the Data

The main data set for this analysis is a subset of the information on the holdings of print archives registered as of December 2014 in the Print Archives Preservation Registry (PAPR) database. This subset includes the records of the serial holdings of the following North American cooperative print archives and shared print programs:

Program	Number of Titles
Core Historical Literature of Agriculture (CHLA)	35
CIC Shared Print Repository (CIC SPR)	2,340
Council of Prairie and Pacific Libraries' Shared Print Archive Network (COPPUL)	3,530
CRL's JSTOR Print Archive	1,851
Linda Hall Library of Science, Engineering and Technology (LHL)	36,301
Law Library Microform Consortium (Legal PAPR)	5,535
Maine Shared Collections Cooperative (Maine SC)	2,660
Pennsylvania Academic Library Consortium Distributed STM Print Serials Archive Project (PALCI)	55
Scholars Trust	18,131
Western Regional Storage Trust (WEST)	14,602
TOTAL	**85,040**

Holdings registered in PAPR were chosen as the basis for this analysis because they are distinctive in an important respect: they are materials for which identified institutions have formally agreed to archive print copies as part of a recognized shared print or cooperative print archiving program; and for which that program has contributed credible, ingestible holdings records to PAPR. The PAPR registry does not include materials which individual libraries have independently committed to maintain

As points of comparison, CRL also factored into its analysis the following additional data sets:

- Title records for print copies maintained in two JSTOR archives: University of California and Harvard (an estimated 2,000 titles)
- Records for CRL's non-JSTOR print serial holdings from CRL's own catalog (52,212 titles)

- Holdings information on digitized serials included in the HathiTrust Digital Library, extracted from the HathiTrust catalog (an estimated 292,480 titles)

- Holdings information supplied by Portico (16,247 titles) and CLOCKSS (6,257 titles) for serials archived by those repositories.

The analysis also took into account a number of other types of information. Among those were the published or otherwise disclosed policies, terms and conditions under which print serials are archived and shared by programs represented in the PAPR registry. This documentation provides important insights into the nature and duration of the commitments made by participating libraries to retain print volumes, about the terms of service provided, and about how those programs are governed. Copious information of this kind is published on the web by the programs themselves, in the form of program MOUs, policies, bylaws, and so forth.

In addition, the analysis considered a significant amount of information about archive practices and challenges revealed in discussions with archive program managers and staff, and in the course of CRL collection analysis and other data-gathering. In addition, the reports and presentations in the series of semi-annual Print Archive Network forums, held over the last six years, provided a wealth of information about planning, best practices, costs, and accomplishments of the major print sharing and archiving efforts.

Limitations of the Data

Readers should bear in mind that this analysis is based on limited data and inexact science. CRL maintains a high standard for inclusion of holdings information in the PAPR database, requiring that data be consistent and. Therefore holdings data from some notable print archiving programs that might prove credible over time are at present excluded. In general, the print preservation community lacks machine-readable, authoritative and comprehensive lists of serials categorized by discipline and subject area that can be aggregated easily with MARC record identifiers (ISSNs, OCLC numbers) and holdings statements. Because of the resultant lack of standardized title names in many records contributed to PAPR, and despite intensive efforts to reconcile and disambiguate titles, our methodology is afflicted by an inherent over-estimation of the number of titles that have actually been comprehensively archived.

Moreover, the lack of uniformity in the expression of serial holdings hampers analysis and comparison of multiple institutions' holdings of the same titles; and widespread inattention to precision in creating local holdings records, undermines our ability to achieve much certainty about the completeness or continuity of holdings of a given title. Most archiving programs were not able to "deliver" detailed volume-level holdings records, let alone issue-level records.

As the basis for the analysis of the comprehensiveness and scope of archiving, we used ProQuest's Ulrich's database of over 400,000 serial titles (excluding newspaper titles) as a point of reference. Ulrich's is often referenced or used to assist in the work print archiving and collection development. However, as an indicator of the total universe of published serials it is limited, because its focus is primarily on scholarly and commercial literature produced in North America and Europe. It does not take into account the vast vernacular and commercial serial literature of South Asia, China, the Middle East, and other less developed world regions. While it encompasses far fewer than the millions of serial records held in OCLC's WorldCat database, its use by print archiving programs, and widespread use in academic libraries for collection development purposes, lends it

legitimacy as a starting point for analysis. OCLC records for serial titles, a more comprehensive albeit less rigorous aggregation of records, suggests that a reasonable proxy for the size of the corpus mostly likely to be 'in scope' for print journal archiving efforts is closer to 5 million titles.

