Center *for* Research Libraries
GLOBAL RESOURCES NETWORK

The "State of the Art": A Comparative Analysis of Newspaper Digitization to Date

April 10, 2015

About this Analysis

As background for discussion at "Framing a Common Agenda for Newspaper Digitization and Preservation: an ICON Summit," CRL has undertaken an assessment of the outputs of the major newspaper digitization efforts to date in the United States and Europe. The purpose of the assessment was to identify the comparative strengths and gaps in the corpus of newspapers digitized to date, relative to the total body of newspapers held by libraries. The data presented in this report are based on information gathered by CRL on major newspaper digitization efforts by national and academic libraries, major producers of commercial academic producers, and other significant stakeholders in the newspaper digitization arena.

To assess the state of newspaper digitization, CRL used both title-level information available from major newspaper digitization projects, and issue-level information aggregated by CRL in the ICON Database of International Newspapers (http://icon.crl.edu). The ICON database is a registry of holdings information on the hard copy, microform and digitized holdings of U.S. and foreign newspapers. Its purpose is to provide reliable information for libraries, publishers and others engaged in preserving, collecting, and digitizing newspapers, about holdings of interest in major research collections and about the contents of "trustworthy" repositories and databases of digitized newspapers.

The largest single repository of information about such holdings, ICON currently contains records for over 171,000 newspaper titles, published in 51 U.S. jurisdictions, nine Canadian provinces, and 159 other nations, and held by hundreds of U.S. research libraries. The publication dates of these holdings range from 1649 to 2015. ICON provides granular information on newspaper holdings, listing over 40 million individual issues of titles held in U.S repositories alone. Issue-level holdings are derived by CRL from bibliographic catalog holdings information and through direct submission by library and commercial partners, or are harvested from digital newspaper repositories considered by CRL to be trustworthy and persistent.

The report will briefly discuss four areas of digitization efforts,
- Global Coverage
- Western Europe
- United States
- Representation of other world areas

Development of the ICON database and this CRL analysis is supported by the National Endowment for the Humanities, Institute of Museum and Library Services, the Andrew W. Mellon Foundation, and the CRL member library community.

Limitations of the Analysis and Next Steps

This report must be read with numerous caveats, foremost among them that the holdings data on which the analysis is based is of varying degrees of accuracy and granularity. This is because, while much of the data in ICON is verified to the issue level, a considerable amount of the data was derived from title lists containing only "span dates," or, dates of the first and last issues of a given title held. Thus, reliable information on gaps and issue-level specificity is not yet available for all titles. CRL will continue to gather more granular information from additional sources through the ICON database, and report on subsequent findings on a periodic basis.

Second, the analysis is based on data on the newspaper holdings of U.S. and European libraries only and on the contents of selected newspaper databases published by commercial producers that CRL considers "trustworthy." CRL did not gather data on newspapers digitized in countries and regions outside Europe, the U.K., and North America. Nor does the report take into account newspapers digitized by genealogical publishers and other aggregators.
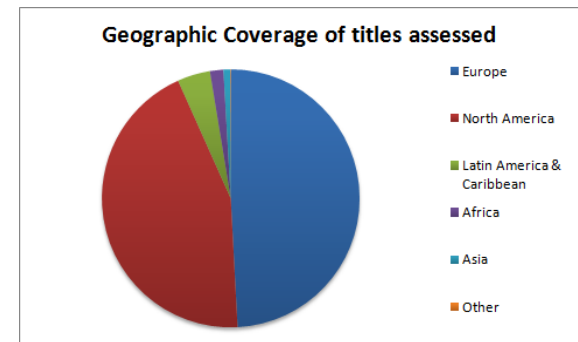
This report makes a case for the active provision of detailed information from trusted repositories, and concludes with some general observations and recommendations for action.

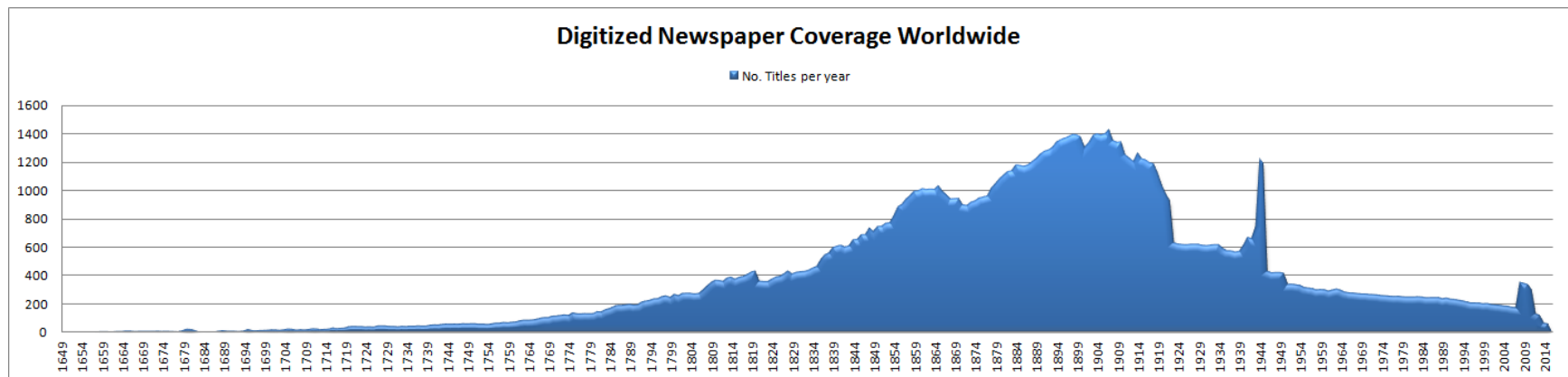## I.   Global Coverage – Number of titles digitized

Based on information gathered by CRL, supplemented by reports from major initiatives such as Europeana, a conservative estimate of the total number of newspapers from all world regions digitized to date (any years) exceeds 30,000 titles. Given that this present study is (as yet) relatively narrowly confined to digitized newspapers held in major repositories in the U.S. the U.K. and Europe, the number of digitized titles available worldwide could easily exceed 45,000 when all countries are taken into account.

CRL has produced a data set of nearly 12,000 digitized news titles drawn from information in the ICON database as well as title-level information gathered from approximately 90 different digital collections. A list of major repositories/producers consulted is available in Appendix A.

