

# How Large Is the “Public Domain”? A Comparative Analysis of Ringer’s 1961 Copyright Renewal Study and HathiTrust CRMS Data

John P. Wilkin

The 1961 Copyright Office study on renewals, authored by Barbara Ringer, has cast an outsized influence on discussions of the U.S. 1923–1963 public domain. As more concrete data emerge from initiatives such as the large-scale determination process in the Copyright Review Management System (CRMS) project, questions are raised about the reliability or meaning of the Ringer data. A closer examination of both the Ringer study and CRMS data demonstrates fundamental misunderstandings and misrepresentations of the Ringer data, as well as possible methodological issues. Estimates of the size of the corpus of public domain books published in the United States from 1923 through 1963 have been inflated by problematic assumptions, and we should be able to correct mistaken conclusions with reasonable effort.



The distinctive nature of U.S. copyright for the period covering 1923–1963 creates opportunities for making available recent creative works and invites speculation about the size of the public domain in that period. U.S. law required compliance with certain rules for works published through most of the 20th century, and many books published in the United States during that period entered the public domain as a consequence of failure to comply with those rules.<sup>1</sup> One of those rules, a requirement that the copyright holder for a work published between 1923 and 1963 renew the copyright of the work 28 years after it was published, was the subject of an important Copyright Office study discussed here.<sup>2</sup> Many cite this study to suggest that 93 percent of books published in the United States during this period are in the public domain.<sup>3</sup> Recent work by the IMLS-funded CRMS project, a HathiTrust project focused on books digitized from research libraries, found a significantly smaller percentage of public domain books for the same period, approximately 50 percent.<sup>4</sup>

The significant difference in the numbers established by these two efforts is puzzling. It should be noted that there is no literature that examines this discrepancy between the two important sources. However, as librarians, copyright experts, and interested observers have discussed this difference, they have offered a number of hypotheses to try to reconcile these two apparently accurate and yet very different pieces of information.

---

*John P. Wilkin is Juanita J. and Robert E. Simpson Dean of Libraries and University Librarian at the University of Illinois at Urbana-Champaign; e-mail: jpwilkin@illinois.edu. ©2017 John P. Wilkin, Attribution-NonCommercial (<http://creativecommons.org/licenses/by-nc/4.0/>) CC BY-NC.*

One may wonder whether one or both of these efforts made errors in their analyses. Setting aside that possibility, in conversations about this issue, most commentators have suggested the discrepancy may be traced to content-type differences between the corpus of materials from research libraries under examination by the CRMS and the corpus reviewed by the Copyright Office report—for example, that the CRMS has reviewed primarily scholarly works while the Copyright Office registered copyright for more popular works.<sup>5</sup>

Why should we be concerned about the apparent discrepancy between these two U.S. copyright data points? After all, each effort was conceived for a different purpose, and one could reasonably argue that each may be entirely accurate as articulated and that each serves its own purpose without conflict with the other. On the other hand, the two studies suggest something very different about the size of the public domain for this period, especially for books published in the United States. The most urgent need for a better understanding of this problem is the need for facts that aid libraries in decision making. A better understanding of the size of the public domain, gaps in the portion of the public domain that has been digitized, the specific characteristics of the in-copyright corpus, and the problems and opportunities in the remainder can help drive digitization and rights clearance efforts. More important, better facts are needed to help shape a host of library programs such as preservation efforts and shared print initiatives. For example, shared print monograph initiatives typically give consideration to whether storage decisions can be shaped by online access to corresponding digital versions; in those discussions, the size of the public domain is typically considered. The most important reason for having a more accurate sense of the extent of the public domain for this period is the way that accurate, evidence-based accounting of the available facts can shape deliberate action by libraries. Especially in digitization and “collective collection” activities, library efforts are hampered by imperfect “facts” or an absence of information.

In the research library community, we struggle with questions such as the extent of the corpus of U.S. federal documents, the size of the collective North American research library monograph collection, or more general questions such as the number of books published in the United States in a given period. Although our literature rarely highlights the problem of a lack of sufficient data for decision making, these gaps in information often color important strategic discussions. For example, in 2012 a group of four consortia met to chart a strategy for U.S. federal government documents digitization. The absence of good bibliographic data on the size and characteristics of the corpus left the group to speculate about the nature of the challenge based on estimates that varied widely: these estimates of the size of the corpus ranged from 1.8 million volumes to 2.2 million volumes, and from an average of 60 pages per volume to 300 pages per volume. The vastness of the range—between 108 million pages and 660 million pages—further complicated efforts by introducing a dramatic difference in cost and difficulty. The 2010 Lyrasis-IMLS meeting on “Developing a North American Strategy to Preserve & Manage Print Collections of Monographs” concluded through a vote of invited participants that the community’s highest priority for print storage efforts should be “public domain and published up to 1976” monographs in HathiTrust and “already in storage.”<sup>6</sup> Discussions at the meeting struggled with determining a goal and strategy at least in part because there were no data on the scope of such an effort. As noted by Demas and Brogdon in their earlier work on copyright determination in support of preservation reformatting, “A curious lack of systematic investigation of [copyright status and permissions seeking] by the library community has effectively stymied large-scale library-based efforts to preserve” this portion of our collective collection.<sup>7</sup>

With regard to the question of the size of the 1923–1963 U.S. public domain book corpus, in conversations with those who have some expertise in this area, there is some sentiment that the assumptions created by Copyright Office study are, even if wrong, helpful in that they might stimulate a “gold rush” looking for the public domain, and perhaps even that a clearer picture of the facts would be a nuisance. In research library strategy meetings, book-related data from Ringer’s Copyright Office study is sometimes invoked to dismiss concerns about the difficulties of copyright determination for the period or to paint a rosy picture about the ease of providing access to this relatively recent corpus of material.

Additional and more accurately interpreted facts will certainly result in our being better able to direct our efforts. As a community, we have been both careless about the challenges and unsystematic about the opportunities. It is hoped that the analysis provided here will shed more light on the question of the size of the U.S. 1923–1963 public domain, as well as the opportunities presented by clearing rights for the remaining in-copyright collection and grappling with challenges in that middle ground of ambiguous and problematic publications.

### **Background Efforts: Ringer’s *Renewal of Copyright* and the CRMS Project**

To better understand why Ringer and the CRMS come to different conclusions and why assertions about the size of the public domain (based on Ringer’s analysis) are problematic, a detailed analysis of Ringer’s work and the CRMS follows. Each section of analysis provides background on the effort in question, either Ringer or the CRMS, followed by an examination of the sources and methodology, the findings, and problems of each.