Ulrich's data was often incomplete, however. It lacked end of publication dates, and beginning publication dates for titles beyond the first title in a family of titles. Moreover, country of publication data listed in Ulrich's did not match data in WorldCat approximately 17% of the time. WorldCat data was substituted for country of publication in Ulrich's, when possible, for analysis. It cannot be emphasized enough that it is impossible to provide consistent, accurate analysis of print archiving unless the source data, that from the programs, is accurate and complete.

Of the 85,040 records in PAPR, 75% had ISSNs. The remaining records all had OCLC numbers. Records with the same ISSN, often had different OCLC numbers, but we did not review each to see if each ISSN was being used correctly or if institutions were just using different OCLC records locally. Title changes were almost certainly masked. OCLC encourages libraries to use a single shared record as appropriate, but WorldCat is full of titles with multiple records. That problem has been transferred to PAPR as libraries register their holdings without reconciling whether their record should be merged into others or if a title change needs to be noted.

To determine the relative concentration of titles archived in various subjects, we relied on the Library of Congress classification scheme, i.e., its published alphabetical class names. Like Ulrich's records, however, the LC subject classification scheme is a blunt instrument: the class names are quite broad and do not map well to lines of current academic research.

In short, therefore, this analysis is a work in progress. Practice is rapidly evolving in the domain of shared print, and the data set continues to grow.

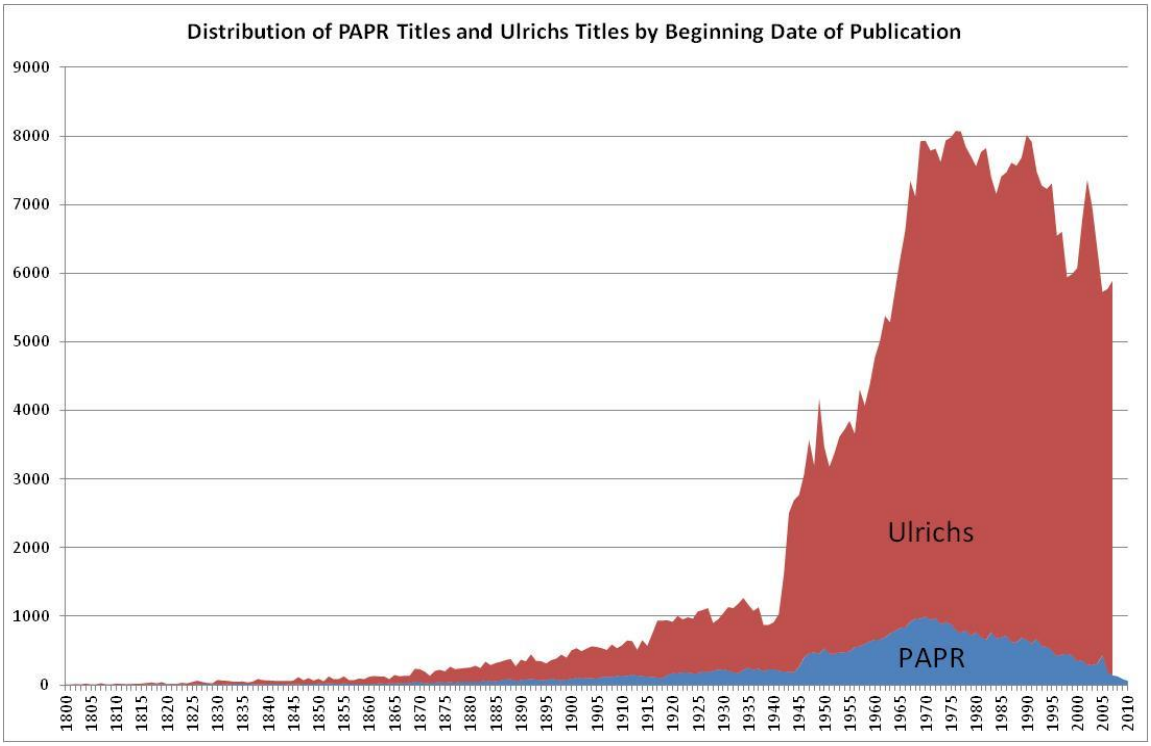
Findings in Detail

I. What the Data Tells Us about the Scope, Strengths and Gaps in the Materials Archived

a) Only a small portion of the universe of published serials is covered by the existing shared print programs.

While the number of volumes archived by the established print archiving and sharing programs is high, the percentage of titles archived is surprisingly low. Fewer than 71,527 (15.5%) of the estimated 460,000 titles listed in Ulrich's are archived by programs registered in the PAPR database. If one accepts the much larger estimate of the number of existing serials based on WorldCat records, 5.5 million titles, the corpus of archived titles shrinks to less than two percent.