The majority of titles in CRL's data set are European- or U.S.-based publications, digitized by the major national and academic libraries in the countries of origin, and by commercial producers (including news publishers and content aggregators). Latin American, African, and Asian (including South, Southeast, and East Asia) titles digitized are largely drawn from CRL/Readex's World Newspaper Archive as well as selected titles digitized by the targeted libraries and commercial providers. The distribution of title coverage by world region is illustrated in the following chart:
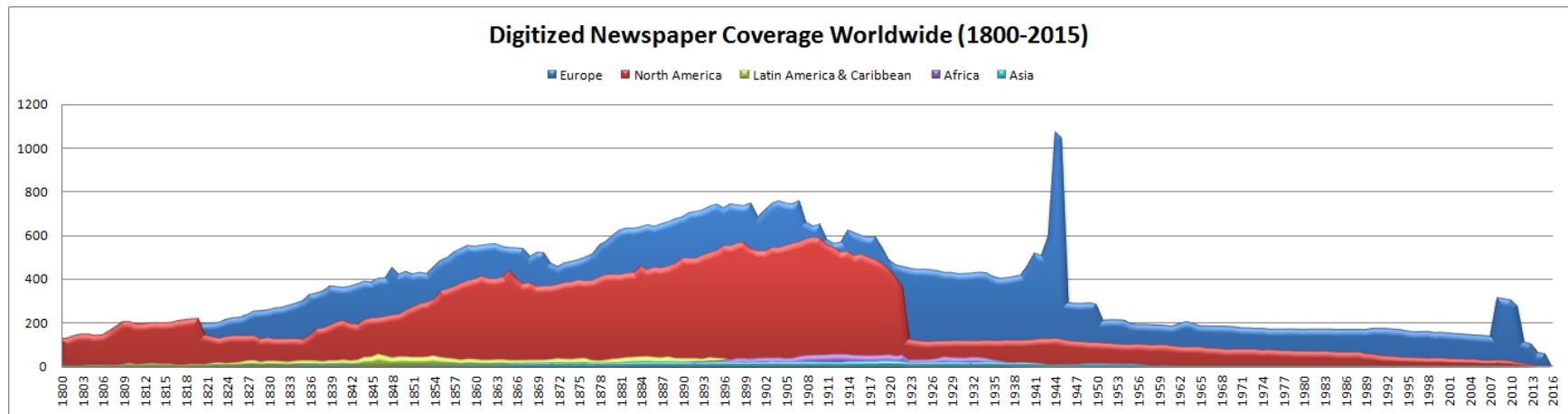

Geographic Coverage of titles assessed

- Europe
- North America
- Latin America & Caribbean
- Africa
- Asia
- Other

Plotting the titles along a time graph, the chronological scope of global coverage of digitized newspapers can be seen as follows:

Digitized Newspaper Coverage Worldwide

No. Titles per year

It is important to note that the representation of titles above should not be taken as a measure of comprehensiveness: for many of the titles, only a few years—or even a few issues—may be represented in online databases. The data presented here shows only whether a title was present in digitized form, even if represented by only a single issue, within a given year and does not take into account the "saturation" of that title (i.e., the total number of issues of the title represented in a given year). Additionally, title counts do not account for whether titles were dailies, weeklies, or other publication frequencies. Issue-level specificity is required to present an accurate account of saturation over time.

In the following area chart, one can observe the relative size and chronological scope of titles broken down and overlaid by world region.
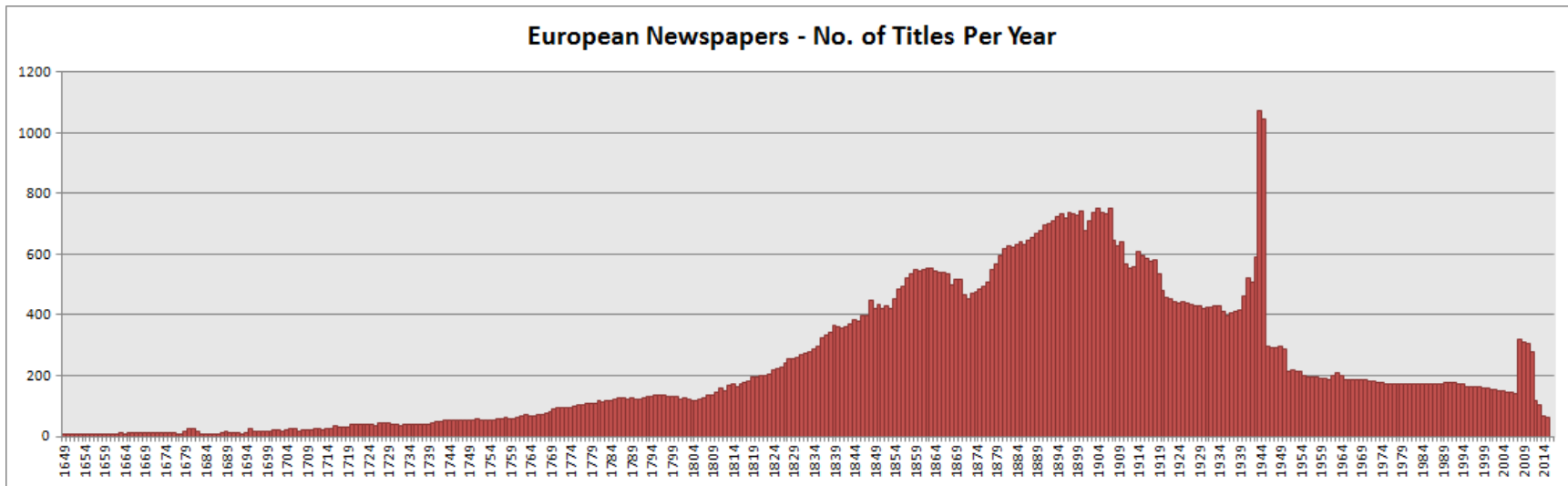

For purposes of readability, the chart above covers only from 1800 to the present In general, the peak coverage for all areas is in the period 1890-1918, though European newspaper coverage peaks slightly earlier (1882-1914), and U.S. newspapers sustain coverage through 1920. The depth of coverage—i.e., the "shape" of newspaper digitization-- reflects priorities of newspaper digitization efforts, driven by cultural, scholarly, and financial interests, but also tempered by copyright restrictions and limited engagement with current publishers. Details of coverage of various world regions are presented below.

II.   *European and U.K. Newspapers Coverage*

An examination of newspaper coverage from Western Europe through the major national library and/or regional academic efforts reveals trends analogous to global newspaper coverage. Peak coverage, as described above, begins in the 1880's and runs through 1918 (likely a product of the European-wide emphasis on digitizing newspapers from the World War I era). Recent projects initiated to scan newspapers from the World War II period produce a spike in coverage from 1944-1945.

As a whole, European approaches to copyright have extended the reach of newspaper coverage up through the 1930's and 1940's (while US copyright rules, in contrast, have generally limited content in the public domain to pre-1923). A survey conducted by Europeana Newspapers in 2012[1] reported that most European newspaper programs use a "sliding-scale" cutoff of 70 years for newspaper digitization, suggesting that content up through 1944 is now eligible for digitization, free of rights.