### **Ringer and the Renewal of Copyright Study**

In 1961, Barbara Ringer published a study on *Renewal of Copyright* for the U.S. Congress on behalf of the Copyright Office. This outstanding work exploring the history, value, and complications related to renewals also contains very important facts intended to help understand patterns of renewals. To help decision makers understand the role and extent of copyright renewal, Ringer’s study sought to determine the frequency with which copyright holders renewed their copyrights for a variety of types of creative works. Ringer’s data and the study itself are frequently cited in support of a number of arguments, including the copyright status of books published in the United States from 1923 through 1963 in research libraries.

### **Ringer’s Sources and Methodology**

Ringer’s study looks at renewals for materials whose copyright was registered in fiscal year 1932. The federal government’s 1932 fiscal year ran from July 1931 through June 1932. Ringer’s data are presented in detail on pages 220–224 of her study.<sup>8</sup> However, Ringer does not clearly identify her sources or her methodology and we are left to speculate about both.

The Copyright Office is the source of all information related to registration and renewal of copyrights in the United States. A limited amount of this information is publicly available outside the Copyright Office itself. One publicly available source of information on registrations and renewals, the *Catalog of Copyright Entries* (CCE), is also a key copyright resource and the foundation for resources such as the *Stanford Copyright Renewal Database*. According to the Copyright Office’s *Circular 22*,

The Copyright Office published the *Catalog of Copyright Entries* (CCE) in printed format from 1891 through 1978. From 1979 through 1982, the CCE was issued in

microfiche format. The CCE is divided into parts according to the classes of works registered. Each CCE segment covers all registrations made during a particular period of time. Renewal registrations made from 1979 through 1982 are found in Section 8 of the catalog. Renewals prior to that time are generally listed at the end of the volume containing the class of work to which they pertained.<sup>9</sup>

While CCE entries include essential information, the Copyright Office's *Circular 22* notes that CCE entries are "not a verbatim transcript of the registration record"<sup>10</sup> and exclude information such as the address of the copyright claimant. The card system at the Copyright Office includes this additional information, and the copyright registration certificates are the most complete and authoritative source of information about a copyright. In their 1997 study, "Determining Copyright Status for Preservation and Access: Defining Reasonable Effort," Demas and Brogdon effectively demonstrated the reliability of using the CCEs in making copyright determinations.<sup>11</sup>

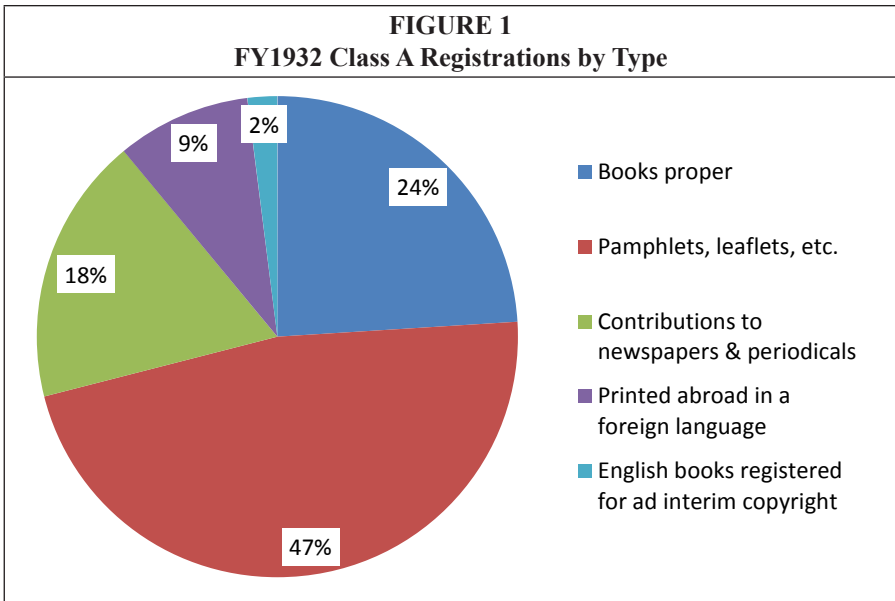
Ringer's analysis probably relied on tallies of renewals (based on Copyright Office catalog card data) performed in 1959–1960 against total numbers of works registered and counted in 1931–1932 and documented in the 1932 *Annual Report* of the Copyright Office. In her 1962 study, Ringer notes that 57,065 items fell into the definition of books referred to as Class A works by the Copyright Office and reports that 3,942 of these works were renewed.<sup>12</sup> The copyrights for these items were registered with the Copyright Office in fiscal year 1932, and their renewals were required 28 years later—in 1958 and 1959. Reproducing the results of the Copyright Office analysis is made challenging by the absence of a methodology in the 1961 report, but we are aided by the 1932 *Annual Report* of the Copyright Office and a footnote in the table included in the 1961 report.

The 1932 *Annual Report* of the Copyright Office spells out the details for Class A registrations that year.<sup>13</sup> Often, the Copyright Office's term "Class A" is used interchangeably with the word "books," though the term includes many other formats. The Copyright Office's 1932 *Annual Report* notes that the copyrights of 57,065 Class A works were registered that year. This is the same number used in the 1961 *Renewal of Copyright* report. Many types of work qualify as Class A works, and the Copyright Office received registrations for non-U.S. works as well. As reported in the 1932 *Annual Report*, the 57,065 Class A registrations included types of works and numbers of each type found in table 1.

A footnote in Ringer's analysis notes that the renewals number "Includes contributions to periodicals," which is consistent with the Class A compilation reported in the

**TABLE 1**  
**1932 Copyright Office *Annual Report* Class A Registrations**

Class	Subject Matter of Copyright	1931–1932
A		Books
	(a) Printed in the United States	
	Books proper	13,460
	Pamphlets, leaflets, etc.	26,995
	Contributions to newspapers and periodicals	10,489
	Total	50,944
	(b) Printed abroad in a foreign language	4,784
	English books registered for ad interim copyright	1,337
	Total	57,065



1932 *Annual Report*. We should assume that Ringer’s number would also include all of the other categories that comprise Class A works, and thus each of the components reported in the 1932 Annual Report. For a breakdown of types of publications in Class A works, see figure 1.