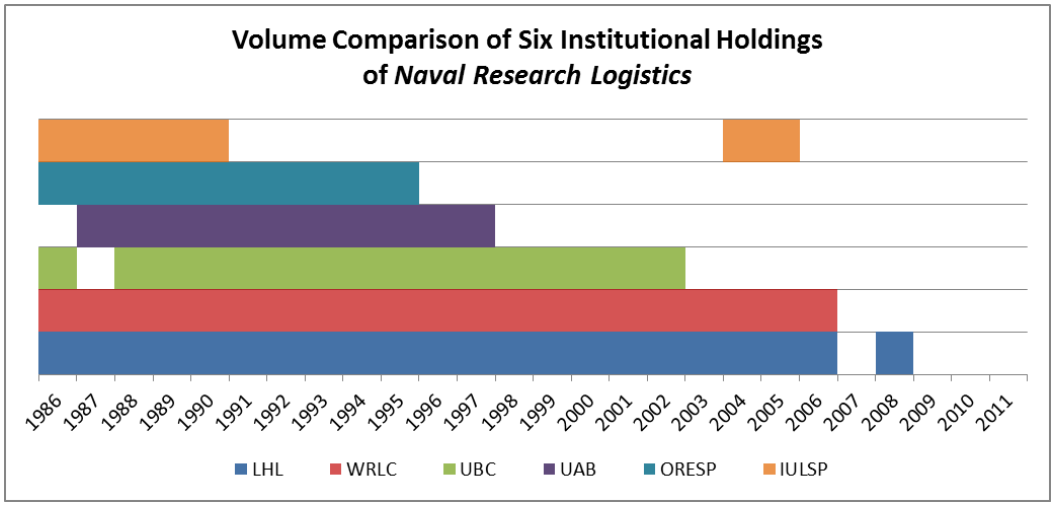
We have no data on the total number of journal titles published from year to year, limiting our ability to measure the coverage of the print archives of publications over time with any precision. The Ulrich's listing provides only the start dates for serial titles, not their end dates. However, based on that limited data the chart below shows the disparity between the number of publications begun over a 200 year period and the number of volumes of those titles that are preserved in print archives.



This does not adjust for the probable over-counting of titles registered in PAPR, in instances where title variants and title changes were recorded by submitting libraries in holdings data as separate titles. Moreover, of the titles registered, many (an indeterminate number) are represented by incomplete runs. When one looks at the issue-

level and even volume-level data -- in the rare instances it is available -- one detects gaps in holdings that are missed when only span (first and last issue) dates are reported.

The chart below illustrates the relative completeness of runs of one journal, Naval Research Logistics, reported by six different archiving institutions from which CRL obtained issue-level holdings data. Missing issues are indicated by a lack of color.



None of the six runs is complete, but the holdings of the Linda Hall Library (LHL), Washington Research Libraries Consortium (WRLC) and the University of British Columbia (UBC) most closely approximate completeness. The University of Alberta (UAB) and Indiana University (IULSP) hold much shorter runs.

b) Redundancy among the archives of print serials is rare and sporadic.

Of the 71,527 unique titles registered in PAPR, only 13,513 (19%) appear to be held by more than one archiving program. All programs hold at least one title that is also held by another program; and with the exception of CHLA, all programs hold multiple copies of at least one title also held by another program. Moreover, hidden in these numbers are multiple records for titles held by individual institutions that are registered with multiple archive programs, as well as accidental duplicate records. Therefore one should assume that the 19% figure may well be inflated. And again many, if not most, of the titles are probably represented by partial holdings.

	Program								
	CHLA	CIC/ SPR	COPPUL	LHL	Legal PAPR	Maine SC	PALCI	ST	WEST
2 copies of the same title (or titles with the same ISSN) held by different programs	0	27	519	396	424	208	39	2615	1101
3 copies of the same title (or titles with the same ISSN) held . . .	0	2	49	29	3	55	0	820	68
4 copies of the same title (or titles with the same ISSN) held	0	0	3	3	0	17	0	126	6
5 copies of the same title (or titles with the same ISSN)	0	0	1	0	0	16	0	22	3