The Europeana Newspapers survey found that the number of digitized newspapers in Europe was estimated at 24,000 titles, with over 129 million pages scanned as of 2012. CRL's assessment covered only a portion of the digital library initiatives in Europe, resulting in approximately 5,800 titles as shown below.
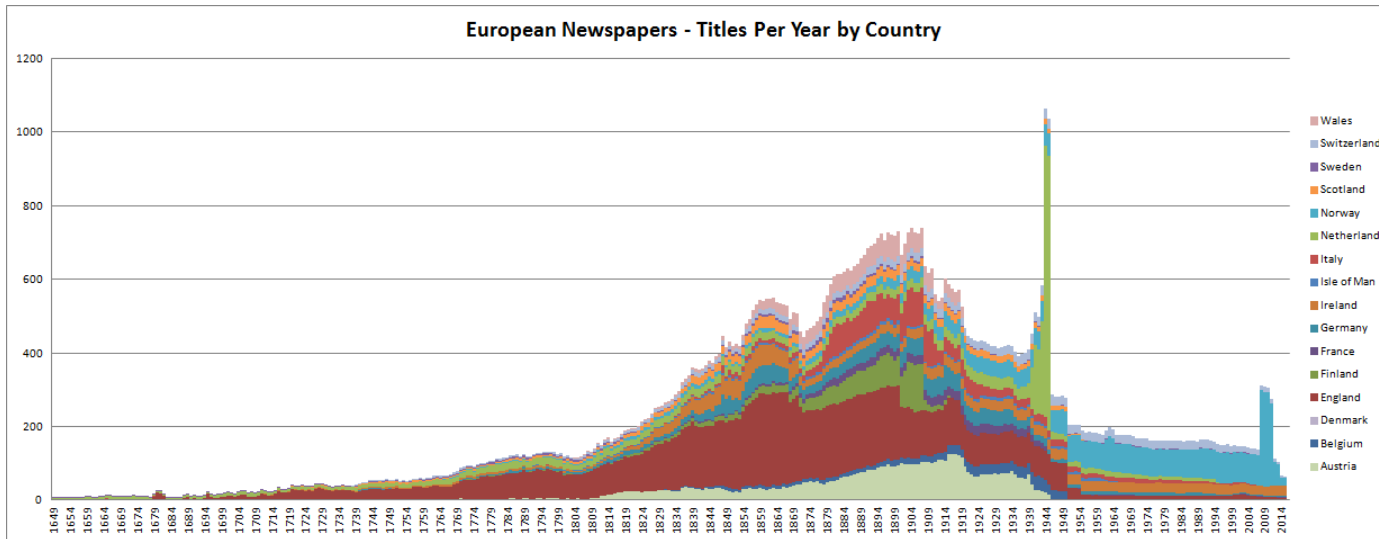

Breakdown by Country

Assessing the total coverage of European and U.K. newspapers broken down by country (below) reveals a number of interesting trends.
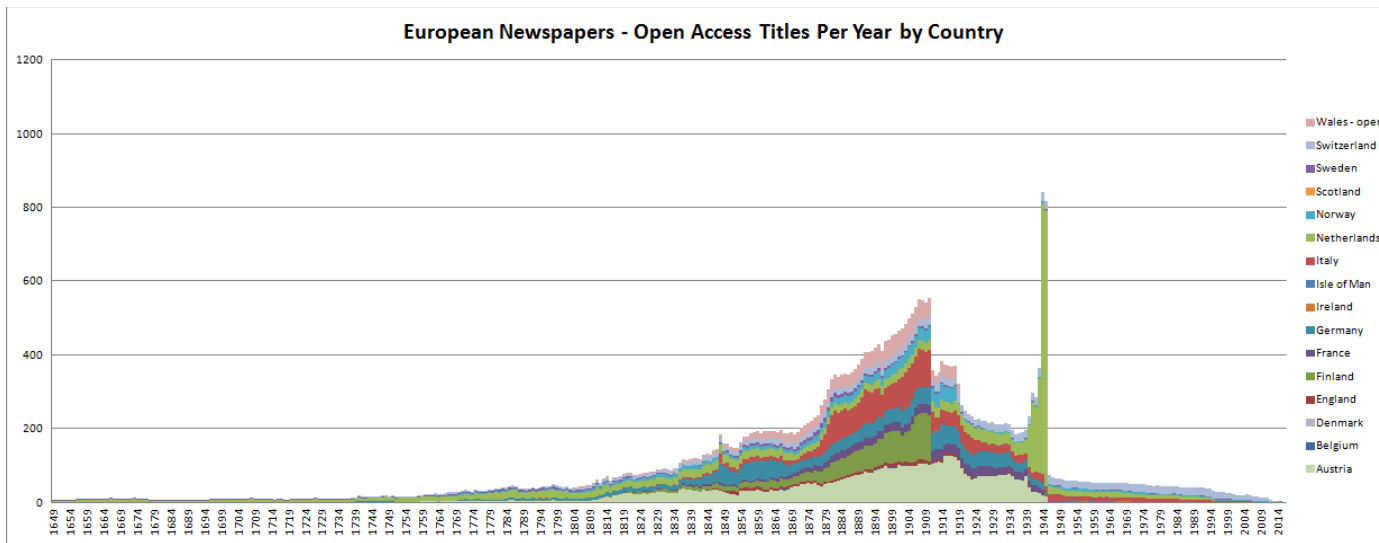- Conservative / restrictive copyright approaches have had an impact on the availability of titles. The various declared cutoff dates (1900 in the UK, 1910 in Finland, for example) cause drop-offs in the overall availability. Norway has taken the approach of scanning content up through to the present day, but only making content openly accessible up through 1918. The overall drop in content post-1944 can be prominently seen in the overall representation of titles available.

---

[1] http://www.europeana-newspapers.eu/wp-content/uploads/2012/04/D4.1-Europeana-newspapers-survey-report.pdf

- Coverage of titles produced through commercial partnerships brings significant benefits in terms of content coverage (though at a price). England's digital newspaper titles dominate the title count for European newspapers, primarily produced in cooperation with commercial publishers (including Gale Cengage, the British Newspaper Archive, and others). While the UK generally employs a cutoff date of 1900 for public domain content, the commercial partners appear to have been successful in negotiating permissions to display content for a fair number of titles up through 1950.
- The Netherlands' coverage of World War II titles far outpaces other libraries' efforts to date. The Royal Library of the Netherlands scanned nearly 1,000 titles of "illegal newspapers" from the period 1940-1945.

When taking into account the restrictions on viewing (whether through access limited to the premises of the national library or selected institutions only, or through commercial licensing arrangements), the number of "open access" titles in Europe appears much smaller.

Most English-language titles digitized (including UK, Ireland, and Scotland) are presently offered under commercial license, thus restricting the open availability of a wide swathe of content from Europe.

How comprehensive is European title coverage? The 2012 Europeana Newspapers survey suggested that few libraries had digitized more than 10% of their collection (either in terms of titles or page numbers). Libraries in Europe have taken varying approaches to digitization, ranging from thematic selection, chronological coverage, and historical significance to comprehensive coverage of cultural heritage from the earliest available material forward. Though exact numbers of newspapers published in Europe are difficult to ascertain, it may be assumed that there is significant room for growth in this area.