It would have been unlikely for Ringer and her staff to have reassembled all of the FY1932 registrations. It may have seemed unnecessary, as the initial reporting process should have been reliable, and recounting would have been labor-intensive. However, after nearly 30 years, it would have been extremely unlikely that a recount performed for the purposes of the 1961 study would have produced precisely the same number as the initial count. Thus, we should assume that Ringer *only counted renewals* and compared this number to the numbers (by category) recorded and reported in FY1932. As we will see later, because Ringer compares the number of renewals to an undocumented and unpublished set of registrations, this sort of procedure is problematic and makes reproducing the conclusions of the original work difficult, if not impossible.

### Ringer’s Findings

It is critical that we understand what Ringer’s study did and *did not* establish. One of the study’s most often noted conclusions is that copyright was renewed for only 15 percent of the works whose copyright was registered in FY1932. Because this 15 percent is the percentage for a variety of types of works, embedded in this average of 15 percent are lower and higher rates of renewal for specific types of works. For example, Class E works (musical compositions) were renewed at a rate of 35 percent. Notably, Ringer reports that “only 7% of books ... are being renewed.”<sup>14</sup> The conclusion that the copyright for only 7 percent of the FY1932 books was renewed is the crux of much of the analysis that follows.

The renewal rates reported by Ringer are critical data points; from them we can certainly learn much about works where renewal was required, especially those published between 1923 and before 1933, the date this study used for its analysis. If we infer from these data, as many do, that 93 percent of the books published from 1923 through 1963 are therefore in the public domain, we make several critical mistakes.

Consider Ringer's statement that "At present about 15% of subsisting copyrights are being renewed; in fiscal 1959, for example, roughly 21,500 copyrights were renewed, as against 124,500 that went into the public domain at the end of their first 28-year term."<sup>15</sup> That is, Ringer suggests that a failure to renew necessarily moves one of these works into the public domain. If we make this assumption, we overlook some of the characteristics of the fairly large and heterogeneous corpus, especially in Class A works. Moreover, if we infer from these data that the rate of renewal for all books published between 1923 and 1963 remained a constant 7 percent, we fail to take into account data suggesting that the rate of renewals increased over time, a fact that Ringer herself reports in the 1961 study.

### Problems with Ringer's Conclusions

#### *Where Nonrenewal Is Irrelevant*

We need first to understand that some works did not enter the public domain as a result of a failure to renew. One of these classes of works is books "Printed abroad in a foreign language." At the time of Ringer's analysis, renewal was required for these works, but subsequent changes to copyright law removes the renewal requirement for these works. Under current law, as a result of the post-Uruguay round of GATT agreements, renewal is not required for foreign-language books published abroad between 1923 and 1963 and for English-language works published abroad (and not published in the United States within a month) by non-U.S. authors for those same years. Peter Hirtle has explored the renewals question extensively, especially with regard to foreign works, and while he concludes that copyright for all 1923–1963 works is complicated, his analysis makes abundantly clear that most non-U.S. works from this period are protected by copyright.<sup>16</sup>

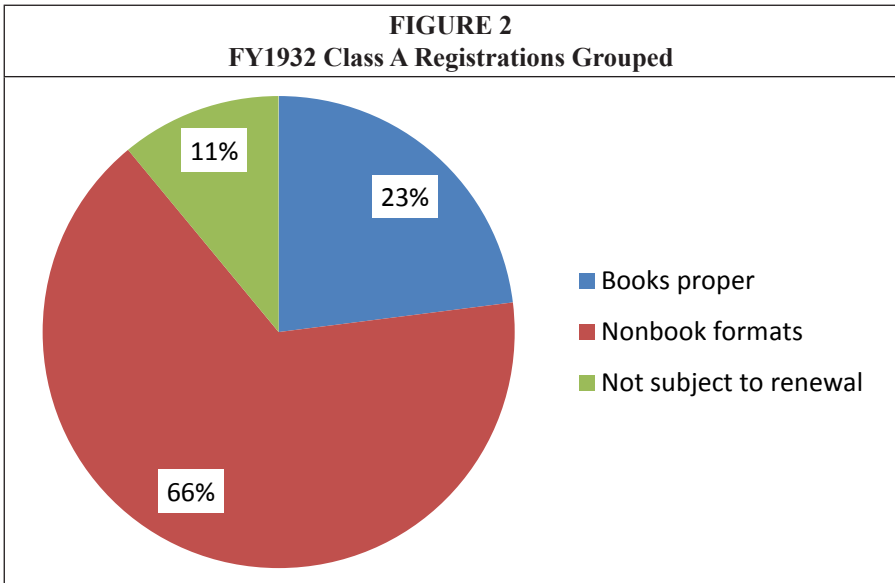
A second class of works to consider here is *ad interim* copyright registrations. Even at the time of Ringer's 1961 report, *ad interim* registrations, defined by Circular 22 as "a special short term of copyright available to certain pre-1978 books and periodicals,"<sup>17</sup> would have been complicated by subsequent publication of the work in the United States. This fairly complicated situation is discussed at length by Peter Hirtle using the example of a Class A work published first in 1939; however, as Hirtle concludes, "Almost by definition, therefore, an *ad interim* copyright means a restored copyright."<sup>18</sup> So, even though both "Printed abroad in a foreign language" and *ad interim* registrations would have been counted among the candidates for renewal, neither could be included *today* to calculate the 7 percent number. Thus, assumptions about public domain status based on the 7 percent renewal rate are related to a corpus that should exclude 11 percent of the works registered in FY1932 (see figure 2). Moreover, in the case of a third class of works, "Contributions to newspapers and periodicals," many would have been afforded copyright protection as a result of separate copyright conditions for the newspapers and periodicals in which they were published. "Contributions to newspapers and periodicals" represents 18.3 percent of the FY1932 Class A works.

#### *The Importance of Terminology*

Terminology, and specifically how we understand "books" and "Class A" works, proves to be one of the most important facets of this analysis. A fuller understanding of what comprises Class A works helps to better understand a major difference between the findings of the Ringer study and what we expect when we discuss the copyright status of "books." Included in Class A works are many items that would not be widely understood to be "books," either in a research library context or popularly. As noted, Class A works include "Contributions to newspapers and periodicals" and "Pamphlets, leaflets, and the like," as well as books. Research libraries do, of course,



collect pamphlets, sometimes in their special collections and sometimes in the general collection as “bound withs.”<sup>19</sup> Leaflets and posters are also collected. And although many of these items may be in research libraries, libraries would not characterize them as “books” for the purposes of cataloging or collection management. Similarly, a popular understanding of “books” would exclude materials that make up 66 percent of the reported FY1932 Class A works. So, importantly, while Ringer says “only 7% of books ... are being renewed,” she means that only 7 percent of Class A works are being renewed and *not* that 7 percent of the 13,460 “books proper” are being renewed.<sup>20</sup> As shown in figure 2, “Books proper” comprise 23 percent of Class A works registered in FY1932, while works not subject to renewal make up 11 percent of the total and nonbook formats comprise the vast majority of works in this classification (66%).