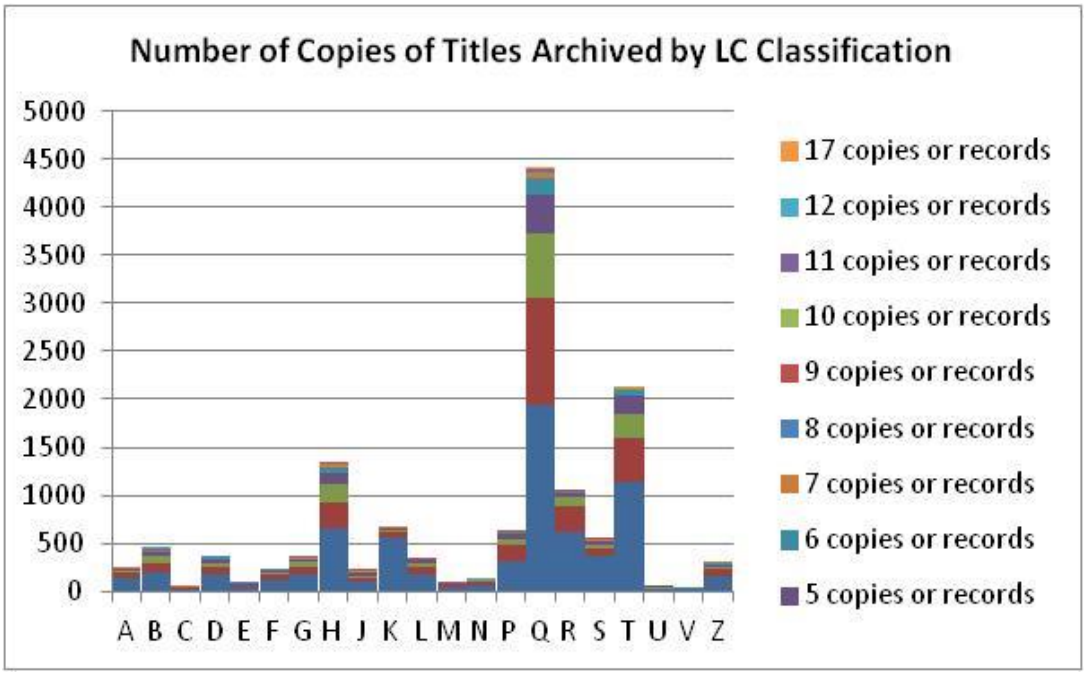
held									
More than 5 copies of the same title (or titles with the same ISSN)	0	0	3	1	0	6	0	14	1

- The archive programs with the greatest amount of overlap in holdings with other programs, as a percentage of the total number of titles held, were Scholars Trust (ST), COPPUL, WEST, PALCI, and Maine.

- The archive programs with the least amount of overlap in holdings with others were the Core Historical Literature of Agriculture (CHLA), CIC Shared Print, and Linda Hall Library.

- Of the 13,513 titles held by multiple PAPR archives, almost 2,000 are JSTOR titles, which are also held by at least two other JSTOR archives not yet represented in the PAPR database: the University of California and Harvard.

In terms of subject matter, most of the redundancy is in Science (LC class Q) and Technology (class T) titles.



On the other hand, because many archive programs focus on titles that are widely held, their holdings are the most likely to be duplicated in libraries that are not registered as archives. Many of these duplicate holdings are titles that also exist, and are subscribed to, in electronic form, such as JSTOR journals and the Elsevier-Springer-Wiley titles. Therefore much of the redundancy is concentrated in materials that are least likely to be at risk.

c) Science and Technology titles are the most widely archived.

Of all titles registered in the PAPR database, the highest concentration of titles, determined using LC classification, is in the field of Science (class Q). Titles in this class make up 25% of the titles archived; followed by Technology (T) at 22%, and Social Science (H) a distant third at 9%. (See chart below.) This concentration is in reverse order of their representation in the Ulrich's title listing, where the clear preponderance is in Social Science (22%), followed by Technology (12%) and Science (10%). No other class constitutes more than 7.1% of the holdings registered in the PAPR database.

The subject classes least well represented among archived titles are class C ("Auxiliary Sciences of History," including archeology, numismatics, genealogy, biography, etc.) and U ("Military Science"). Among the archived titles included in the PAPR database, the K class (Law) titles are heavily concentrated in the holdings of the LegalPAPR program.

LC Class	PAPR		Ulrich's	
	No. of Titles	Percentage	No. of Titles	Percentage
H (Social Science)	4,794	9.45 %	105,749	22.39 %
T (Technology)	11,541	22.75	58,772	12.44
Q (Science)	12,869	25.36	49,900	10.56
R (Medicine)	3,002	5.91	40,554	8.58
G (Geography)	1,284	2.53	29,285	6.20
S (Agriculture)	3,374	6.65	22,355	4.73
P (Literature)	1,565	3.09	21,326	4.51
B (Philosophy)	1,078	2.13	18,225	3.85
A (General Works)	1,275	2.51	17,550	3.71
L (Education)	881	1.73	16,993	3.59
K (Law)	3,636	7.16	16,446	3.48
Z (Bibliography)	1,363	2.68	16,136	3.41
J (Political Science)	687	1.35	15,823	3.35
D (World History)	1,093	2.15	12,734	2.70
N (Fine Arts)	515	1.01	12,413	2.63
M (Music)	264	0.52	5,616	1.19
C (Aux. History)	171	0.33	4,730	1.00
F (Local History)	691	1.36	4,309	0.91
U (Military)	164	0.32	1,472	0.31
E (American History)	296	0.58	1,073	0.23
V (Naval Science)	183	0.36	790	0.17

d) The overwhelming majority of titles archived are in English.