III.   *United States Newspapers Coverage (a sampling)*

Newspaper coverage in the United States is less deep in terms of chronological coverage and more diffuse (in terms of stakeholders and access models), compared to European models. While several universities and commercial producers began digitization efforts in the late 1990s and early 2000s, library participation in digitization began in earnest with the federally-funded National Digital Newspapers Program (NDNP) initiated in 2004. Coordinated by the Library of Congress, and funded
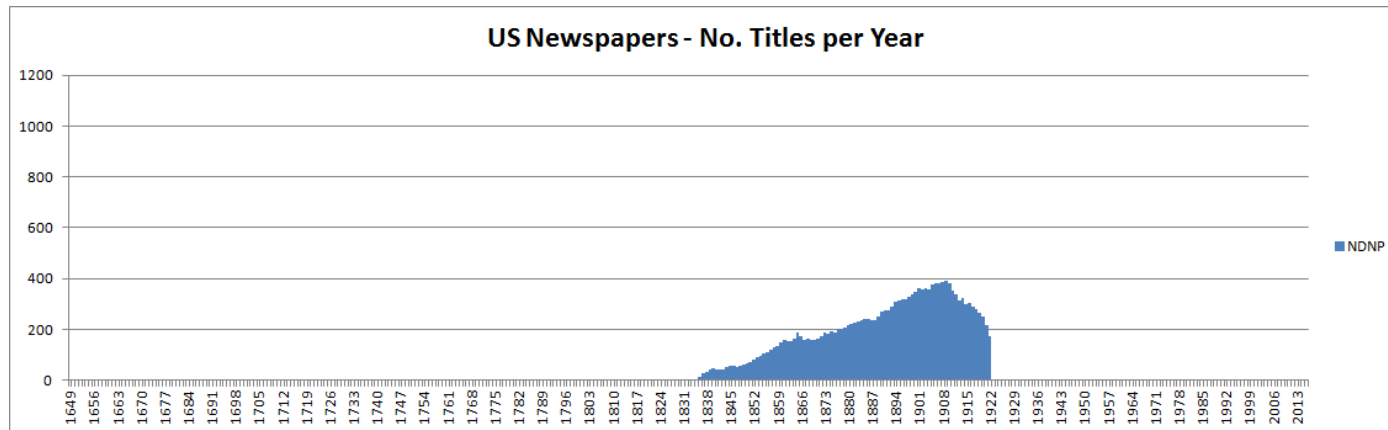
Source: NEH.

by the National Endowment for the Humanities, the NDNP has awarded grants to institutions in 39 states and territories to date, for digitization of newspapers published during the period 1836–1922.

Annual appropriations for the NDNP program average $3.3 million per year. NEH awards funding to 10–15 state programs per year to digitize 100,000 pages of newspapers per award. Should support for the program continue at a consistent rate, NDNP stands to reach ca. 15 million pages (representing all 50 states with 300,000 pages per state) by the year 2020. Costs for the program (direct awards) could reach $46 million U.S. dollars. [2]

As of April 6, 2014, 1,728 newspaper titles are represented in Chronicling America, the Library of Congress' platform for access to the NDNP newspapers. The coverage of NDNP, shown on the same scale as the European newspapers chart, can be seen below.

"Beyond NDNP"

As funding and scope for the NDNP is limited, many of the current and former participants have continued digitizing content beyond the scope of the NDNP. A 2014 survey of 35 current and former participants in the program—academic libraries, state historical societies, and state libraries—revealed that these institutions have produced to date more than 15 million pages of content beyond the scope of NDNP. The vast majority of this content is served open access through the repositories designated for their newspaper programs.[3] A number of the participants have begun to negotiate permissions to provide access to current newspaper content directly from publishers or through state press agencies. This may include digitizing backfiles in addition to ingesting PDF versions of current newspapers.

---

[2] Total funding to date (2004-2016) is $28.5 million, and 9.52 million pages have been (or are in the process of being) digitized.
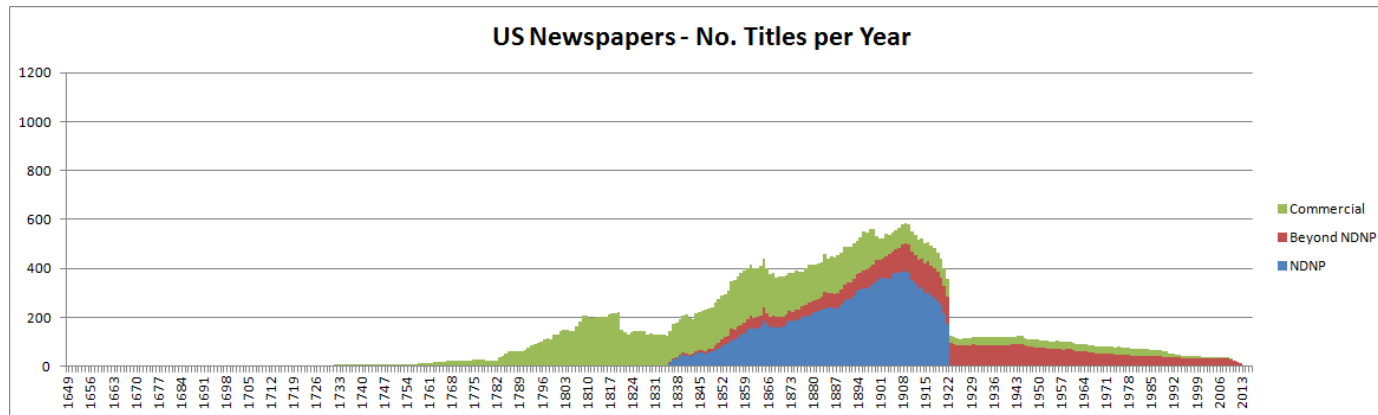[3] The range of repository platforms utilized is diverse, including ContentDM, ChronAm (LC's open source platform), DSpace, Veridian, locally-developed platforms, and other proprietary systems.

CRL's assessment included 10 current or former participants of the NDNP that digitize newspapers beyond the scope of NDNP. The institutions represent a range of types (academic, state, historical), with varying intensity of digitization activities. From this sampling, 1,250 titles "beyond NDNP" were identified, representing publications dating from as early as 1787 and as recent as the present day.

Commercial Producers
In the United States, numerous commercial providers have also been actively digitizing and hosting newspaper content, in databases made available mostly to academic and public libraries through one-time purchase or ongoing annual subscriptions. Major providers include East View Information Services, Gale Cengage, ProQuest, and Readex (a division of NewsBank). These providers play an important role in the ecosystem of digitized newspapers by bearing the initial cost of digitizing and hosting historical and recent materials and (in the case of copyrighted content) obtaining the necessary rights to provide public access. CRL's assessment included multiple collections from these providers (whether single-title products or historical news aggregations), identifying 2,212 additional titles.

The following chart illustrates the complementary role played by the diverse players in U.S. newspaper digitization, including Chronicling America, other libraries, and commercial partners. Results are stacked to show total coverage from the three types of producers.