### The Rate of Renewal for “Books Proper”

We can only speculate as to the rate of renewal for different portions of the Class A works. Rights holders are likely to show less interest in renewing leaflets, for example. Book authors and publishers have the greatest economic incentive to renew copyrights, and so it is entirely likely that the preponderance of the 3,942 Class A renewals received were made for the 13,460 “books proper” whose copyright was registered in FY1932. If it were the case that the 3,942 renewals were made primarily against the 13,460 “books proper,” this would be a renewal rate of 29 percent. While the rate of renewal for “books proper” may not be as high as 29 percent (for example, because some foreign works and some submissions to periodicals *were* renewed), the renewal rate for “books” is certain to be several times higher than the 7 percent reported. Based only on the information provided in the 1961 report, we are not able to determine the rate at which the copyright of “books proper” was renewed; however, we can be absolutely certain that Ringer’s statement that “only 7% of books ... are being renewed” is not accurate. We should make some effort to confirm the rate of renewal for “books proper,” but this will prove challenging.

### **The Copyright Review Management System (CRMS)**

In December 2008, a group of institutions led by the University of Michigan began a systematic and large-scale review of the copyright status of books in HathiTrust published in the United States between 1923 and 1963. That initiative received generous support from the Institute of Library and Museum Services (IMLS) for the creation of a strengthened copyright determination system and process.<sup>21</sup> The challenge for these processes is, in a sense, the challenge of proving a negative: those who are interested in establishing the public domain status of a U.S. book published in this period must prove definitively that the book *did not* renew its copyright or *did not* follow other required copyright formalities (such as including a copyright statement). The IMLS-funded effort, referred to as the Copyright Review Management System (CRMS), introduced a unique methodology to help build reliability and thus confidence in making copyright determinations. Although the bulk of the reviews have now been completed (as of December 2015), as previously undigitized candidate U.S. works are digitized and become available in the HathiTrust Digital Library, they are reviewed in the CRMS using the same process. By 2016, more than 300,000 books published in the U.S. between 1923 and 1963 had been reviewed.

### **The CRMS's Sources**

The corpus of materials reviewed by the CRMS is drawn entirely from the HathiTrust Digital Library. The HathiTrust corpus is notable for its size and comprehensiveness. At 13.8 million volumes (in December 2015), the HathiTrust collection is larger than all but the largest research library collections. Overlap analysis performed by OCLC using 2009 and 2010 HathiTrust data demonstrated significant overlap with ARL library print collections.<sup>22</sup> As noted by Malpas in 2011, "In June 2009, an average of 20% of titles held in any given ARL library was duplicated in the HathiTrust Digital Library; by June 2010, the average duplication rate had increased to 30%."<sup>23</sup> The 2010 OCLC analysis was performed when the HathiTrust collection was fewer than 6 million volumes. Now that the HathiTrust Digital Library has more than doubled in size, the median overlap between all ARL collections and HathiTrust will have grown significantly. The immense size and diversity of materials available to the CRMS, as well as the consistency of its findings, suggests that its data for books published in the U.S. from 1923 to 1963 are characteristic of the copyright status for these works in ARL libraries.

### **The CRMS Methodology**

The CRMS methodology was designed to minimize human error, and the project confirmed its reliability by contracting with the U.S. Copyright Office for three independent reviews of samples. The CRMS, through its online system, brings together reviewers from several different institutions (initially Indiana University and the Universities of Illinois, Michigan, and Wisconsin), and reviewers interact with materials through an interface that prioritizes candidate volumes for review. Books published in the United States between 1923 and 1963 comprise the candidates for review work and are assigned by the system to reviewers. The CRMS system makes available to each reviewer a variety of tools, including the Stanford Copyright Renewal database and a digitized copy of the work itself. Once a determination is made by one reviewer, the work is prioritized for review by a second reviewer at another institution. If the system finds that the two reviewers agree on the determination, that conclusion (that is, public domain or in-copyright), along with a reason, is registered in the CRMS system with metadata for the book. If the two reviewers disagree, the disputed determination is sent to an expert reviewer for arbitration. The arbitrated conclusion is then stored in the CRMS database. Because the goal of the CRMS is increasing the reliability of



determinations, the project regularly sampled determinations and paid the Copyright Office to perform research on these conclusions. This metareview of CRMS work found the process to be extremely reliable, ultimately exceeding a 99 percent accuracy rate.<sup>24</sup>

### **The CRMS’s Findings**

The CRMS process has been extremely important for a number of reasons. Perhaps most important is its success in opening access to U.S. books published from 1923 to 1963. It also establishes some very important facts: HathiTrust contains some 14 million books digitized by partner libraries, typically in cooperation with Google, and within that corpus there are some 300,000 books published in the U.S. between 1923 and 1963. By early 2016, these 300,000 books in HathiTrust had been reviewed by staff at HathiTrust partner institutions to determine their copyright status, and, of these, approximately 150,000 had been determined to be in the public domain, either because their copyright had not been renewed or because the work did not include a copyright notice. While roughly 50 percent of the books reviewed were definitively determined to be in the public domain, only about 16 percent were conclusively determined to be in-copyright as the result of a copyright renewal record being located. The remainder, more than 30 percent, remain closed for public access due to challenges in determining their copyright status: for example, the work may also include significant amounts of material likely to be protected by copyright. The findings by the CRMS have been remarkably consistent during the course of the project. That is, from month to month, the CRMS has found roughly 50 percent of the materials being reviewed to be in the public domain and approximately 16 percent to be in-copyright.

The CRMS project stores and shares detailed, database-based data on copyright determinations.<sup>25</sup> To better understand the characteristics of U.S. works published in 1932, a period consistent with Ringer’s analysis, the CRMS project provided the author cumulative CRMS data for determinations made throughout the project until the end of 2015. By January 2016, the CRMS reviewed 5,004 books characterized by catalogers as books having been published in the United States in 1932. In its selection routines, the CRMS tries to exclude books published simultaneously both in and outside the United States. The CRMS report shows that 634 (12.7%) of these volumes were found to be in-copyright as a result of renewal. A work can be determined to be in the public domain for several reasons: for the same period, 67 (1.3%) of the volumes were determined to be in the public domain because they omitted a copyright statement, and 2,759 (55.1%) were determined to be in the public domain because of a failure to renew. The remainder of the volumes remain closed because of complications such as in the inclusion of possibly in-copyright content within the book.