English-language serials represent 78% of the titles archived in programs registered in PAPR; but only 60% of titles listed in Ulrich's. Russian-language works are more heavily represented among the archived titles (4.5%), than they are in Ulrich's (1.2%), due to a rich concentration of those holdings at the Linda Hall Library. Again, using Ulrich's as a basis for comparison is somewhat misleading, given that listing's emphasis on the literature of North America and Europe. If one counts the vast vernacular and commercial publishing of South Asia and the Middle East that is not included in Ulrich's, the percentage of non-Western language materials covered is probably much lower. As a result, foreign language materials, particularly non-Western language materials except for Japan and China, are at risk. These are materials least likely to be preserved in their own countries.

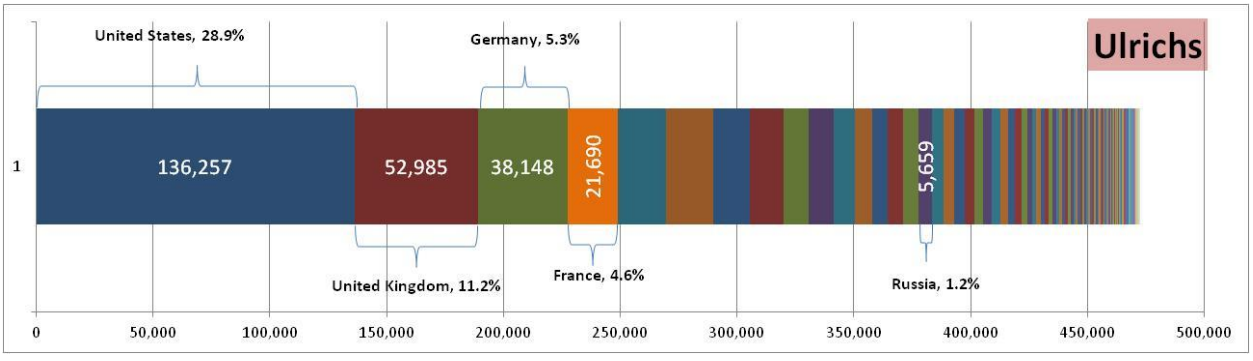
Ulrich's			**PAPR**		
Language	**Title Count**	**Percentage**	**Language**	**Title Count**	**Percentage**
English	283705	60.1 %	English	39,858	78.5 %
German	41342	8.8	Russian	2,289	4.5
French	26607	5.6	German	2,239	4.4
Spanish	18485	3.9	French	1,824	3.6
Dutch	14544	3.1	Spanish	1,050	2.1
Italian	12208	2.6	Japanese	488	1.0
Chinese	11579	2.5	Italian	473	0.9
Japanese	7204	1.5	Chinese	360	0.7
Undetermined	6266	1.3	Polish	352	0.7
Portuguese	5605	1.2	Portuguese	275	0.5
Russian	5592	1.2	Dutch	199	0.4
Danish	5511	1.2	Romanian	189	0.4
Swedish	5231	1.1	Czech	188	0.4%
Polish	3795	0.8	Swedish	103	0.2
Norwegian	2731	0.6	Bulgarian	100	0.2

e) Serials published in the U.S. are more widely archived than non-U.S. titles.

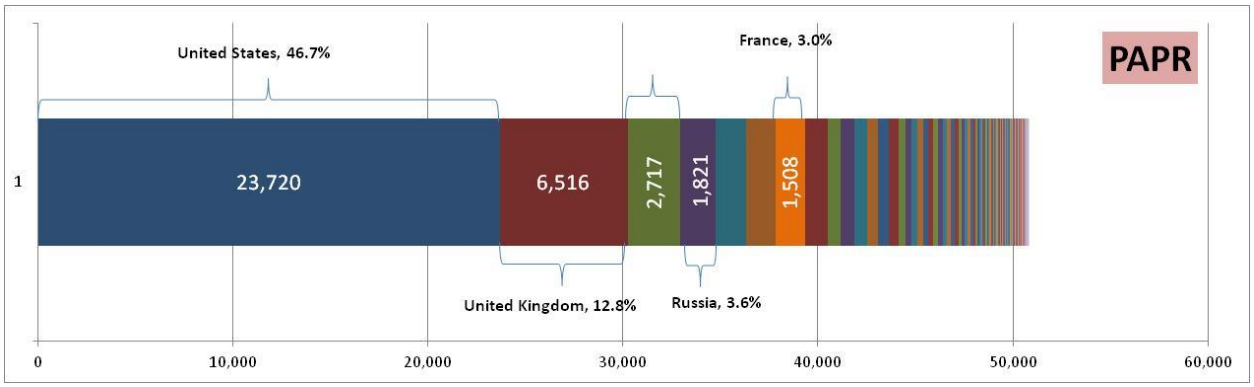
Analysis of archived titles by country of publication suggests that titles published in the U.S. are more heavily represented among the archived titles (46.7%) than in Ulrich's (28.9%). This probably owes to the emphasis on U.S. titles among the LegalPAPR holdings. This may be appropriate because the U.S. institutions that make up the majority of archiving participants are the natural "libraries of first resort" for domestic titles, and because Canadian programs tend to be reluctant to rely on U.S.-based repositories for their materials.