Overall, how comprehensive are U.S. newspaper digitization efforts? While data gathered by CRL to date covers only a representative sampling of U.S. titles digitized, even a generous tripling of the estimated number of titles digitized would suggest that the U.S. has not yet reached 10% of total coverage of pre-1923 newspapers held by libraries. For instance, based on the number of titles cataloged through the U.S. Newspapers program, it is estimated that in 1922 there were more than 15,000 titles currently being published in the United States. CRL's sampling for 1922

indicated only 360 titles available in digitized format. Even at the peak of digitized news coverage (1908) in the sample, the number of titles represented was less than 4% of the total press output in the U.S.[4]
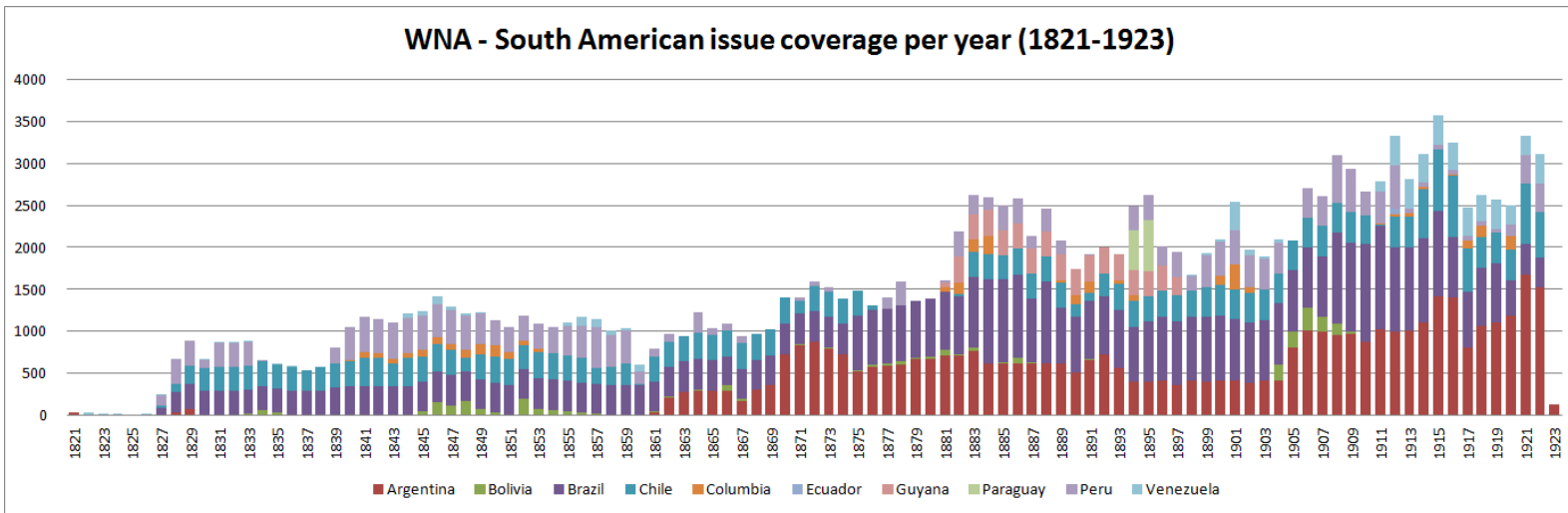
IV.   *Representation of Other World Areas*

Beyond the United States and Europe, the number of large-scale, sustained programs of newspaper digitization is relatively small, but continues to grow with each passing year. ICON's [directory of newspaper digitization projects](#) identifies and describes many of the existing digital newspaper programs. National and academic libraries in Australia, Israel, Korea, New Zealand, and Singapore have developed moderate- to large-scale digitization programs. In Latin America, Mexico's National Library has scanned 545 titles published between 1822 and 1876, while Brazil's Digital Newspaper Library lists 5161 newspapers and magazines from the 19th and 20th century. In China and Japan, commercial ventures have largely led the foray into digitized news, offering a variety of long-run titles on proprietary platforms. These collections were not analyzed for the present study, but will be added at a later date.
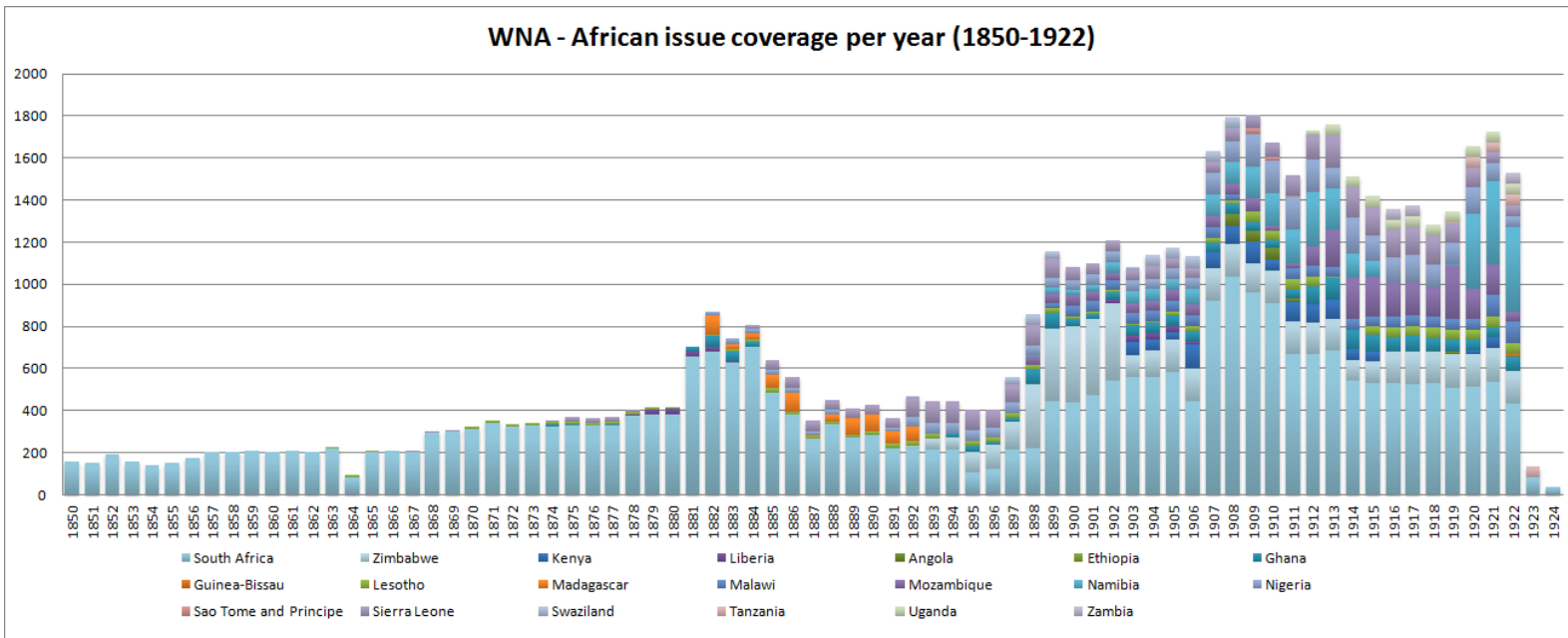
In the absence of library-based programs in other parts of the world, commercial providers have begun to expand the availability of titles. ProQuest has scanned long-running titles such as the *Times of India*, *Jerusalem Post*, and the *South China Morning Post* through its Historical Newspapers program. East View Information Service has digitized backfiles from a number of Russian newspapers, including *Pravda* (1912-2009), *Izvestiia* (1917-2010), *Literaturnaia gazeta*, and *Argumenty i fakty*. These producers provide (fee-based) access to content using platforms familiar to researchers and to libraries. Together, these sources add more than 200,000 issues to the global newspaper count.