### **Apparent Contradictions in Determining the Size of the Public Domain**

The Ringer study and the CRMS process have been extremely helpful and, indeed, influential in shaping our understanding of the U.S. public domain. And yet the conclusions of the two efforts are at odds with each other. In our work as librarians and digital collection builders, our focus tends to be on the public domain, so the biggest clash is between the putative 93 percent public domain suggested by the Ringer study and the 50 percent public domain found by the CRMS.

### **Errors Based on Problematic Assumptions**

The conclusion that those volumes that were not renewed are in the public domain, encouraged by Ringer herself and a host of sources, is a logical problem. Although Ringer found 7 percent of the books reviewed to have been renewed, we cannot then conclude that the other 93 percent are in the public domain. For example, as noted here,

some of the items reviewed were foreign works with foreign copyrights and others were works with *ad interim* copyright registrations, registrations for works that would have been published and renewed later.

Indeed, this kind of logic is often used to argue that the entire public domain for this period (that is, including all formats) is 85 percent or greater. One university copyright resource page states that:

Estimates are that 85% of copyrights were not renewed (93% in the case of books), most likely because the works were no longer commercially valuable. In addition, works were not protected unless authors included a basic copyright notice—the word “copyright” or © with one’s name and year next to it (this notice requirement was eliminated in 1989). By some estimates, 90% of works did not include this copyright notice and immediately entered the public domain. So, before 1978, only 10% of works might have been subject to copyright at all, and of the works that were, up to 85% only used the first 28-year term, with 15% renewing for the full 56-year term. That’s 1.5% of all works with 56 years of protection, 8.5% with 28 years, and 90% completely free.<sup>26</sup>

Critically important copyright resources imply similar conclusions (that is to say that works not renewed were thus in the public domain) or are not careful to make clear that nonrenewal is only one consideration. For example, the Copyright Sherpa web page implies the nonrenewed works are in the public domain in its statement that “The copyrights for many, many works were not renewed. In fact, the U.S. Copyright Office has estimated that less than 15% of works eligible for renewal were, in fact, renewed. That means *a lot* of works are in the public domain ... but it also means you have to find out whether copyright renewal happened, or didn’t.”<sup>27</sup>

Just as one cannot conclude from a failure to renew that something is in the public domain, we cannot conclude from the CRMS’s data (in other words, that 50 percent of the works are in the public domain) that 50 percent are in-copyright. Only approximately 16 percent have been found to be definitively in copyright, and only a portion of the more than 30 percent that are problematic will be protected by copyright. Thus we find that we should not logically conclude from one set of facts (for example, that a percentage is in the public domain) the other (for instance, the remainder are in copyright).

### Data and Terminology Errors

The reliability of data from both the CRMS process and the 1961 report should be interrogated. As reported earlier, the CRMS subjected itself to ongoing scrutiny and confirmed the reliability of its methodology. It would be helpful to assess the reliability of the conclusions drawn by the Copyright Office in its 1961 analysis by reviewing the renewal data for the 13,460 “books proper.” Aside from card records at the Copyright Office, records from the Copyright Office for this period are only available in the *Catalog of Copyright Entries* (CCE), an annual published by the Copyright Office. Although these annual publications are organized chronologically, their relationship to registrations is only approximate. For example, a volume (such as 1931) may contain works published in a previous year (such as 1930). Moreover, an annual CCE volume will also include *renewals* in addition to *registrations* for U.S. and foreign works, and *ad interim* registrations.

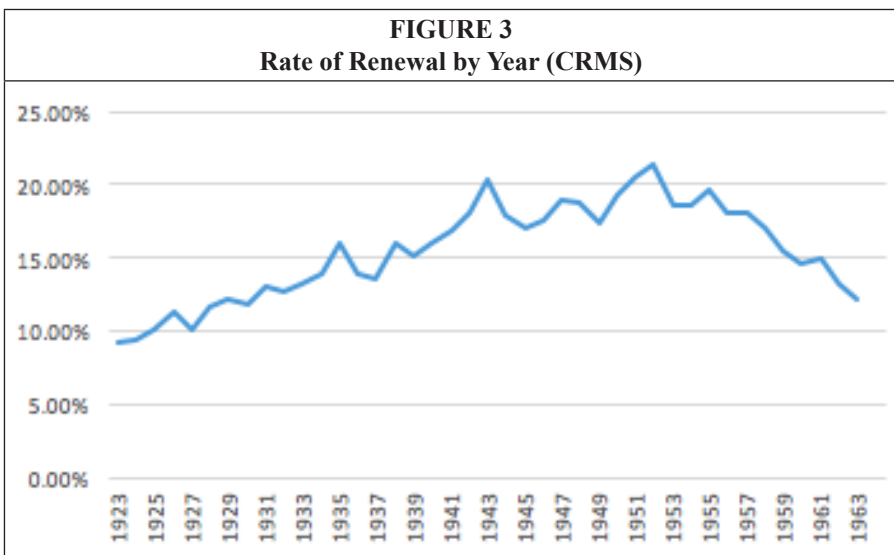
Although these CCE volumes do not correspond directly to the registrations reported in the annual reports of the Copyright Office, they should have a rough correspondence. Despite this, the numbers reported in the CCE volumes for the years corresponding to the 1932 *Annual Report* are curiously inconsistent with the *Annual Report*. The 1932 *Annual Report* tells us that 13,460 “books proper” were registered. The 1931 CCE for

books reports only 9,837 U.S. publications, and the 1932 CCE for books reports only 8,994 U.S. publications. A year later, in 1933, we see no significant variation and, instead, a slight decline: only 8,268 U.S. publications are recorded. Similarly, while the 1932 *Annual Report* notes 4,784 books “printed abroad in a foreign publication,” the CCE volumes for 1931, 1932, and 1933 report respectively 3,357, 2,471, and 3,170 “foreign books in foreign languages.” We are at a loss to explain the roughly 30 percent larger number of “books proper” in the *Annual Report* as compared to the CCE volumes for the corresponding years. Only by comparing card-based records at the Copyright Office to the relevant CCE volumes can we hope to explain the discrepancy between the numbers in the CCEs and the *Annual Report*, but it is likely that the renewals Ringer’s study counted for the FY1932 works corresponds to a smaller candidate group than the 13,460 “books proper” reported by Ringer and the 1932 *Annual Report* (that is, approximately 9,000 books instead), thus increasing the percentage of the rate of renewal further.