Ulrich's			PAPR		
Country	**Title Count**	**Percentage**	**Country**	**Title Count**	**Percentage**
United States	136,257	28.9%	United States	23,720	46.7%
United Kingdom	52,985	11.2%	United Kingdom	6,516	12.8%
Germany	38,148	8.1%	Germany	2,717	5.3%
France	21,690	4.6%	Russia	1,821	3.6%
Netherlands	20,450	4.3%	Canada	1,525	3.0%
Canada	20,120	4.3%	Netherlands	1,516	3.0%
Australia	15,678	3.3%	France	1,508	3.0%
Italy	14,636	3.1%	Japan	1,183	2.3%
China	10,769	2.3%	Italy	680	1.3%
Japan	10,480	2.2%	India	677	1.3%
Spain	9,151	1.9%	Australia	658	1.3%
India	7,600	1.6%	Switzerland	562	1.1%
Switzerland	6,718	1.4%	Poland	547	1.1%
Sweden	6,502	1.4%	China	509	1.0%

Distribution of Titles Across Country of Publication – Not relative



The relatively small representation of U.K. titles in registered U.S. and Canadian archives suggests the potential for a cooperative archiving arrangement with the U.K. Research Reserve program.



e) Most shared print and print archiving focuses on secondary, rather than primary, literature.

Not surprisingly, scholarly and scientific journals dominate the holdings of most print archive and shared print programs, while trade journals, popular serials, and industrial literature are archived by relatively few programs. This despite the fact that much of the latter literature has been retired from use, if not disposed of, by many libraries, and thus may be more at risk of loss.

The archiving of at-risk, primary source materials takes place most often in independent research libraries such as the New York Public Library, Linda Hall Library and the American Antiquarian Society, and in the special collections of academic libraries. Such efforts tend to be driven and supported by the specialized interests or focus of individual institutions rather than the result of cooperative efforts, and are therefore tend not to be coordinated with print sharing and archiving programs. However, they offer potential economies to an archiving effort: the historical investment of the independent research libraries in security (closed stacks, monitored service), climate control, acquisition with a focus on completeness and best quality edition (curatorship), and limited use and handling over the years, minimize the costs and even the necessity of validating such holdings.

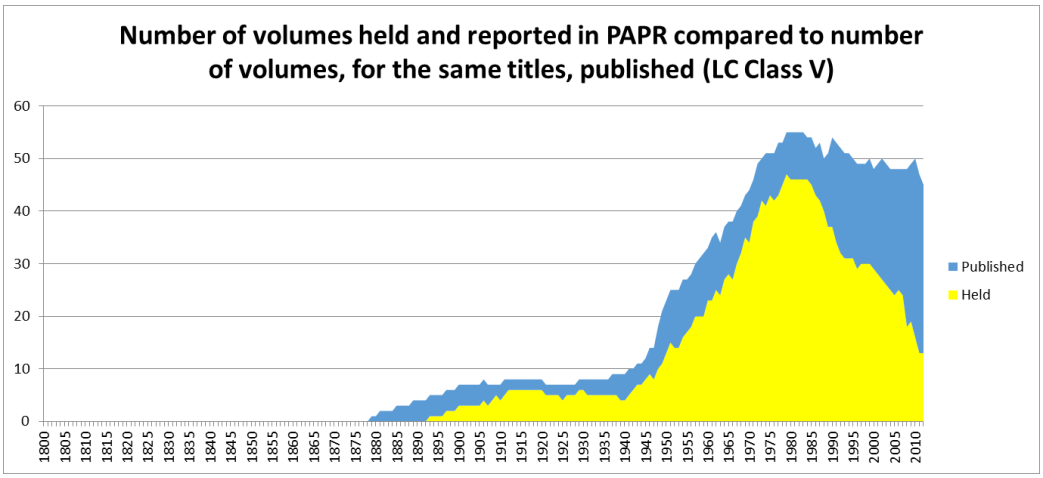
f) Historical and at-risk materials are not well covered.