CRL has collected granular information on newspapers digitized through the World Newspaper Archive (WNA), a cooperative effort of CRL and Readex to make accessible the wealth of international collections amassed over a century of sustained acquisition and preservation by CRL and its member libraries. To date, this effort has digitized nearly 3.5 million pages of content from Latin America and the Caribbean (2 series), Sub-Saharan Africa, and South Asia. The representation of issues per country over time can help shed light on the depth of coverage for each country. Examples below show coverage for Africa and for South America within WNA.

---

[4] "The Growth of Newspapers Across the U.S., 1690-2011" - http://web.stanford.edu/group/ruralwest/cgi-bin/drupal/visualizations/us_newspapers

WNA - South American issue coverage per year (1821-1923)

Issue-level metadata provided courtesy of Readex via ICON

WNA - African issue coverage per year (1850-1922)

Issue-level metadata provided courtesy of Readex via ICON

How comprehensive is the World Newspaper Archive? Despite growing interest and investment in this area, the digitization of newspaper content from the southern hemisphere is still in its early stages. As a library/private sector model, the WNA is financially supported by the North American academic library community rather than through government or other external support. As such, the approach to large-scale digitization is constrained by the level of collection funds academic libraries can manage.

As for depth of coverage, content is sourced primarily from U.S. institutions with rich collections of regional material from the 19th through 21st centuries. However, a conservative approach to intellectual property from CRL's partner has limited the content thus far to pre-1923 material. The majority of WNA content was scanned from microfilm, the predominant medium of most early international news collections. Thus, for coverage up to 1922, extrapolating from title/issue counts within ICON and comparing total digital issues to issues in all formats, it can be estimated that CRL and WNA have scanned 25–30% of available content held by participating North American libraries.

However, the volume of content post-1923 dwarfs that of the preceding period (both in terms of numbers of titles and in issue/page counts) and remains, as yet, largely un-scanned. As a telling statistic, estimates of CRL's Latin American newspaper collection put the total number of available pages in microform and paper at over 30 million Thus, the amount of content available digitally to date represents only a fraction of the total holdings of international news content held in North America (not to mention the rich heritage collections in the regions).
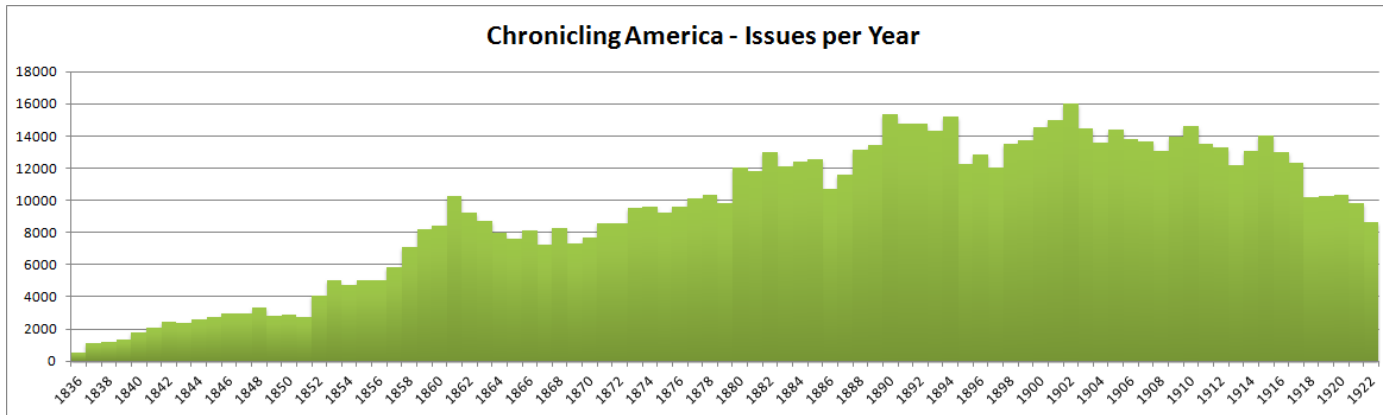
V.   Comparison of Content : The Case for More Data

The ICON database was developed in response to concerns expressed by libraries regarding the availability of trustworthy and detailed information on newspapers held in digital, microform, and print formats. Libraries require such information to make informed investment decisions on the purchase of newspaper databases, and to enable them to decide whether to invest resources in microform and digital reformatting of titles in their own collections. They also need better data to help determine whether to deacidify or conserve materials in their own collections, or whether to fill gaps or even extend incomplete runs of those print titles. In order for libraries to make these decisions the information consulted must be detailed and reliable.

The ICON database is designed to be a collection management tool for libraries and commercial producers and publishers of news databases. It provides granular information on the availability, location and access options of newspaper titles and issues in digital format, as well as in print and microform. It provides much-needed transparency of existing resources and products, and can support planning and prioritization for digitization of future collections

Case Example: Assessment of database coverage

Based on a harvest of title and granular issue information for a database of digitized newspapers, CRL can assess the distribution of coverage of digital content. For example, an assessment of issue coverage in Chronicling America demonstrates consistent coverage of papers over the course of its intended scope (1836–1922), adjusting for publishing trends of the respective time periods.
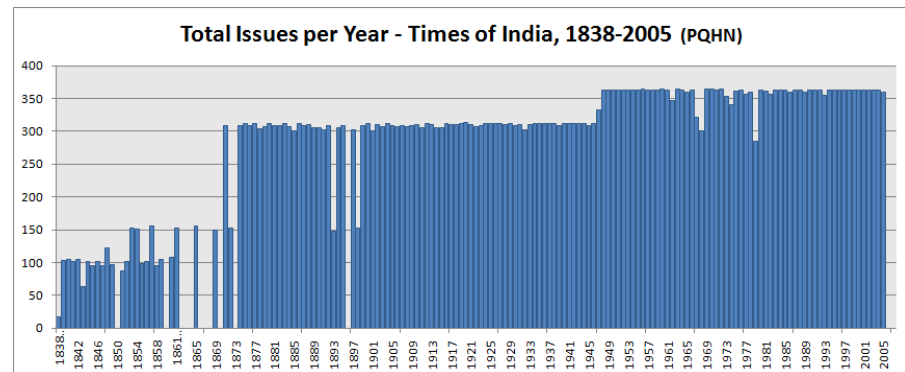

Issue-level metadata provided courtesy of Chronicling America via ICON

This type of assessment can be further broken down into states, or even individual title coverage, if desired.