**Failure to Take into Account Changes in the Rate of Renewal over Time**

Ringer and the CRMS effort both provide some very important chronological or date-based information about renewals. As Ringer notes, the trend was for renewals to increase with the passage of time. She reports a “dramatic rise in the total percentage of copyrights renewed.” Her report provides data showing annual rates of renewal beginning in 1883, and notes that the rate of renewals for works published between 1940 and 1959 had doubled.<sup>28</sup> She includes a graph, “Appendix C: Graph to Accompany Table 2,” that shows growth in a compelling way.<sup>29</sup>

The CRMS also supports the notion of a general historical trend of renewals increasing after 1932, though not a trend of constant increase. Renewals for books published in 1923 were approximately 9 percent and had grown to 13 percent by 1932. In many years following 1932, the CRMS found renewal rates of 20 percent or more. One of the more peculiar findings of the CRMS project is that rates declined toward the end of the 1923–1963 period after a relatively steady growth. Figure 3 shows this very effectively. A table of renewals and other related data found by the CRMS process is included in appendix A.



Popular resources commonly conclude or imply that Ringer's analysis applies to the entire period before 1964, or even 1978 as in the case of the university copyright web page quoted earlier. A commercially published book by Stephen Fishman guiding readers on exploiting the public domain makes both the problematic chronological and the problematic logical argument: "The Copyright Office estimates that only 15% of all works published *between 1923–1963* were ever renewed. This means that all works first published in the United States from 1923 through 1963 *for which no renewal was filed are in the public domain*" (emphasis added).<sup>30</sup> In fact, Ringer was careful to limit her conclusions to FY1932 works and pointed out that rates of renewal were dramatically increasing. The Stanford Fair Use page appears to imply the findings from renewal rates for FY1932 works are applicable to the entire 1923–1963 period when it writes that "If a work published after 1922 and before 1964 was not renewed, it fell into the public domain. According to Copyright Office surveys, the great majority of *pre-1964* works were never renewed and, therefore, are in the public domain" (emphasis added).<sup>31</sup> In fact, as we know, the Copyright Office report only focused on renewals for works published before 1933.

It is reasonable to assume that the rate of renewals would increase over time and could not be generalized from the 1932 data. Indeed, both Ringer's data about increasing rates of renewal and data from the CRMS process support this likelihood. Ringer does not provide data on renewals for years subsequent to the FY1932 publications, but CRMS data show an increase in renewals after 1932. In their analysis of copyright renewal for core agricultural monographs, Demas and Brogdon also found increasing rates of renewal over time, from "11% for titles published in the 1920s versus 39% for titles published in the 1940s."<sup>32</sup> It would also be reasonable to assume that publication of Ringer's study would stimulate *an increase* in renewals, for example, because publicity around renewal rates would motivate rights holders. Regardless, it is important to keep in mind that the rates of renewal reported by Ringer are only for the FY1932 publications and not for all publishing in the 1923–1963 period.

### Conclusions and Future Work

Estimates of the size of the corpus of public domain books published in the U.S. from 1923 through 1963 have been inflated by problematic assumptions. The apparent differences between numbers reported in Barbara Ringer's 1961 Copyright Office study and data generated by the Copyright Review Management System are, on closer examination, not great. Indeed, the corpus under review by the CRMS process may represent a major portion of the U.S. *book* publishing output for the period; thus, the CRMS data help us better understand the copyright characteristics of books examined by the Ringer study.

Problematic assumptions based on Ringer's data are the source of misunderstandings about the size of the 1923–1963 U.S. book public domain. Those assumptions include the following mistakes:

- Problematic (logical) assumptions, where we assume that nonrenewal necessarily means a work is in the public domain.
- A problem of misinterpreting data, typically by conflating terminology, first, where we assume that Class A works are synonymous with "books" and, second, where Ringer assumes that the corpus of registered works is as large as was reported in the 1932 *Annual Report*.
- A failure to take into account changes in the rate of renewal over time, principally where we assume the conclusion collected for the FY1932 publications applies to publications from the entire period from 1923 through 1963.

Nonetheless, the percentage of the U.S. 1923–1963 books in the public domain is extremely large, and those that are not definitively in the public domain remain an

important corpus for collective attention. If we assume no significant differences from the data generated by the CRMS, we can conclude that approximately 50 percent of the U.S. books from this period are in the public domain (and that many are currently openly accessible in HathiTrust). A further 16 percent are protected by copyright as a result of renewals. This may in fact be a feasibly small body of materials on which the community can perform rights clearance and collective licensing work.

The CRMS has made a tremendous contribution by reviewing what may be the majority of the books published in the United States between 1923 and 1963. With more than 300,000 books reviewed, and many clear determinations about copyright status, future work should be much easier. Still, this analysis of Ringer’s 1961 study and comparison with CRMS data raise many questions and can help guide subsequent work. Suggested areas include:

- **Closer examination of the 1932 *Catalog of Copyright Entries*:** It would be helpful to more carefully compare 1932 copyright registrations to data from Ringer’s study. The Copyright Office has embarked on a process to digitize cards from their catalog, and this would be an ideal source to perform that examination. Alternatively, one could convert the entries in the 1932 CCE to a database so that entries can be categorized (for instance, “books proper” as one category) and conclusions further compared to data from Ringer’s study. It should be possible to determine the rate of copyright renewal for these U.S. books published in 1932. It should also be possible to determine overlap and difference with HathiTrust, helping to better understand the extent to which HathiTrust includes U.S. publishing output and perhaps guiding efforts to augment HathiTrust. Should this analysis find a rate of renewal similar to that found by the CRMS, we may be able to extrapolate CRMS findings to other years.
- **Comparison of the 1932 *Catalog of Copyright Entries* and WorldCat:** WorldCat is an interesting but problematic point of comparison, compiling, as it does, books cataloged rather than books published. Still, it is another data point, albeit a noisy one. The 1932 CCE reports 8,994 U.S. books, while OCLC’s WorldCat reports 55,657 U.S. 1932 books.<sup>33</sup> To what can we attribute this surprising and significant difference? A closer examination of the WorldCat output should help us determine the extent to which that number is inflated by, for example, nonbook material cataloged as books<sup>34</sup> and should also help us understand the extent to which U.S. publishing for the period is not represented in the CCE.
- **Rights clearance/licensing of in-copyright works:** The CRMS’s conclusions about the percentage of works whose copyright has been renewed inspires confidence that this body of relatively contemporary publishing can be opened more generally. In 1997, Demas and Brogdon urged establishing a “reasonable effort procedure for contacting copyright holders and seeking permissions” to advance our collective preservation efforts.<sup>35</sup> A coordinated effort to secure permissions for works that are not on the market may be feasible. Roughly 55,000 such volumes are in HathiTrust now. What is the financial and practical feasibility of such an effort?
- **Determining the ROI for further work on works with a complicated copyright status:** Again, based on the work done by the CRMS, it appears that approximately 30 percent of U.S. 1923–1963 book publishing has complicated rights issues—neither definitively public domain nor definitively in-copyright. This body of material should be analyzed (perhaps sampled) by experts and subjected to collective review. Many of these works will be in the public domain and can add to our collective public good.