A disproportionate number of the serial titles archived, as a percentage of the total number of serials published in a given year, date from the late twentieth century. Even accounting for the explosive growth in serial publishing during the later period, there is a larger disparity between the number of titles published prior to 1950 and the number of titles from that period that are held by archives registered in PAPR. This is particularly apparent when one takes into account the fact that earlier titles are less likely to have ISSNs or to be represented in the Ulrichs title list.

The early volumes in the Linda Hall Library, CRL JSTOR, and the LegalPAPR archives account for many of the early holdings in our data.

- The largest number of serial volumes archived date from the post-war period, or the second half of the twentieth-century. The chart below reflects the presence of archived copies of 2,171 volumes of the 3,331 published volumes in the 88 titles held by all registered archives in total. CRL's sampling of 88 titles in LC class V (Naval Science) suggests that the rise in titles archived during that period roughly mirrors the

published output of serials – but only up to a point. The completeness of these aggregate holdings begins to decline precipitously around 1980.



II. What the Data Tells Us about the North American Archiving Programs in General

a) Actionable information about archived holdings is scarce.

While much information is available today about the terms of archiving and shared print, that information still falls short of the granularity and detail needed to support sound risk assessments and decision-making on the preservation, retention and disposition of serial materials. This deficiency is largely the product of cost decisions made by archiving programs: creating, updating, standardizing, and verifying accurate holdings information is resource-intensive.

- Few archiving libraries provide consistently granular data on their local holdings of serial titles. In many instances holdings statements reflect only the dates of first and last issues held, and thereby obscure the existence of gaps in runs. Moreover, holdings information is expressed in a variety of formats, which are not easily analyzed in a uniform way.

- Often even the data that is reported is unreliable, due to inadequate validation of holdings. Validation of individual issues of serial titles, for example, is not required by most archive programs. Apart from the JSTOR archives validation, where it exists, at the volume level appears to be more common than issue-level validation. The absence of validation makes it difficult to verify completeness of a given series or the physical state of the copies archived. This means that there are varying levels of certainty as to whether titles some libraries agree to archive are complete or intact. The exception to this rule is when archiving is undertaken in tandem with other certain types of collection-related work, such as relocating and digitizing collections, or as part of active collection development and gap-filling efforts.

- Redundancy is often used to compensate for the lack of granularity in holdings statements and relatively minimal validation: if multiple holdings of a given title are held, it is reasoned that gaps in the holdings of some archiving libraries will be filled by the holdings of others. However it is not certain that gaps in one archive's holdings of a given title will not be repeated elsewhere.

Eventually, this kind of information will be needed to promote confidence in decision-making about "last copies." Today, however, the absence of this data already makes it difficult for libraries to demonstrate that consequential decisions and actions on management of print are made on an objective, empirical basis. One practical benefit of having this kind of data could be the ability to reduce the number of copies necessary to provide assurance of the survival of complete and intact runs of important titles.

b) Actionable information about archive programs is scarce.

Information about the nature of the commitments made by archiving libraries to verify and maintain archived runs of serials is all relevant to the level of certainty those archives can provide to stakeholders. Also relevant is information about the measures in place to fulfill those archiving obligations. Such information is often unavailable or dispersed.

- While the major print archiving / shared print programs openly publish their policies and principles, aggregation and effective evaluation of the information disclosed are impeded by a lack of uniformity in

how that information is expressed. Terminology regarding retention commitments, terms of access and service, and program organization, for example, varies from one archive program to the next and is thus difficult to codify and compare.

- There is relatively little precise, structured data available on the environmental conditions in which archiving libraries maintain archived materials. Conditions reported to CRL indicate wide variation in the security, climate control, and hazard prevention measures provided for archived materials. Facilities range from open, university stacks to remote underground storage. The MARC 583 field provides a receptacle for such information, but may not be extensible enough to handle and structure the data at scale. As time goes on and print becomes less common, this kind of data will become more important.

- Many libraries are reluctant to disclose documentation of the validation of titles thy archive. Information and evidence of the validation of JSTOR volumes by some programs, for example, was not readily available.

- The obligations of archiving institutions are often not well defined. For example, it is not always clear whether an archiving institution has committed merely to retaining designated titles, or whether it has also agreed to monitor and complete runs of those titles, or to replace and/or restore archived copies if they become missing or damaged.