Similarly, ICON data can be used to assess completeness of coverage of a given product or digitized title. An issue-level assessment of ProQuest's Historical Newspaper digitization of *The Times of India* using data supplied to CRL for ICON can illustrate depth of coverage across years of the product, as below:

Detailed examination of the apparent issue gaps can reveal areas where ProQuest might seek alternative sources to fill in unavailable content (for example, holdings gaps in the early years of 1850, 1862-70, 1873, 1896, and so on).
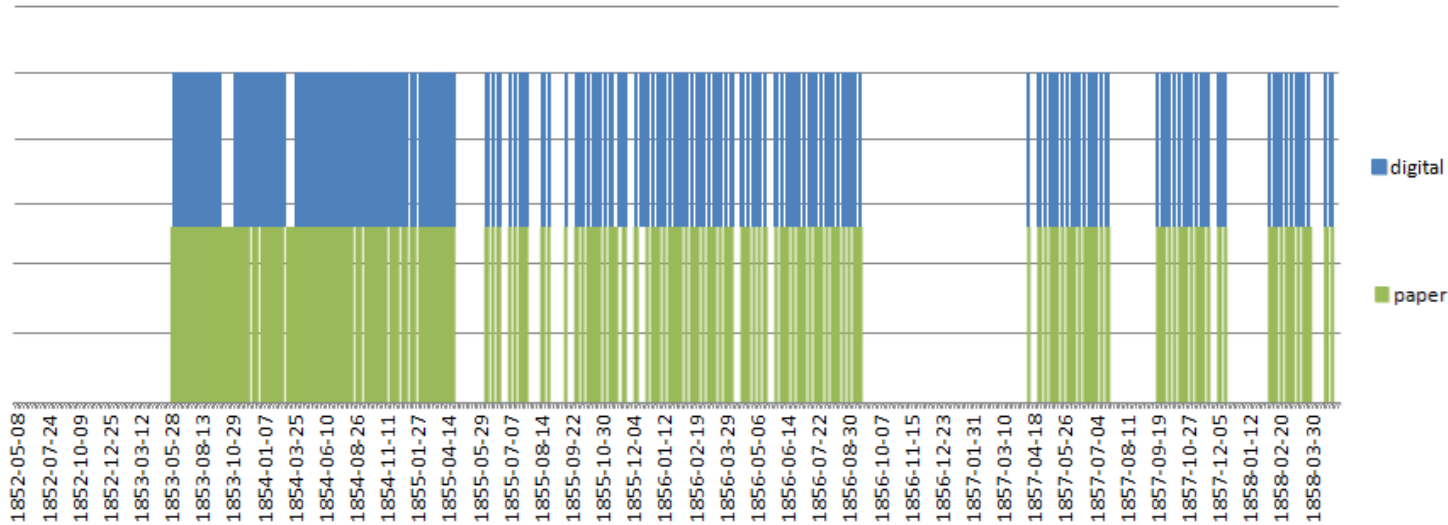

Issue-level metadata provided courtesy of ProQuest via ICON

Case Example: comparison of digital coverage with print source content

Granular data will permit broad comparisons of the coverage of individual titles in a database, to the available print copies held by one (or more) institutions. The graph below compares the digitized title "The Weekly Minnesotan," available in Readex's *Early American Newspapers* series. Content was originally sourced from the holdings of the American Antiquarian Society. In the graphical representation below, we can note differences between the reported print holdings of the AAS and the digitized version in Readex. ICON data shows nine issues (1853-1854) of the paper held by AAS that are not represented in the digital version. Conversely, Readex's data shows five issues digitized that are not held in the AAS repository, suggesting that Readex may have filled gaps from an alternative source.

Issue-level metadata provided courtesy of Readex and American Antiquarian Society via ICON

VI.   Observations and Recommendations

CRL's assessment of 90 distinct collections of digitized newspapers reveals the tremendous diversity of content available across the globe. The following observations draw from this assessment and suggest areas for cooperative action.

1. Too many newspapers remain "at risk."
With the data CRL has gathered thus far, it is difficult to estimate with any precision the amount of newspaper content digitized relative to the total number of newspapers held by libraries worldwide. However, our sampling indicates that the percentage is quite small. Titles that have not been scanned or otherwise preserved remain "at risk" due to the highly acidic nature of much newsprint, potential damage through circulation and handling, and sheer neglect. Many digitization projects are working from microfilm "vaults," which leaves the most at-risk material unavailable.

The scale of the at-risk corpus has to be better understood and to be communicated to funders and other stakeholders. To do this, libraries and publisher must share existing data on both digital and print holdings more broadly. The ICON database is a natural repository for such information.

2. Newspaper digitization remains "selective"
Comprehensive digitization remains a distant goal. It is true that the scale of many digital projects has increased exponentially: in 2005, four million pages seemed ambitious; in 2015, projects tenfold in scale seem achievable. Still, the volume of content still "on the shelves" and limitations on local capacity mean most institutions are limited in how much—and in what timeframe—they can make their holdings digitally accessible.

The overwhelming majority (87%) of materials digitized thus far dates from prior to the mid-twentieth century. The bulk of newspapers in print and microform are, as yet, relatively untouched. Given the scope of the remaining challenge (and the risks to the collection described above), libraries and commercial producers alike should be more strategic in their investment in newspaper digitization. Again, greater sharing of granular data on library newspaper holdings, on the contents of newspaper databases, and even on titles in the digitization pipeline will make those investment decisions more informed.

3. Newspaper digitization is driven by "local" and market imperatives.
Newspaper content, much of which is born of local events and produced for personal consumption, remains an inherently "local" phenomenon. Digitization of domestic newspaper content by regional or national-level institutions tends to be driven foremost by regional and national imperatives. Kentucky digitizes newspapers from Kentucky: England digitizes content published in the U.K.

Digitization projects that serve local and national agendas are, of course, the most likely to attract public funding. Unfortunately, public funding is in decline in many economies. And this dynamic may result in the loss or permanent inaccessibility of a great deal of historical material relating to regions and locales that do not have well-developed public library and archive infrastructure and funding to begin with.

Digitization by commercial aggregators and publishers, on the other hand, is naturally driven by market imperatives. This results most often in a focus on newspapers for which sizable communities of consumers exist, and neglect of newspapers produced in communities whose citizens do not constitute a potentially significant consumer base. Newspapers of tremendous research value published in

developing regions, in languages other than English, and by immigrant and diaspora communities in developed regions, are therefore at considerable risk.