We should, with reasonable effort, be able to *correct* the mistaken conclusions drawn from Ringer's 1931 study with some precision. A better understanding of U.S. publishing between 1923 and 1963 can be extremely beneficial for managing and providing access to the total collection.

### **Acknowledgements**

I owe a special debt of thanks to several people who aided me in this analysis. Peter Hirtle patiently worked through many of the issues and challenges with the methodology and sources. John Mark Ockerbloom was generous, both as a guide to nuances in the CCEs and as a reader. The CRMS staff at Michigan, Melissa Levine and Moses Hall, helped clarify CRMS issues and provided invaluable data from the project. And of course my best friends and readers, my wife Maria Bonn and colleague Aaron McCollough, helped untangle my prose.



## APPENDIX A. CRMS Renewal Data

The table included below, table 2, drawn from the CRMS database on January 21, 2016, removes duplicate digitized books, that is copies of the same book that were reviewed twice, and provides a count of total candidate volumes for each year, along with the number of works determined to be in copyright: (1) as a result of renewal; (2) in the public domain as a result of no copyright notice; and (3) in the public domain as a result of nonrenewal. Column headings use documented CRMS terminology.<sup>36</sup> The first code is an "attribute" code and the second is a "reason code." Attributes in this table include IC ("in-copyright") and PD ("public domain"). Reasons include REN ("Copyright renewal research was conducted") and NCN ("no printed copyright notice").

<b>Year</b>	<b>Total</b>	<b>IC/REN</b>	<b>PD/NCN</b>	<b>PD/REN</b>	<b>Pct Renews</b>
1923	6,246	578	80	3,320	9.25%
1924	5,686	529	74	3,034	9.30%
1925	5,694	572	80	3,073	10.05%
1926	5,785	658	95	3,078	11.37%
1927	6,541	659	70	3,375	10.07%
1928	6,379	740	88	3,547	11.60%
1929	6,203	759	89	3,338	12.24%
1930	6,752	794	102	3,602	11.76%
1931	5,813	759	60	3,092	13.06%
1932	5,004	634	67	2,759	12.67%
1933	4,693	618	78	2,452	13.17%
1934	4,827	668	64	2,506	13.84%
1935	5,488	879	86	2,779	16.02%
1936	7,525	1,041	111	3,042	13.83%
1937	7,074	962	145	3,862	13.60%
1938	6,354	1,016	87	3,303	15.99%
1939	6,651	1,011	111	3,436	15.20%
1940	6,891	1,097	117	3,616	15.92%
1941	6,804	1,142	104	3,486	16.78%
1942	6,177	1,113	91	3,114	18.02%
1943	5,486	1,114	85	2,654	20.31%
1944	5,147	924	77	2,537	17.95%
1945	5,355	914	77	2,673	17.07%
1946	6,311	1,108	77	3,101	17.56%
1947	7,404	1,406	165	3,412	18.99%
1948	7,603	1,426	77	3,515	18.76%
1949	8,225	1,435	86	3,887	17.45%
1950	8,735	1,690	103	3,763	19.35%

**TABLE 2**  
**CRMS Renewal Data (January 21, 2016)**

Year	Total	IC/REN	PD/NCN	PD/REN	Pct Renews
1951	7,521	1,546	86	3,369	20.56%
1952	7,838	1,681	98	3,371	21.45%
1953	7,748	1,439	108	3,273	18.57%
1954	8,679	1,616	89	3,529	18.62%
1955	8,655	1,696	105	3,729	19.60%
1956	8,677	1,561	98	3,766	17.99%
1957	9,377	1,692	108	3,886	18.04%
1958	9,528	1,619	110	4,364	16.99%
1959	10,563	1,639	106	4,392	15.52%
1960	12,586	1,831	115	5,585	14.55%
1961	12,628	1,892	131	5,557	14.98%
1962	14,106	1,855	124	5,940	13.15%
1963	14,797	1,800	184	6,224	12.16%

### Notes

1. In addition to the rules discussed here for works published in the United States between 1923 and 1963, U.S. law required a copyright notice for all works published between 1923 and 1977, and registration within five years of publication for works published without notice between 1978 and March 1, 1989.

2. Barbara Ringer, "Study No. 31: Renewal of Copyright" (1960), reprinted in *Library of Congress Copyright Office. Copyright Law Revision: Studies Prepared for the Subcommittee on Patents, Trademarks, and Copyrights of the Committee on the Judiciary, United States Senate, Eighty-sixth Congress, first [-second] session.* (Washington, D.C.: U.S. Govt. Print. Office, 1961).

3. James Boyle and Jennifer Jenkins, *Intellectual Property: Law & The Information Society: Cases & Materials* (Durham, N.C.: Center for the Study of the Public Domain, 2014), 294. While other examples will be cited later, significant works like Boyle and Jenkins refer to the Ringer work to argue that "the majority of works [entered] the public domain after the first 28-year term (studies put the rate of nonrenewal for all works at 85 percent, and for books alone at 93 percent)."

4. The CRMS project is documented online at [www.lib.umich.edu/copyright-review-management-system-ims-national-leadership-grant](http://www.lib.umich.edu/copyright-review-management-system-ims-national-leadership-grant) [accessed 12 December 2016]. A full description of the project is included there.

5. Samuel Demas and Jennie L. Brogdon, "Determining Copyright Status for Preservation and Access: Defining Reasonable Effort," *Library Resources and Technical Services* 41, no. 4 (Oct. 1997): 323-34. A similar line of reasoning was advanced by Demas and Brogdon in this work on copyright determination. In this analysis comparing agriculture monographs to the figures promulgated by the Copyright Office, they note: "The renewal rate on a group of qualitatively selected, core scholarly monographs might reasonably be far higher than that of books and pamphlets as a whole" (329). They may have been correct, and were surely right if we include pamphlets.