- Archive obligations, moreover, are formalized through a wide variety of instruments and with varying degrees of specificity. Many are expressed only in non-binding or short-term memoranda of understanding; in other instances published guidelines and policies are the only governing documents. Archiving provisions can therefore lack "teeth" or enforceability.

This suggests that there is relatively little certainty that the titles archived will be preserved intact and complete, and will be available for the long term.

c) Volume is more common than value as a criterion for a criterion for selecting materials for archiving.

Many archive programs are designed to achieve coverage of the largest amount of material, in terms of space reclaimed or the number of volumes archived. Because recovering space used for journals storage is an imperative of many libraries, many print archive programs focus on materials that occupy the greatest cubic and linear footage.

- Many programs take advantage of "low-hanging fruit", i.e., journals that are available, and primarily used in, electronic form, and therefore can be removed with relatively little impact on users. However, since space reclamation is a one-time benefit, the incentive for libraries to support the retention of these titles may decline over time.

- Many programs also favor materials that are widely held by participating libraries, such as JSTOR serials and journals published by Elsevier, Springer, Wiley and other major scholarly and scientific publishers, over materials that are intrinsically at-risk of loss. This generates the greatest immediate return on

participant investment in terms of space recovered, and maximizes the number of potential beneficiary libraries, but may be risking the loss of less common materials and "last copies."

- Only a few print archives are highly "curated", or focused on assembling collections of materials identified by scholars and specialists as critical materials for research. Exceptions include:

    - The Linda Hall Library, a strategically developed body of materials in science, technology and engineering, with deep historical back files;

    - The Core Historical Literature of Agriculture (CHLA), a collection of agriculture journals prioritized for preservation by scholars and bibliographers;

    - The JSTOR print archives, covering materials that are widely held (low-risk) but also of high value, because of the highly strategic and consultative nature of the JSTOR selection process.

d) Most programs archive materials "in place", as opposed to extracting and isolating archived materials in dedicated environments.

Reliance upon extant runs of titles currently in campus stacks or remote storage, rather than assembling and isolating holdings for the purpose of archiving, is a common practice. This is the least expensive and least disruptive approach to creating designated archives of titles, particularly when archived holdings are formed based on the depth of the archiving libraries' existing holdings. In such instances verification of holdings is based on existing records, which can be imprecise, incomplete or outdated. Shelf-reading is rarely done.

Better controls are normally implemented in the process of preparing library materials for relocation and transport. Inventorying, inspection, counting, and rehousing materials often take place at these times. Absent this closer verification, the likelihood increases that gaps will exist, undetected and unreported, in the archived holdings.

Archiving in place also complicates the process, and thereby increases the future costs, of adding missing volumes to complete the series and later adding new volumes to existing runs of current titles.

Anecdotal evidence suggests, however, that pressures to vacate library space devoted to storing collections, particularly in the fields of law, agriculture, and government information, are changing the cost-benefit calculation of extracting and isolating materials for archiving, increasing the opportunity costs of retaining collections in place. This is evidenced by the accelerating pace of retirement of titles to the Law Library Microform Consortium's underground storage and ReCAP, and the recent investment of major universities like the University of Chicago, University of Pennsylvania, and Northwestern University in new dedicated storage facilities.

e) Print sharing and archiving programs vary considerably in the benefits and incentives they provide.

The benefits archiving programs provide to participating libraries vary from one program to another. A variety of incentives for participation exist, some weaker and shorter-lived than others.

By far, the main incentive for libraries to participate in shared print and print archiving is the ability to reclaim storage space through withdrawal of volumes archived elsewhere. This encourages programs to focus on archiving the greatest number of volumes of titles held by the largest number of participating libraries, thereby maximizing the amount of storage space that can be reclaimed. Unfortunately this means that titles that are rare, and by definition most at risk, are least likely to be archived.

Other programs provide services based on their pooling and sharing of print collections: interlibrary loan of archived volumes, document delivery of articles from those titles, and digitization on demand from archived materials as well.

The emphasis on storage space reduction provides considerable value, but in the form of a one-time benefit: stakeholder libraries can reclaim storage space by deaccessioning or relocating titles archived elsewhere. This might make sustainability of such programs difficult, once supporting libraries have realized their initial gains and accrue fewer ongoing benefits from the program thereafter.

Programs that focus on pooling and sharing print collections provide an additional benefit: enlarging the pool of resources available to participating libraries. This creates a stronger incentive for participating libraries to provide ongoing support for the programs.

A few programs actually compensate archiving institutions for retaining volumes, but this compensation tends to be relatively insignificant in the overall operating finances of most archiving institutions.

Conversely, some programs put in place other incentives for archiving libraries to maintain and protect the designated titles, such as penalties for the loss of archived volumes. But these provisions are rare.