Some libraries and archives have digitized content beyond their regional mandates. The Bibliotheque nationale de France includes in its *Gallica* digital library selective news titles from Algeria, Madagascar, Vietnam, and other former French colonies. The National Library of the Netherlands has digitized content from Nederlands-Indië and related areas, and the Berlin State Library has included a selection German-language papers published internationally. The CRL/Readex World Newspaper Archive emphasizes digitization of titles from world regions receiving less preservation attention. Given the scarcity of funding available to sustain this kind of activity, it will require coordination of efforts, and even cooperation, among U.S., U.K., and European libraries to make an impact. To subsidize digitization of newspaper from outside the developed world a sizable, international pool of investors with specialized interests in these regions will be necessary. The research universities in North America, the U.K., and Europe constitute the most likely pool of such investors.

Providing data to ICON on international and foreign-language newspaper holdings and on such materials currently in the digitization pipeline can be a positive first step toward such coordination.

4. Digitization and access are not synonymous
Digital access has become the preferred—and expected—medium for researchers, students, genealogists, and other primary source users. However, even as libraries move to adopt electronic legal deposit and put processes in place for digital preservation, widespread access to recent content is still hampered by publisher and wire service copyright, author rights, and other restrictions on intellectual property. While print content continues to stack up, the limitations on digitization beyond the mid-twentieth century means the percentage of newspaper back files that remain "offline" is growing.

While national libraries are beginning to scan/capture an array of current domestic newspapers, most of those titles are inaccessible to users beyond the library premises or select participating libraries. For example, most of Norway's digitized newspapers from the copyright period are available only onsite in the national library. Similarly, Germany's digitized East German newspapers are inaccessible to users outside Germany, due to restrictions put in place by agreement with the publishers.

5. We need more data
Effectively managing and providing access to historical newspapers are matters of consequence for libraries. Academic and national libraries invest considerable sums to digitize newspapers to make them accessible to historians and other researchers. And each year, research libraries in the aggregate spend millions of dollars to purchase databases of digitized historical newspapers from commercial publishers.

As the thrust of this report suggests, objectively measuring and assessing the state of newspaper digitization is challenging. To measure across collections and gauge our collective progress, we need more information about titles, years, and issues produced. CRL maintains that exposure of metadata at the issue level should be considered a basic prerequisite of transparency—and thus trustworthiness—in library and commercial databases.

Optimal access to structured content metadata would be through call technologies such as APIs, Open Access metadata, or other means that allow information to be freely harvested (metadata CC:0) and to keep information on repositories up to date. Failing that, libraries and commercial publishers should agree on a common set of metadata that can be shared and exchanged.

ICON offers a mechanism for sharing such information, and should be considered by libraries and commercial providers alike as a platform for exposing the outcomes of our shared effort. ICON's metadata specifications for submission are included here as Appendix B.

Appendix A – Digital Newspaper Products / Libraries surveyed in this assessment:

Libraries / Repositories
Austria	Austrian National Library
Belgium	Bibliothèque royale de Belgique
Denmark	Danish State and University Library
	Kongelige Bibliotek
England/United Kingdom	British Library
	British Newspaper Archive, Ltd.
	Cambridge Public Library
	King's College London
	National Library of Wales
Finland	National Library of Finland
France	Bibliothèque nationale de France
Germany	Bavarian State Library
	Hamburg State Library
	Landesbibliothek Wiesbaden
	Staats- und Universitätsbibliothek Dresden
	Staatsbibliothek zu Berlin
	Thüringer Universitäts- und Landesbibliothek
	Universitäts- und Landesbibliothek Bonn
	Universitäts- und Landesbibliothek Darmstadt
	Universitätsbibliothek Augsburg
	Universitätsbibliothek Freiburg
	Universitätsbibliothek Heidelberg
Netherlands	Koninklijke Bibliotheek
Isle of Man	Isle of Man National Library
Italy	Biblioteca comunale dell'Archiginnasio
	Biblioteca Nazionale Braidense
	Biblioteca Panizzi Reggio Emilia
	Biblioteca Sportiva Nazionale
	Biblioteca Universitaria di Pisa
	Dr. Friedrich Teßmann Library
Norway	National Library of Norway
Sweden	National Library of Sweden
Switzerland	Biblioteca cantonale di Lugano
	Bibliothèque cantonale et universitaire – Lausanne
	Swiss National Library
	Société Neuchâteloise de Presse SA
United States	Center for Research Libraries
	Library of Congress (Chronicling America)
	Library of Virginia
	Louisiana State University
	Montana Memory
	Portal to Texas History
	University of California-Riverside
	University of Hawaii
	University of Kentucky Libraries

University of Oregon
University of Utah
Indiana State Library

<u>Commercial Publishers/Providers</u>

Frankfurter Allgemeine Zeitung (Germany)
Gale Cengage
Hamburger Abendblatt (Germany)
Irish Newspaper Archives Ltd
Le Monde (France)
ProQuest
La Provincia (Italy)
Readex
Schaffhauser Nachrichten (Switzerland)
La Stampa (Italy)
Le Temps (Switzerland)
UKPressOnline
L'Unita' (Italy)
Die Zeit (Germany)

Appendix B – Metadata Field List for ICON Database

Guidelines for submission by third-party publishers.

Title-Level Metadata

Title ID	Unique identifier for Title assigned by provider	Required
OCLC	OCLC Record number(s)	At least one is required. Any/all such numbers for the given title are desired.
LCCN	Library of Congress Control Number	
ISSN	International Standard Serial Number	
Publication Title	Title assigned by provider—uniform, family or title for entire publication.	Required
City	City of publication	Required
Country	Country of publication	Required
Publication Date One	Beginning publication year (YYYY)	Required if present
Publication Date Two	Ending publication year (YYYY)	Required if present
Frequency	Publication frequency or pattern of publication. Include frequencies (current, former) and date ranges for when each frequency applies. Corresponds to MARC 310 and 321 fields	Required if present
Other information about dates of publication	Information from MARC 362 or 310 fields, or other sources of dates of publication [beginning date, end date, or variances in publication]. Submit as separate date fields	Required if present
First Date Held	First issue held by provider (YYYY-MM-DD)	Required
Last Date Held	Last issue held by provider (YYYY-MM-DD)	Required
Collection	Name of product or collection	Required

Issue Level Metadata

Issue ID	Unique identifier for Issue assigned by provider	Required
Masthead data	Data recorded from keyed masthead [including title as published on given Issue; other fields?]	Required if present
Date	YYYY-MM-DD	Required
Volume Number	Volume number	Required if present
Issue Number	Issue number	Required if present
Edition	Edition designation	Required if present
Number of pages	Number of pages included in issue.	Required
Source*	Source of the film/print from which the issue was digitized. May include "Provenance" (source from which Issue was procured), or	Required if present

	"Reproducer" (production details for the filmer of given Issue)	
Additional Source Information	Cataloging or identifying information known about source	Required if present
Completeness	Any information about incomplete issues due to missing or damaged pages.	Required if present

Preferred format to receive data:

- CSV text file

- UTF-8 character set (so as to correctly represent diacritics or non-English characters)

- fields separated by the 'pipe' character **|**