6. From a November 2010 summary document circulated to participants.

7. Samuel Demas and Jennie L. Brogdon, "Determining Copyright Status for Preservation and Access," 324.

8. Barbara Ringer, "Study No. 31: Renewal of Copyright." See "Appendix C: A Statistical Survey of Renewal Registrations."

9. *How to Investigate the Copyright Status of a Work: Circular 22* (Library of Congress, Copyright Office, 2004), available online at [www.copyright.gov/circs/circ22.pdf](http://www.copyright.gov/circs/circ22.pdf) [accessed 12 December 2016].

10. *Circular 22*, 2.

11. Demas and Brogdon, "Determining Copyright Status for Preservation and Access: Defining Reasonable Effort," 323-34.

12. Ringer, "Study No. 31: Renewal of Copyright," 221.

13. *Thirty-Fifth Annual Report of the Register of Copyrights for the Fiscal Year Ending June 30, 1932*

(Library of Congress, Copyright Office, 1932), 17, available online at [www.copyright.gov/reports/annual/archive/ar-1932.pdf](http://www.copyright.gov/reports/annual/archive/ar-1932.pdf) [accessed 12 December 2016].

14. Ringer, “Study No. 31: Renewal of Copyright,” 220.

15. Ringer, “Study No. 31: Renewal of Copyright,” 187. Emphasis added to highlight the way that Ringer suggests a status of public domain from a failure to renew.

16. Peter Hirtle, “Copyright Renewal, Copyright Restoration, and the Difficulty of Determining Copyright Status” *D-Lib Magazine* 14, no. 7/8 (July/Aug. 2008), available online at [www.dlib.org/dlib/july08/hirtle/07hirtle.html](http://www.dlib.org/dlib/july08/hirtle/07hirtle.html) [accessed 12 December 2016]. Hirtle’s primary argument is that the effect of “restoration” creates confusion for all copyright determinations for this period. He writes that “it is almost impossible to determine with certainty whether a work published between 1923 through 1963 in the US is in the public domain because of copyright restoration of foreign works.” Considerable successful work (such as by the CRMS) has been done to make determinations of the copyright status of works published in this period. We can agree, however, that the complications Hirtle documents with regard to foreign works are immense and that, in most cases, these works would be protected by copyright.

17. *Circular 22*, 6.

18. Hirtle, “Copyright Renewal, Copyright Restoration, and the Difficulty of Determining Copyright Status.”

19. As one library notes, “bound withs” are bound volumes “containing two or more works bound together after publication by someone other than the publisher,” frequently a library. These volumes present discovery problems because the works included are usually not cataloged separately. Note, for example, the University of Illinois Rare Book and Manuscript Library instructions for cataloging these individual works, available online at [www.library.illinois.edu/rbx/qnc/procedures/bound\\_withs.html](http://www.library.illinois.edu/rbx/qnc/procedures/bound_withs.html) [accessed 12 December 2016]. The CRMS process excludes “bound withs” in its review process.

20. Ringer, “Study No. 31: Renewal of Copyright,” 220.

21. The author was the Principal Investigator of the first IMLS grant awarded for the CRMS.  
22. Lorcan Dempsey, Brian Lavoie, and Constance Malpas, *Understanding the Collective Collection: Towards a System-wide Perspective on Library Print Collections* (2013), available online at [www.oclc.org/content/dam/research/publications/library/2013/2013-09.pdf](http://www.oclc.org/content/dam/research/publications/library/2013/2013-09.pdf) [accessed 12 December 2016].

23. <http://www.oclc.org/content/dam/research/publications/library/2011/2011-01.pdf> [accessed 12 December 2016].

24. The final report by the CRMS project to IMLS is available online at [www.lib.umich.edu/files/services/copyright/2012-03-06\\_CRMS-US\\_Final\\_Report\\_to\\_IMLS-narrative\\_only.pdf](http://www.lib.umich.edu/files/services/copyright/2012-03-06_CRMS-US_Final_Report_to_IMLS-narrative_only.pdf) [accessed 12 December 2016] and notes that “two audits we undertook with the US Copyright Office’s own records showed marked improvements in the accuracy and reliability of our reviews” (2). The results of those reviews are not available in published documents.

25. Detailed CRMS data is publicly available to us through HathiTrust’s monthly bibliographic data reports, the hathifiles. The compact monthly and cumulative reports include a single line record for each item in HathiTrust and its copyright status. The hathifiles are documented online at <https://www.hathitrust.org/hathifiles> [accessed 12 December 2016], and codes for CRMS categories are documented online at [https://www.hathitrust.org/rights\\_database](https://www.hathitrust.org/rights_database) [accessed 12 December 2016].

26. Duke Law Center for the Study of the Public Domain, “Public Domain Day—Frequently Asked Questions,” available online at <https://web.law.duke.edu/cspd/publicdomainday/2014/faqs> [accessed 31 January 2016].

27. Public Domain Sherpa, “Copyright Renewal: When It Had to Happen, or Else,” available online at [www.publicdomainsherpa.com/copyright-renewal.html](http://www.publicdomainsherpa.com/copyright-renewal.html) [accessed 31 January 2016].

28. Ringer, “Study No. 31: Renewal of Copyright,” 221.

29. Ringer, “Study No. 31: Renewal of Copyright,” 223.

30. Stephen Fishman, *The Public Domain: How to Find & Use Copyright-Free Writings, Music, Art & More* (Berkeley, Calif.: Nolo, 2008), 336.

31. Stanford University Libraries, Copyright & Fair Use, “Searching the Copyright Office and Library of Congress Records,” available online at <http://fairuse.stanford.edu/overview/copyright-research/searching-records/> [accessed 31 January 2016].

32. Demas and Brogdon, “Determining Copyright Status for Preservation and Access: Defining Reasonable Effort,” 328. Note, too, that they found an overall rate of renewal of 18 percent, a number very similar to the rate of renewal found by the CRMS effort.

33. Work performed in December 2015 by the University of Illinois Library’s Cataloging and Access Management unit, using GLIMIR-based deduplication algorithms.

34. The cataloging of nonbook material as monographs or “books” is a common problem in WorldCat.

35. Demas and Brogdon, "Determining Copyright Status for Preservation and Access: Defining Reasonable Effort," 333.

36. For example, [www.lib.umich.edu/files/services/copyright/CRMS-World\\_Cheat\\_Sheet-v2-0.pdf](http://www.lib.umich.edu/files/services/copyright/CRMS-World_Cheat_Sheet-v2-0.pdf) and [https://www.hathitrust.org/rights\\_database](https://www.hathitrust.org/rights_database) [accessed 12 December 2016].